LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Fakultät für Medizin

Institut für Psychiatrische Phänomik & Genomik

TOWARDS A FULLER PICTURE OF THE GENETIC ARCHITECTURE OF NEUROPSYCHIATRIC DISORDERS

THE RARE, THE COMMON, AND EVERYTHING IN BETWEEN

Kumulative Habilitationsschrift

vorgelegt von

Dr. med. Eva Christina Schulte, Ph.D.

(2022)

Für Lisa, Horst & Marianne

This work is based on the following publications:

Zimprich A, Benet-Pages A, Struhal W, Graf E, Eck SH, Offman MN, Haubenberger D, Spielberger S, **Schulte EC**, Lichtner P, Rossle SC, Klopp N, Wolf E, Seppi K, Pirker W, Presslauer S, Mollenhauer B, Katzenschlager R, Foki T, Hotzy C, Reinthaler E, Harutyunyan A, Kralovics R, Peters A, Zimprich F, Brücke T, Poewe W, Auff E, Trenkwalder C, Rost B, Ransmayr G, Winkelmann J, Meitinger T, Strom TM. Exome sequencing reveals mutations in the retromer protein VPS35 as cause for Parkinson`s disease. Am J Hum Genet 2011, 89: 168-175. PMID: 21763483 **(IF 2011: 10.6)**

Schulte EC, Ellwanger DC, Dihanich S, Manzoni C, Stangl K, Schormair B, Graf E, Eck S, Mollenhauer B, Haubenberger D, Pirker W, Zimprich A, Brücke T, Lichtner P, Peters A, Gieger C, Trenkwalder C, Mewes HW, Meitinger T, Lewis PA, Klünemann HH, Winkelmann J. Rare variants in LRRK1 in Parkinson`s disease. Neurogenetics 2014, 15:49-57. PMID: 24241507 **(IF 2014: 3.5)**

Schulte EC, Stahl I, Czamara D, Ellwanger DC, Eck S, Graf E, Mollenhauer B, Zimprich A, Lichtner P, Haubenberger D, Pirker W, Brücke T, Bereznai B, Molnar MJ, Peters A, Gieger C, Müller-Myhsok B, Trenkwalder C, Winkelmann J. Rare variants in PLXNA4 and Parkinson's diasease. PLoS One 2013, 8:e79145. PMID: 24244438 **(IF 2013: 3.5)**

Schulte EC*, Claussen MC*, Jochim A, Haack T, Hartig M, Hempel M, Prokisch H, Haun-Jünger U, Winkelmann J, Hemmer B, Förschler A, Ilg R. Mitochondrial membrane protein associated neurodegeneration: a novel variant of neurodegeneration with brain iron accumulation. Mov Disord 2013, 28: 224-227. PMID: 23436634 **(IF 2013: 5.6)**

Schulte EC, Knauf F, Kemlink D, Schormair B, Gieger C, Lichtner P, Meitinger T, Winkelmann J. Variant Screening of the Coding Regions of MEIS1 in Patients with Restless Legs Syndrome. Neurology 2011,76:1106-1108. PMID: 21422461. **(IF 2011: 8.3)**

Schulte EC*, Kousi M*, Tan P, Schormair B, Knauf F, Lichtner P, Trenkwalder C, Högl B, Frauscher B, Berger K, Fietze I, Gross N, Stiasny-Kolster K, Oertel W, Bachmann C.G, Paulus W, Zimprich A, Peters A, Gieger C, Meitinger T, Müller-Myshok B, Katsanis N, Winkelmann J. Targeted Resequencing and Systematic In Vivo Functional Testing Identifies Rare Variants in MEIS1 as Significant Contributors to Restless Legs Syndrome. Am J Hum Genet 2014, 95:85-95. PMID: 24995868 **(IF 2014: 10.9)**

Schulte EC, Kurz A, Alexopoulos P, Hampel H, Peters A, Gieger C, Rujescu D, Diehl-Schmid J, Winkelmann J. Excess of rare coding variants in PLD3 in late but not early-onset Alzheimer's disease. Hum Genome Var 2015, 2:14028. PMID: 27081517. **(IF 2019: 1.3)**

Schulte EC, Mollenhauer B, Zimprich A, Bereznai B, Lichtner P, Haubenberger D, Pirker W, Brücke T, Molnar MJ, Peters A, Gieger C, Trenkwalder C, Winkelmann J. Variants in eurkaryotic translation initiation factor 4G1 in sporadic Parkinson's disease. Neurogenetics 2012, 13: 281-285. PMID: 22707335 **(IF 2012: 3.4)**

Schulte EC, Fukumori A, Mollenhauer B, Hor H, Arzberger T, Perneczky R, Kurz A, Diehl-Schmid J, Hüll M, Lichtner P, Eckstein G, Zimprich A, Haubenberger D, Pirker W, Brücke T, Bereznai B, Molnar MJ, Lorenzo-Betancor O, Pastor P, Peters A, Gieger C, Estivill X, Meitinger T, Kretzschmar HA, Trenkwalder C, Haass C, Winkelmann J. Rare variants in β-Amyloid precursor protein (APP) and Parkinson's disease. Eur J Hum Genet 2015, 23: 1328-1333. PMID: 25604855. **(IF 2015: 4.6)**

Zech M, Nübling G, Castrop F, Jochim A, **Schulte EC**, Mollenhauer B, Lichtner P, Peters A, Gieger C, Marquardt T, Vanier MT, Latour P, Klünemann HH, Trenkwalder C, Diehl-Schmid J, Perneczky R, Meitinger T, Oexle K, Haslinger B, Lorenzl S, Winkelmann J. Niemann-Pick C disease gene mutations and age-related neurodegenerative disorders. PLoS One 2013,8:e82879. PMID: 24386122 **(IF 2013: 3.5)**

Winkelmann J, Schormair B, Czamara D, Knauf F, **Schulte EC**, Trenkwalder C, Dauvilliers Y, Polo O, Högl B, Berger K, Fuhs A, Gross N, Stiasny-Kolster K, Oertel W, Bachmann CG, Paulus W, Xiong L, Montplaisir J, Rouleau GA, Fietze I, Vavrova J, Kemlink D, Sonka K, Nevsimalova S, Lin SC, Wszolek Z, Vilarino-Guell C, Farrer MJ, Gschliesser V, Frauscher B, Falkenstetter T, Poewe W, Allen RP, Earley CJ, Ondo WG, Le WD, Spieler D, Kaffe M, Zimprich A, Kettunen J, Perola M, Silander K, Cournu-Rebeix I, Francavilla M, Fontenille C, Fontaine B, Vodicka P, Prokisch H, Lichtner P, Peppard P, Faraco J, Mignot E, Gieger C, Illig T, Wichmann HE, Müller-Myhsok B, Meitinger T. Genome-wide association study identifies novel Restless Legs Syndrome susceptibility loci on 2p14 and 16q12.1. PLoS Genetics 2011, 7: e1002171. PMID: 21779176 **(IF 2011: 10.2)**

Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, Adams MJ, Agerbo E, Air TM, Andlauer TMF, Bacanu SA, Baekvad-Hansen M, Beekman AFT, Bigdeli TB, Binder EB, Blackwood DRH, Bryois J, Buttenschon HN, Bybjerg-Grauholm J, Cai N, Castelao E, Christensen JH, Clarke TK, Coleman JIR, Colodro-Conde L, Couvy-Duchesne B, Craddock N, Crawford GE, Crowley CA, Dashti HS, Davies G, Deary IJ, Degenhardt F, Derks EM, Direk N, Dolan CV, Dunn EC, Fley TC, Eriksson N, Escott-Price V, Kideh FHF, Finucane HK, Forstner AJ, Frank J, Gaspar HA, Gill M, Giusti-Rodrigues P, Goes FS, Gordon SD, Grove J, Hall LS, Hannon E, Hansen CS, Hansen TF, Herms S, Hickie IB, Hoffmann P, Homuth G, Horn C, Hottenga JJ, Hougaard DM, Hu M, Hyde CL, Ising M, Jansen R, Jin F, Jorgenson E, Knowles JA, Kohane IS, Kraft J, Kretzschmar WW, Krogh J, Kutalik Z, Lane JM, Li Y, Li Y, Lind PA, Liu X, Lu L, MacIntyre DJ, MacKinnon DF, Maier RM, Maier W, Marchini J, Mbarek H, McGrath P, McGuffin P, Medland SE, Mehta D, Middeldorp CM, Mihailov E, Milaneschi Y, Milani L, Mill J, Mondimore FM, Montgomery GW, Mostafavi S, Mullins N, Nauck M, Ng B, Nivard MG, Nyholt DR, O'Reilly PF, Oskarsson H, Owen MJ, Painter JN, Pedersen CB, Pedersen MG, Peterson RE, Pettersson E, Peyrot WJ, Pistis G, Posthuma D, Purcell SM, Quiroz JA, Qvist P, Rice JP, Riley BP, Rivera M, Saeed Mirza S, Saxena R, Schoevers R, **Schulte EC**, Shen L, Shi J, Shyn SI, Sigurdsson E, Sinnamon GBC, Smit JH, Smith DJ, Stefansson H, Steinberg S, Stockmeier CA, Streit F, Strohmaier J, Tansey KE, Teismann H, Teumer A, Thompson W, Thomson PA, Thorgeirsson TE, Tian C, Traylor M, Treutlein J, Trubetskoy V, Uitterlinden AG, Umbricht D, Van der Auwera S, van Hemert AM, Viktorin A, Visscher PM, Wang Y, Webb BT, Weinsheimer SM, Wellmann J, Willemsen G, Witt SH, Wu Y, Xi HS, Yang J, Zhang F; eQTLGen; 23andMe, Arolt V, Baune BT, Berger K, Boomsma DI, Cichon S, Dannlowski U, de Geus ECJ, DePaulo JR, Domenici E, Domschke K, Esko T, Grabe HJ, Hamilton SP, Hayward C, Heath AC, Hinds DA, Kendler KS, Kloiber S, Lewis G, Li QS, Lucae S, Madden PFA, Magnusson PK, Martin NG, McIntosh AM, Metspalu A, Mors O, Mortensen PB, Müller-Myhsok B, Nordentoft M, Nöthen MM, O'Donovan

MC, Paciga SA, Pedersen NL, Penninx BWJH, Perlis RH, Porteous DJ, Potash JB, Preisig M, Rietschel M, Schaefer C, Schulze TG, Smoller JW, Stefansson K, Tiemeier H, Uher R, Völzke H, Weissman MM, Werge T, Winslow AR, Lewis CM, Levinson DF, Breen G, Børglum AD, Sullivan PF; **Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium**. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 2018, 50: 668-681. PMID: 29700475. **(IF 2018: 25.5)**

Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, Als TD, Bigdeli TB, Børte S, Bryois J, Charney AW, Drange OK, Gandal MJ, Hagenaars SP, Ikeda M, Kamitaki N, Kim M, Krebs K, Panagiotaropoulou G, Schilder BM, Sloofman LG, Steinberg S, Trubetskoy V, Winsvold BS, Won HH, Abramova L, Adorjan K, Agerbo E, Al Eissa M, Albani D, Alliey-Rodriguez N, Anjorin A, Antilla V, Antoniou A, Awasthi S, Baek JH, Bækvad-Hansen M, Bass N, Bauer M, Beins EC, Bergen SE, Birner A, Bøcker Pedersen C, Bøen E, Boks MP, Bosch R, Brum M, Brumpton BM, Brunkhorst-Kanaan N, Budde M, Bybjerg-Grauholm J, Byerley W, Cairns M, Casas M, Cervantes P, Clarke TK, Cruceanu C, Cuellar-Barboza A, Cunningham J, Curtis D, Czerski PM, Dale AM, Dalkner N, David FS, Degenhardt F, Djurovic S, Dobbyn AL, Douzenis A, Elvsåshagen T, Escott-Price V, Ferrier IN, Fiorentino A, Foroud TM, Forty L, Frank J, Frei O, Freimer NB, Frisén L, Gade K, Garnham J, Gelernter J, Giørtz Pedersen M, Gizer IR, Gordon SD, Gordon-Smith K, Greenwood TA, Grove J, Guzman-Parra J, Ha K, Haraldsson M, Hautzinger M, Heilbronner U, Hellgren D, Herms S, Hoffmann P, Holmans PA, Huckins L, Jamain S, Johnson JS, Kalman JL, Kamatani Y, Kennedy JL, Kittel-Schneider S, Knowles JA, Kogevinas M, Koromina M, Kranz TM, Kranzler HR, Kubo M, Kupka R, Kushner SA, Lavebratt C, Lawrence J, Leber M, Lee HJ, Lee PH, Levy SE, Lewis C, Liao C, Lucae S, Lundberg M, MacIntyre DJ, Magnusson SH, Maier W, Maihofer A, Malaspina D, Maratou E, Martinsson L, Mattheisen M, McCarroll SA, McGregor NW, McGuffin P, McKay JD, Medeiros H, Medland SE, Millischer V, Montgomery GW, Moran JL, Morris DW, Mühleisen TW, O'Brien N, O'Donovan C, Olde Loohuis LM, Oruc L, Papiol S, Pardiñas AF, Perry A, Pfennig A, Porichi E, Potash JB, Quested D, Raj T, Rapaport MH, DePaulo JR, Regeer EJ, Rice JP, Rivas F, Rivera M, Roth J, Roussos P, Ruderfer DM, Sánchez-Mora C, **Schulte EC**, Senner F, Sharp S, Shilling PD, Sigurdsson E, Sirignano L, Slaney C, Smeland OB, Smith DJ, Sobell JL, Søholm Hansen C, Soler Artigas M, Spijker AT, Stein DJ, Strauss JS, Świątkowska B, Terao C, Thorgeirsson TE, Toma C, Tooney P, Tsermpini EE, Vawter MP, Vedder H, Walters JTR, Witt SH, Xi S, Xu W, Yang JMK, Young AH, Young H, Zandi PP, Zhou H, Zillich L; HUNT All-In Psychiatry, Adolfsson R, Agartz I, Alda M, Alfredsson L, Babadjanova G, Backlund L, Baune BT, Bellivier F, Bengesser S, Berrettini WH, Blackwood DHR, Boehnke M, Børglum AD, Breen G, Carr VJ, Catts S, Corvin A, Craddock N, Dannlowski U, Dikeos D, Esko T, Etain B, Ferentinos P, Frye M, Fullerton JM, Gawlik M, Gershon ES, Goes FS, Green MJ, Grigoroiu-Serbanescu M, Hauser J, Henskens F, Hillert J, Hong KS, Hougaard DM, Hultman CM, Hveem K, Iwata N, Jablensky AV, Jones I, Jones LA, Kahn RS, Kelsoe JR, Kirov G, Landén M, Leboyer M, Lewis CM, Li QS, Lissowska J, Lochner C, Loughland C, Martin NG, Mathews CA, Mayoral F, McElroy SL, McIntosh AM, McMahon FJ, Melle I, Michie P, Milani L, Mitchell PB, Morken G, Mors O, Mortensen PB, Mowry B, Müller-Myhsok B, Myers RM, Neale BM, Nievergelt CM, Nordentoft M, Nöthen MM, O'Donovan MC, Oedegaard KJ, Olsson T, Owen MJ, Paciga SA, Pantelis C, Pato C, Pato MT, Patrinos GP, Perlis RH, Posthuma D, Ramos-Quiroga JA, Reif A, Reininghaus EZ, Ribasés M, Rietschel M, Ripke S, Rouleau GA, Saito T, Schall U, Schalling M, Schofield PR, Schulze TG, Scott LJ, Scott RJ, Serretti A, Shannon Weickert C, Smoller JW, Stefansson H, Stefansson K, Stordal E, Streit F, Sullivan PF, Turecki G, Vaaler AE,

Vieta E, Vincent JB, Waldman ID, Weickert TW, Werge T, Wray NR, Zwart JA, Biernacka JM, Nurnberger JI, Cichon S, Edenberg HJ, Stahl EA, McQuillin A, Di Florio A, Ophoff RA, Andreassen OA. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet 2021, 53:817-829. PMID: 34002096. **(IF 2021: 23.1)**

TABLE of CONTENT

ABBREVIATIONS

AD	Alzheimer's disease
APP	*β-amyloid precursor protein*
Arg	arginine
bp	base pairs
BPD	bipolar disorder
BTBD9	*BTB/POZ-domain-containing protein 9*
BWA	Burrows-Wheeler Aligner
Cas9	CRISPR-associated protein 9
CNV	copy-number variant
CRISPR	clustered regularly interspaced short palindromic repeats
COVID-19	coronavirus disease-19
EIF4G1	*eukaryotic translation initiation factor 4G1*
ENCODE	Encyclopedia of DNA Elements
FDR	false discovery rate
FTLD	frontotemporal lobar degeneration
FUS	*fused in sarcoma*
Glu	glutamate
Gly	glycine
GRN	*granulin*
GTEx	genotype-tissue expression project
GWAS	genome-wide association study
HEK	human embryonic kidney
His	histidine
KORA	Kooperative Gesundheitsforschung in der Region Augsburg
LD	linkage disequilibrium
Lys	lysine
MAF	minor allele frequency
MALDI-TOF	matrix-assisted laser desorption time-of-flight mass spectrometry
MAPT	*microtubule-associated protein tau*
MAP2K5	mitogen-activated protein kinase kinase 5
MEIS1	*myeloid ecotropic viral integration site 1 homolog*
MDD	major depressive disorder
NHLBI-GO-ESP	National Heart, Lung, and Blood Institute Grand Opportunity Exome Sequencing Project
NPC	Niemann-Pick type C
OD	odds ratio
ORF	open reading frame
PD	Parkinson's disease
PGC	Psychiatric Genomics Consortium
PLD3	*phospholipase D3*
PRS	polygenic risk scores
PSEN1	*presenilin 1*
PSEN2	*presenilin 2*
PTPRD	protein tyrosine phosphatase receptor type D
RLS	restless legs syndrome

RVAS	rare variant association study
Ser	serine
SCZ	schizophrenia
SKAT	sequence kernel association test
SKOR1	*SKI family transcriptional corepressor 1*
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
TDP-43	*TAR DNA-binding protein-43*
TOX3	*TOX high mobility group box family member 3*
UTR	untranslated region
WES	whole exome sequencing
WGS	whole genome sequencing

1 INTRODUCTION

1.1 SETTING THE STAGE

The past decade has witnessed a striking development of the scope of analyses possible in genetics. Analyses have progressed from single variant and candidate gene studies via genome-wide association studies (GWAS) utilizing genetic information from millions of single nucleotide polymorphisms (SNPs) to studies assessing all variants—including structural variants—across an exome or genome of an individual(McGuire et al., 2020; Smoller, 2017; Sullivan and Geschwind, 2019). With the continuously decreasing costs of sequencing, increasing computational power(Antonarakis et al., 2010; Mardis, 2011), and ever larger fully sequenced, deeply phenotyped cohorts—such as in the UK Biobank(Bycroft et al., 2018), the Million Veteran Program(Gaziano et al., 2016) or All of Us(All of Us Research Program et al., 2019)—available, this trend towards the superlatives in genetics is sure to continue for some time. However, this will also present challenges because the road from knowledge of a genomic region harboring factors of increased disease risk to understanding its biological implications and to translating this knowledge to patient care is still long and rocky(Sullivan and Geschwind, 2019; Tolosa et al., 2020).

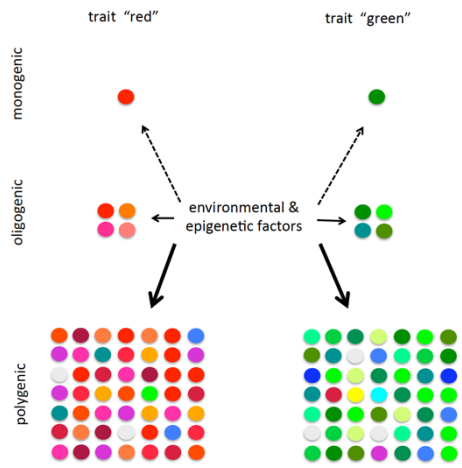


Figure 1: Schematic representation of different genetic architectures that can underlie different traits or diseases. In reality, these should be viewed as a continuum more than as three separate entities. (taken from (Schulte, 2013))

Most neuropsychiatric disorders encountered in quotidian clinical practice are genetically complex. That is to say that not one single genetic alteration is sufficient to cause disease but that a number of different genetic factors act in unison to increase an individual`s likelihood of developing a certain condition. Even among the genetically complex disorders, the degree of polygenicity and of variant frequency and effect size is not the same in all common, complex disorders, yielding a unique disease-specific (or subphenotype-specific) genetic architecture (Kendall et al., 2021) (Figure 1).

Although there can be significant overlap in the genetic factors contributing to a group of disorders such as in severe psychiatric disorders spanning the affective and

psychotic spectrum(PGC, 2019), hardly any two genetic architectures are alike. Even among common, complex neuropsychiatric conditions, the relative contributions of genetic variants vary widely with regard to the number of involved loci(Kendall et al., 2021; Zhang et al., 2018), the different variants' effect sizes on a given phenotype(Winkelmann et al., 2011; Wray et al., 2018; Zhang et al., 2018), and variant frequencies(Schulte et al., 2014b; Winkelmann et al., 2011; Winkelmann et al., 2007). One aspect of special interest with regard to the study of common, complex neuropsychiatric diseases is the fact that despite massive efforts by large international consortia (International Parkinson Disease Genomics, 2020; Schizophrenia Working Group of the Psychiatric Genomics, 2014; Stahl et al., 2019; Wray et al., 2018) to decipher the genetic architecture, there is still a large amount of "missing heritability"(Maher, 2008), that cannot be accounted for in current studies. One long-standing hypothesis to explain this conundrum which is observed across all common, complex disorders is that not only a very large number of common variants with generally small effect sizes contribute to the genetic architecture ("common variant, common disease hypothesis"(Pritchard and Cox, 2002)) but that rare or intermediate frequency variants—which were not represented on the most widely used genotyping arrays employed in GWAS could account for some of this missing heritability ("rare variant, common disease hypothesis") (Figure 2). It is becoming increasingly clear that in many neuropsychiatric disorders the genetic architecture will be a combination of common, rare, and intermediate frequency variants with different effect sizes on the phenotype and in different proportions. This conception could have important implications both for the study of the underlying pathophysiology—which in many

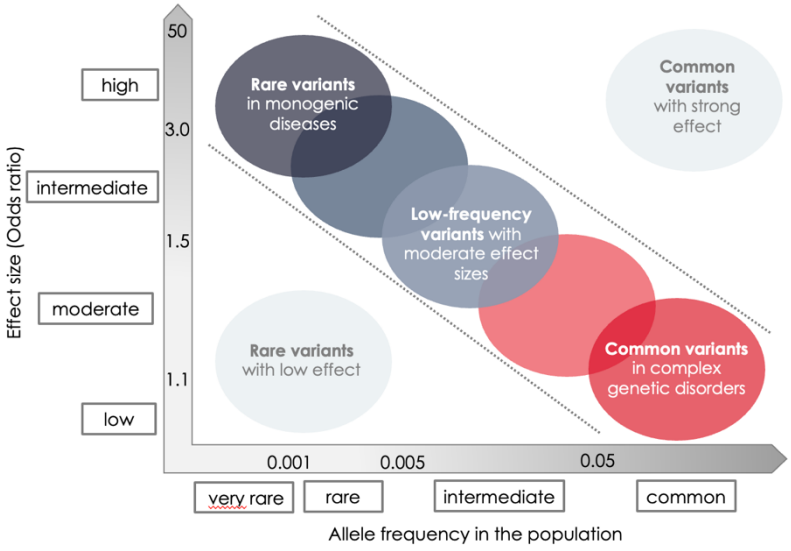


Figure 2: Relationship between allele frequency and effect size of a genetic variant. Generally, the rarer the variant, the stronger its effect on a given phenotype. (adapted from (Manolio et al., 2009))

cases is still only incompletely understood—but also for potential applications of genetics in common neuropsychiatric disorders in the clinical setting.
Consequently, it is vitally important to understand the contributions of different classes of genetic variation to neuropsychiatric disorders. It is upon this backdrop that the studies depicted herein were conducted.

1.2 GENETIC VARIATION

Errors in DNA replication in humans occur at a rate of roughly one in every 10 to 100 million bases(Strachan, 2011). On the whole, estimates of the mutation rate range around $1.1x10^{-8}$ per site across the entire genome per generation(Kondrashov, 2003; Nachman and Crowell, 2000; Roach et al., 2010). The resulting genetic variants differ in size (i.e. from one to several tens of thousands of bases or entire chromosomes) and composition (i.e. structural variants like copy-number variants (CNVs) vs. non-structural variants like single nucleotide variants (SNVs)). Genetic variants can further be classified according to their genomic location with respect to annotated genes (i.e. within inter- vs. intragenic regions or in intronic vs. exonic regions).
Beyond these well-known forms of genetic variation, additional levels of genetic variation exist that include, but are not limited to, mobile genetic elements like retrotransposons(Kazazian et al., 1988; Solyom and Kazazian, 2012) or the introduction of additional changes at the RNA level (RNA editing)(Song et al., 2012). These support the notion that our current understanding of the diverse forms of genetic alterations that play a role in bringing about human phenotypes is still fragmented and far from complete.
Next to classification schemes largely rooted in the "physical" characteristics of a variant, variant frequency in a population (usually reported using the frequency of the less common of two alleles at a given locus (i.e. the minor allele frequency (MAF)) and the strength of effect on a given phenotype are two additional, very important means to classify genetic variation (also cf. Figure 2). The MAF is utilized to assess how frequently a given variant occurs in a given population. Although often times grouped into two ("common" and "rare") or three ("common", "low-frequency", and "rare") categories, variant frequency actually spans the entire frequency spectrum from variants of near equal distribution in the population to those found in only one single individual in the population. Although definitions for the categories above have been subject to change with the field's increasing ability to evaluate variants of lower and

lower frequency, as a general rule, today, variants with MAF >1% can be classified as "common", those with MAF < 0.1% as "rare", and those with 0.1% ≥ MAF ≤ 1% as "low-frequency" variants. It should be noted, however, that when the entire human genome is assessed, rare variants—although individually rare—are by far the most frequent type of genetic variation. At the same time, common variants with MAF > 5% are responsible for the largest proportion of genetic differences between individuals(Fu et al., 2013; Tennessen et al., 2012).

Next to a variant's frequency, its effect size is of great importance to consider when evaluating its contribution to the genetic architecture of a given trait. Effect sizes are commonly measured as odds ratios (ORs). For example, an OR of 2 would indicate that an individual with the given variant would be twice as likely to display a given phenotype as an individual not harboring that variant.

Although exceptions prove the rule(Klein et al., 2005), an inverse correlation appears to exist between the frequency of a given variant and its effect size in the vast majority of cases(Manolio et al., 2009). This is predominantly due to the fact that variants of large effect are subject to purifying selection. Deleterious variants are selected against and prevented from becoming increasingly frequent and eventually fixed in a given population. Today, methodologies are in place to—sample size and financial resources permitting—examine all variants along the frequency and effect size axes.

2 METHODS

The large spectrum of different kinds of genetic variation that contribute to the genetic architecture of biological traits also means that a number of different methodologies and varied approaches are needed to address these different forms of genetic variation. The fact that the past decade has seen enormous technological advances in the field of genetics has further increased the number of potential investigative avenues. At the same time, financial constraints have limited the large-scale roll-out of some of these like, for example, whole exome (WES) or whole genome sequencing (WGS). Here, it has only recently and in very large global consortial efforts been possible to successfully perform so-called "rare variant association studies" (RVAS), that seeking to establish population-scale associations between rare variants

identified by means of WES and specific disorders (e.g. (Fu, 2021a; Palmer, 2020; Satterstrom et al., 2019; Singh, 2020)).

The studies conducted as part of this work also reflect these changes in the methodologies used and the scale at which to address the overarching question of how different genetic variants contribute to the genetic make-up of different neuropsychiatric disorders (Gloyn and McCarthy, 2010). Nonetheless, the conducted works can be subdivided into those geared at investigating the role of rare genetic variants and those primarily focusing on common genetic variants.

Targeted resequencing as well as short-read WES on Illumina® platforms was used to investigate rare variants, both in classical family-based approaches as well as in population-based association designs (Hopfner et al., 2013; Schulte et al., 2013a; Schulte et al., 2014a; Schulte et al., 2015a; Schulte et al., 2011; Schulte et al., 2014b; Schulte et al., 2015b; Schulte et al., 2012; Schulte et al., 2013b; Zimprich et al., 2011). Targeted resequencing also highlighted low-frequency variants of intermediate effect size with a role in restless legs syndrome (Schulte et al., 2014b). Common variants were associated with specific phenotypes like RLS, major depressive disorder (MDD), or bipolar disorder (BPD) using array-based genome-wide genotyping and genome-wide association studies (Mullins et al., 2021; Winkelmann et al., 2011; Wray et al., 2018).

Classical molecular functional follow-up studies of the identified variants represent important tools to bridge the gap between the genetic underpinnings and an understanding of the pathophysiologic changes underlying a given disorder. In the projects conducted as part of this work, *in vitro* cell culture models and primary patient fibroblasts as well as *in vivo* approaches in zebrafish were used to this end(Schulte et al., 2014a; Schulte et al., 2015a; Schulte et al., 2014b; Schulte et al., 2013b). For example, systematic functional follow-up of rare non-synonymous coding variants in genes harboring common risk factors for RLS by *in vivo* complementation in zebrafish identified a specific isoform of the gene *MEIS1* of particular relevance to RLS(Schulte et al., 2014b). Tracking of this isoform can then be used to uncover the developmental time points and locations of greatest importance to the development of RLS, which largely still remain to be uncovered.

3 PROJECTS

In the following, the individual projects relevant to this work will be subsumed into three topic areas highlighting the contribution of different classes of genetic variation to the investigated common, complex neuropsychiatric phenotypes.

3.1 Rare Variants

The first group of projects assesses the role of rare genetic variation—at the time, defined as having a MAF < 5%—in common and genetically complex neuropsychiatric disorders like Parkinson's disease (PD) or restless legs syndrome (RLS).

3.1.1 Family studies to link novel rare variants to disease

Zimprich A, Benet-Pages A, Struhal W, Graf E, Eck SH, Offman MN, Haubenberger D, Spielberger S, **Schulte EC**, Lichtner P, Rossle SC, Klopp N, Wolf E, Seppi K, Pirker W, Presslauer S, Mollenhauer B, Katzenschlager R, Foki T, Hotzy C, Reinthaler E, Harutyunyan A, Kralovics R, Peters A, Zimprich F, Brücke T, Poewe W, Auff E, Trenkwalder C, Rost B, Ransmayr G, Winkelmann J, Meitinger T, Strom TM. Exome sequencing reveals mutations in the retromer protein VPS35 as cause for Parkinson`s disease. Am J Hum Genet 2011, 89: 168-175. PMID: 21763483 **(IF 2011: 10.6)**

Schulte EC, Ellwanger DC, Dihanich S, Manzoni C, Stangl K, Schormair B, Graf E, Eck S, Mollenhauer B, Haubenberger D, Pirker W, Zimprich A, Brücke T, Lichtner P, Peters A, Gieger C, Trenkwalder C, Mewes HW, Meitinger T, Lewis PA, Klünemann HH, Winkelmann J. Rare variants in LRRK1 in Parkinson`s disease. Neurogenetics 2014, 15:49-57. PMID: 24241507 **(IF 2014: 3.5)**

Schulte EC, Stahl I, Czamara D, Ellwanger DC, Eck S, Graf E, Mollenhauer B, Zimprich A, Lichtner P, Haubenberger D, Pirker W, Brücke T, Bereznai B, Molnar MJ, Peters A, Gieger C, Müller-Myhsok B, Trenkwalder C, Winkelmann J. Rare variants in PLXNA4 and Parkinson's diasease. PLoS One 2013, 8:e79145. PMID: 24244438 **(IF 2013: 3.5)**

Schulte EC*, Claussen MC*, Jochim A, Haack T, Hartig M, Hempel M, Prokisch H, Haun-Jünger U, Winkelmann J, Hemmer B, Förschler A, Ilg R. Mitochondrial membrane protein associated neurodegeneration: a novel variant of neurodegeneration with brain iron accumulation. Mov Disord 2013, 28: 224-227. PMID: 23436634 **(IF 2013: 5.6)**

One of the most classical approaches to identify genetic variants underlying disease is to study families. The four studies depicted here used strategies like whole exome sequencing in families with familial PD or other rare neurologic disorders (e.g. neurodegeneration with brain iron accumulation (NBIA)) to identify rare genetic

variants of putatively strong effect and with high penetrance that cause or modify the clinical phenotype. Families were recruited both in our own outpatient clinic at Klinikum rechts der Isar, Technical University Munich, but also at collaborating centers specializing in movement disorders at Schön Klinik Schwabing (Prof. Andres Ceballos-Baumann, Munich, Germany), Paracelsus-Elena-Klinik (Prof. Claudia Trenkwalder, Kassel, Germany), and Allgemeines Krankenhaus Wien (Prof. Alexander Zimprich, Vienna, Austria). Extensive families were subsequently phenotyped and biosampled in field studies.

Short read, paired-end sequencing was performed on the Genome Analyzer IIx platform from Illumina®. Agilent® SureSelect Human All Exon kits were used for in-solution enrichment. Reads were aligned with BWA and variants called with SAMtools. Filtering strategies in order to identify the most relevant rare genetic variants included—besides an overlap with the pedigree—filtering out variants with MAF >0.01 in dbSNP, in 1076 to 1739 in-house exomes, the 1,000 genomes, or the NHLBI-GO-ESP exomes (n=4250). Further, variants predicted to change the amino acid sequence (i.e. missense, nonsense, stop-loss, splice site, or frameshift variants or insertions and deletions) were selected and, in most instances, *in silico* prediction scores of variant deleteriousness by prediction algorithms like PolyPhen(Ramensky et al., 2002), PolyPhen2(Adzhubei et al., 2010), Mutation Taster(Schwarz et al., 2010), or SIFT(Ng and Henikoff, 2003) were included in the filtering strategy. Depending on the specific study, linkage analysis of families using SNP array genotypes and the software MERLIN(Schulte et al., 2013b) or custom bioinformatics algorithms combining predictions from a number of publicly available prediction tools(Schulte et al., 2014a) were employed in addition to enhance the filtering strategy. All prioritized variants were verified by Sanger sequencing and segregation analyses were performed in the families to assess penetrance.

Through this approach, rare non-synonymous variants in *C19orf1*, *EEF1D*, *LRRK1*, *PLXNA4*, and *VPS35*, were prioritized for further analysis (Table). At a time, when large public repositories of genetic variation where not yet available, the next step in the case of prioritized variants for PD entailed a frequency assessment of the candidate variants in large sets of 860 to 975 case and 1014 to 1568 control individuals. These were carried out using MALDI-TOF mass spectrometry on the Sequenom® platform. Association was tested using the allelic test in PLINK. Genes with (suggestive) evidence for association in frequency assessments were taken further into targeted resequencing. Idaho® Light Scanner high-resolution melting curve analysis was repeatedly used to this end to screen the coding regions as well as the exon-intron boundaries of prioritized genes for additional rare variants.

Gene	Reference
EEF1D	(Schulte et al., 2014a)
LRRK1	(Schulte et al., 2014a)
PLXNA4	(Schulte et al., 2013b)
VPS35	(Zimprich et al., 2011)

Table: Genes harboring rare variants prioritized for a role in familial PD using WES

Burden testing to investigate the cumulative burden of rare genetic variation across a given gene, a functional domain of a gene, or a set of genes was carried out using Fisher's Exact and $\chi 2$ statistics. For example, for *PLXNA4*, the rarer and the more likely predicted to be "deleterious" the variant was, the more likely it was found in individuals with PD (p=0.033 to p=0.018; Fisher's Exact and $\chi 2$ tests as appropriate)(Schulte et al., 2013b).

Especially in cases where statistical support for the variant alone was insufficient, functional follow-up experiments were performed. However, if the precise function of a gene in relevance to a given phenotype is unknown, the first stratum of investigations comprises very general experiments. In these, which included assessments of protein expression levels, subcellular localization of proteins, and cell viability in both established cell lines like HEK293 cells as well as primary patient fibroblasts, none of the prioritized variants were linked to significant alterations likely to be related to the assessed neuropsychiatric phenotype.

The projects above led to the identification of potentially and likely causal novel variants associated with PD or NBIA. The role of variants in *EEF1D*, *LRRK1*, and *PLXNA4* in PD is still a matter of debate. However, the identification of rare variants in *PLXNA4* in individuals with PD(Schulte et al., 2013b) has implicated axonal guidance as a novel mechanism to be considered in the study of the pathophysiology underlying PD. After

the initial back-to-back identification in our and another study(Vilarino-Guell et al., 2011; Zimprich et al., 2011), *VPS35*, on the other hand, has become an established PD gene. Since its first description, dysfunction of the retromer complex has evolved into an established pathomechanism not only for PD but also for other neurodegenerative disorders and mouse models as well as cerebral organoids of individuals carrying pathogenic variants in *VPS35* have recapitulated many phenotypic aspects of PD (e.g.(Chen et al., 2019; Chiu et al., 2020; Dhungel et al., 2015; Hanss et al., 2021; Niu et al., 2021; Tang et al., 2015; Wang et al., 2016; Zavodszky et al., 2014)).

A Mutation in *VPS35*, Encoding a Subunit of the Retromer Complex, Causes Late-Onset Parkinson Disease

Alexander Zimprich,[1,14,*] Anna Benet-Pagès,[2,14] Walter Struhal,[3,14] Elisabeth Graf,[2,14] Sebastian H. Eck,[2] Marc N. Offman,[4] Dietrich Haubenberger,[1] Sabine Spielberger,[5] Eva C. Schulte,[2,6] Peter Lichtner,[2] Shaila C. Rossle,[4] Norman Klopp,[7] Elisabeth Wolf,[5] Klaus Seppi,[5] Walter Pirker,[1] Stefan Presslauer,[8] Brit Mollenhauer,[9] Regina Katzenschlager,[10] Thomas Foki,[1] Christoph Hotzy,[1] Eva Reinthaler,[1] Ashot Harutyunyan,[11] Robert Kralovics,[11] Annette Peters,[7] Fritz Zimprich,[1] Thomas Brücke,[8] Werner Poewe,[5] Eduard Auff,[1] Claudia Trenkwalder,[9,12] Burkhard Rost,[4] Gerhard Ransmayr,[3] Juliane Winkelmann,[2,6,13] Thomas Meitinger,[2,13] and Tim M. Strom[2,13,*]

To identify rare causal variants in late-onset Parkinson disease (PD), we investigated an Austrian family with 16 affected individuals by exome sequencing. We found a missense mutation, c.1858G>A (p.Asp620Asn), in the *VPS35* gene in all seven affected family members who are alive. By screening additional PD cases, we saw the same variant cosegregating with the disease in an autosomal-dominant mode with high but incomplete penetrance in two further families with five and ten affected members, respectively. The mean age of onset in the affected individuals was 53 years. Genotyping showed that the shared haplotype extends across 65 kilobases around *VPS35*. Screening the entire *VPS35* coding sequence in an additional 860 cases and 1014 controls revealed six further nonsynonymous missense variants. Three were only present in cases, two were only present in controls, and one was present in cases and controls. The familial mutation p.Asp620Asn and a further variant, c.1570C>T (p.Arg524Trp), detected in a sporadic PD case were predicted to be damaging by sequence-based and molecular-dynamics analyses. VPS35 is a component of the retromer complex and mediates retrograde transport between endosomes and the trans-Golgi network, and it has recently been found to be involved in Alzheimer disease.

Parkinson's disease (PD [MIM 168600]) is the second-most common neurodegenerative disorder; it affects 1%–2% of the population above the age of 60.[1] It is characterized by degeneration of dopaminergic neurons in the nigrostriatal pathway and other monoaminergic cell groups in the brainstem. This degeneration leads to bradykinesia, resting tremor, muscular rigidity, and postural instability as well as nonmotor symptoms. Up to 20% of cases with PD are reported to be familial,[2,3] but extended pedigrees with clear Mendelian inheritance are rare. Genetic studies have so far revealed mutations in five genes causing autosomal-recessive (*PARK2* [MIM 602544], *PINK1* [MIM 608309], *PARK7* [MIM 602533]) or autosomal-dominant (*SNCA* [MIM 163890], *LRRK2* [MIM 609007]) forms of PD.[4–9] Whereas the autosomal-recessive forms with early onset and *SNCA* missense mutations or duplications[10] are rare, a single *LRRK2* mutation (RefSeq number NM_198578.3: c.6055G>A [p.Gly2019Ser]) accounts for approximately 1% of sporadic cases of European origin.[11–13] A recent study revealed a strong association of PD with glucocerebrosidase (*GBA*) mutations in carriers for Gaucher [MIM 230800] disease, thus implicating a lysosomal enzyme in the pathogenesis of PD.[14,15] Genome-wide association studies revealed several low-risk susceptibility loci, among them *LAMP3* [MIM 605883] and *HIP1R* [MIM 605613], which have been reported to be implicated in the lysosomal pathway.[16–18]

We identified an Austrian family in which 16 members were affected by PD (family A, Figure 1). PD seemed to be inherited in an autosomal-dominant mode with high penetrance. Seven affected members were available for clinical and DNA investigations. Six of them exhibited at least three of the four cardinal signs of PD (akinesia, resting tremor, rigidity, and postural instability) and showed improvement after dopaminergic treatment. A single affected individual had displayed action tremors since childhood but developed L-Dopa-responsive resting tremors and akinesia only at the age of 62 years. The mean age of onset was 53 years (range 40–68 years) (Table 1). The clinical diagnosis of idiopathic PD was made by movement-disorder specialists who used UK brain bank criteria for PD.[19] All participants gave written informed

[1]Department of Neurology, Medizinische Universität Wien, 1090 Vienna, Austria; [2]Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany; [3]Department of Neurology and Psychiatry, Allgemeines Krankenhaus, 4021 Linz, Austria; [4]Institute of Bioinformatics, Technische Universität München, 85748 Garching, Germany; [5]Department of Neurology, Medizinische Universität Innsbruck, 6020 Innsbruck, Austria; [6]Department of Neurology, Technische Universität München, 81675 Munich, Germany; [7]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany; [8]Department of Neurology, Wilhelminenspital, 1160 Vienna, Austria; [9]Paracelsus-Elena Klinik, 34128 Kassel, Germany; [10]Department of Neurology, Sozialmedizinisches Zentrum Ost-Donauspital, 1220Vienna, Austria; [11]Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria; [12]Department of Clinical Neurophysiology, Georg-August-Universität Göttingen, 37079 Göttingen, Germany; [13]Institute of Human Genetics, Technische Universität München, 81675 Munich, Germany
[14]These authors contributed equally to this work
*Correspondence: alexander.zimprich@meduniwien.ac.at (A.Z.), timstrom@helmholtz-muenchen.de (T.M.S.)
DOI 10.1016/j.ajhg.2011.06.008.

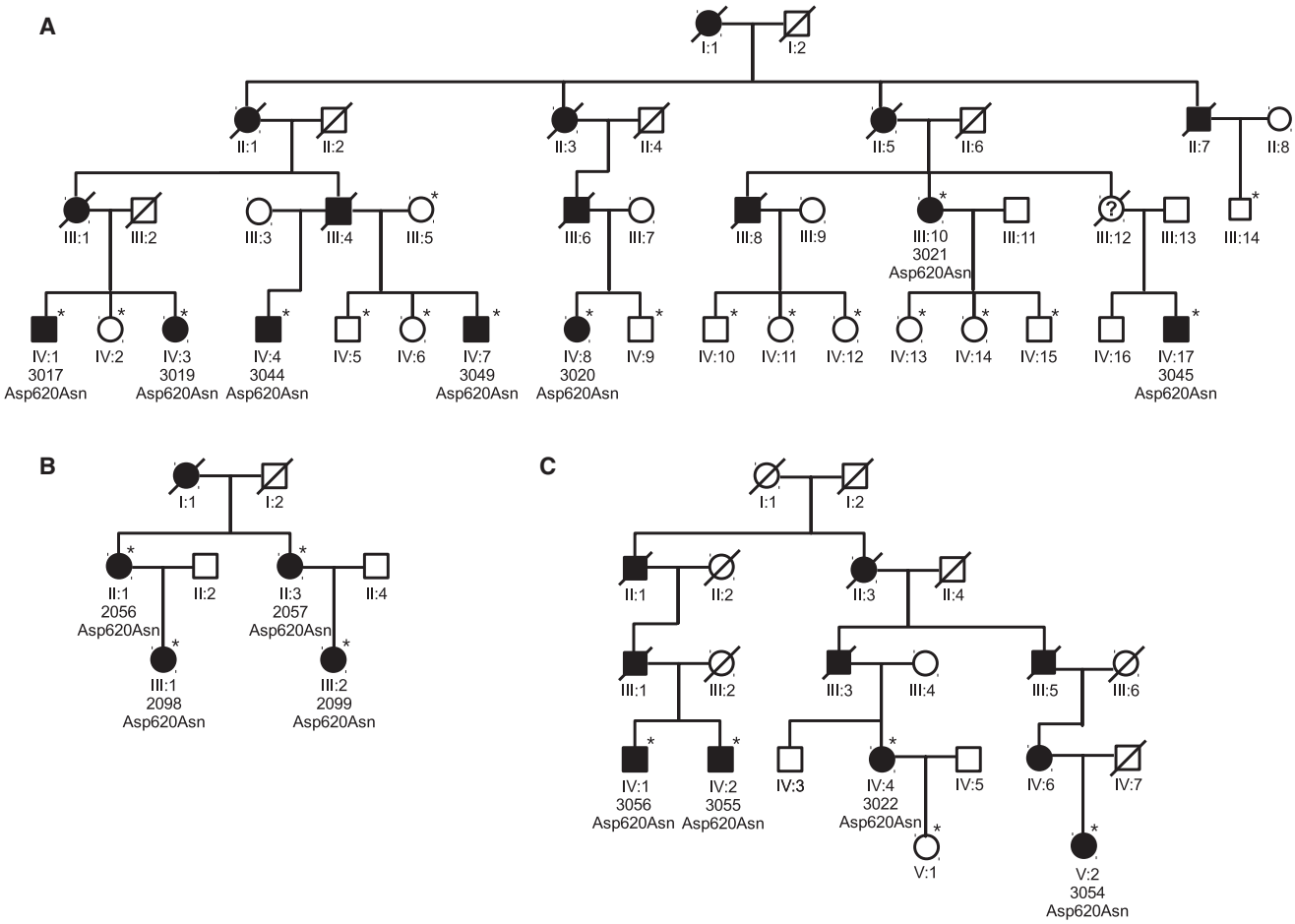


Figure 1. Pedigrees of Families A, B, and C
Unaffected family members are indicated by open symbols, affected members by closed symbols. Asterisks denote individuals genotyped for p.Asp620Asn. To maintain confidentiality, we have not shown genotypes of unaffected individuals. A question mark within a symbol denotes an unknown phenotype. Diagonal bars through symbols denote deceased individuals.

consent. The study was approved by the institutional review board of the Medizinische Universität Wien and the Hessische Landesärztekammer Wiesbaden.

To identify the disease-causing variant, we selected two second cousins (#3017 and #3020) for exome sequencing. We assumed that any rare variants common in both individuals would be disease-causing candidates. Selecting distantly related members of the pedigree should minimize the proportion of alleles shared by descent. Exome sequencing was performed on a Genome Analyzer IIx system (Illumina) after in-solution enrichment of exonic sequences (SureSelect Human All Exon 38 Mb kit, Agilent). We sequenced two lanes of a flowcell for both samples, each as 54 bp paired-end runs. Read alignment was performed with BWA (version 0.5.8) to the human genome assembly hg19 (Table S1, available online). Single-nucleotide variants and small insertions and deletions (indels) were detected with SAMtools (v 0.1.7). We filtered called variants to exclude those present in 72 control exomes from patients with other unrelated diseases. We further excluded all variants that were present in dbSNP 131 and had an average heterozygosity of more than 0.02. Variant annotation was performed with custom scripts. This approach left ten heterozygous nonsynonymous variants shared by both affected individuals (Table 2; see also Table S2).

Only a single heterozygous variant in the *VPS35* gene (RefSeq number NM_018206.4: c.1858G>A [p.Asp620Asn]) fulfilled two further criteria of being possibly causative: (1) it was found in all seven affected members investigated and (2) was absent in approximately 680 KORA S4 general-population samples (Tables 2 and 3).[20] We next screened 486 unrelated PD patients from Austria for the p.Asp620Asn variant by MALDI-TOF mass spectroscopy (Sequenom MassArray system). We detected two additional index patients carrying this mutation (families B and C; Figure 1 and Table 1). The variant was detected in all eight affected individuals investigated in both families. It was not present in a second set of 554 Austrian controls or in an additional 1014 KORA-AGE controls (Table 3). The variant was further detected in three clinically unaffected family members in families A, B, and C. Because the unaffected individuals are all younger than 60 years of age, either they are all presymptomatic or the mutation is nonpenetrant in these subjects.

Table 1. Clinical Findings for PD Patients Carrying Variants in VPS35

Family	Patient	Variation	AaO	DD	IS	B	R	RT	PI	L-Dopa/DA	Other Features
A	3017	p.Asp620Asn	48	7	B	+	+	−	+	+	
A	3019	p.Asp620Asn	40	5	B	+	+	+	+	+	
A	3020	p.Asp620Asn	46	7	PI	+	+	−	+	+	
A	3021	p.Asp620Asn	68	16	PI	+	+	+	+	+	
A	3049	p.Asp620Asn	49	4	RT	+	+	+	−	+	
A	3044	p.Asp620Asn	64	3	PI	+	+	+	+	+	
A	3045	p.Asp620Asn	63	1	RT	+	−	+	−	+	action tremor since childhood
B	2056	p.Asp620Asn	61	15	RT	+	+	+	+	+	fluctuations, dyskinesias
B	2057	p.Asp620Asn	56	8	RT	+	+	+	+	+	fluctuations, dyskinesias
B	2098	p.Asp620Asn	46	0.5	RT	−	−	+	−	untreated	depression, action tremor, pathologic DAT SPECT
B	2099	p.Asp620Asn	51	5	B	+	+	+	−	+	fluctuations, pathologic DAT SPECT
C	3022	p.Asp620Asn	61	5	RT	+	+	+	−	+	dyskinesias
C	3055	p.Asp620Asn	46	12	RT	+	+	+	−	+	
C	3054	p.Asp620Asn	53	9	B	+	+	−	−	+	dyskinesias
C	3056	p.Asp620Asn	43	10	B	+	+	+	+	+	dyskinesias
	211	p.Arg524Trp	37	9	MG	+	+	+	−	+	mild action tremor since youth; 75% motor improvement on levodopa-test; DBS for fluctuations and dyskinesias; pathologic DAT SPECT
	524	p.Leu774Met	51	7	RT	+	+	+	−	+	marked postural tremor
	243	p.Leu774Met	73	9	RT	+	+	+	+	+	dyskinesias, pathologic DAT SPECT
	806	p.Ile241Met	72	2	Postural tremor	+	−	+	+	+	hyposmia (6/12 sniffing sticks), DAT SPECT pathologic, pathologic crying
	90/05	p.Met57Ile	62	13	RT	+	+	+	+	+	dementia (MMSE 23), dysphagia and dysarthria, hyposmia by history, depression

Abbreviations are as follows: AaO, age at onset; DD, disease duration in years; IS, initial symptoms; B, bradykiesia; R, rigidity; RT, resting tremor; PI, postural instability; L-Dopa/DA, response to L-Dopa and/or dopamine agonist; MG, micrographia; DBS, deep brain stimulation.

Cross-species alignment of VPS35 from plants, fungi, invertebrates, and vertebrates showed complete conservation of amino acid Asp620 (Figure S1). The likely consequence of the p.Asp620Asn variant was predicted to be damaging by PolyPhen2,[21] SNAP,[22] and SIFT.[23] We therefore concluded that the variant p.Asp620Asn is indeed very likely to be causative for PD in families A, B, and C.

To determine whether the variant p.Asp620Asn occurred on the same haplotype, we genotyped 20 individuals from families A–C with oligonucleotide SNP arrays (HumanOmni2.5-Quad, Illumina). Haplotyping and linkage analysis were performed with the Merlin software.[24] The haplotypes carrying the variant p.Asp620Asn in families A–C are depicted in Table S3. Family A and B shared a common haplotype across 21 Mb between markers rs1072594 and rs4444336. Family C, however, showed only a common region of 65 kb across *VPS35*. Different alleles were located at markers rs56168099 and rs74459547, 25 kb upstream and 11 kb downstream of *VPS35*, respectively (Table S3). Because the two intragenic markers did not differ, we could not determine whether the three families shared an old common haplotype or whether the mutation has recently arisen on two different haplotypes.

To assess the prevalence of other *VPS35* mutations among PD cases and the general population, we screened all 17 coding exons for variations by dye-binding/high-resolution DNA melting curve analysis (LightScanner HR I 384, Idaho Technology) in 860 cases (484 Austrian and

Table 2. Exome Sequencing: Rare, Heterozygous, Nonsynonymous Variations Shared by Two Individuals of Pedigree A

				Variations		Control Genotypes			
Gene	**Position (hg19)**	**dbSNP**	**Transcript**	**Nucleotide**	**Amino Acid**	**1/1**	**1/2**	**2/2**	**Segregation**
PLK3	chr1:45270359		NM_004073.2	c.1543T>A	p.Ser515Thr	669	0	0	4 of 7
C8A	chr1:57383357	rs41285938	NM_000562.2	c.1723C>T	p.Pro575Ser				5 of 7
ADCY10	chr1:167787479	rs41270737	NM_018417.4	c.4313A>G	p.Asn1438Ser				2 of 7
LAMB2	chr3:49166460		NM_002292.3	c.1724G>A	p.Arg575Gln	647	28	0	5 of 7
NOM1	chr7:156762317		NM_138400.1	c.2503G>A	p.Ala835Thr	670	0	0	3 of 7
KIF22	chr16:29816237		NM_007317.1	c.1780G>A	p.Asp594Asn	665	6	0	6 of 7
SEZ6L2	chr16:29899021		NM_012410.2	c.947G>A	p.Arg316His	660	4	0	7 of 7
VPS35	chr16:46696364		NM_018206.4	c. 1858G>A	p.Asp620Asn	1069[a]	0	0	7 of 7
NLRP1	chr17:5421150		NM_001033053.2	c.3985G>A	p.Val1329Ile	666	4	0	3 of 7
NEURL4	chr17:7221197		NM_001005408.1	c.4109G>A	p.Arg1370Gln				3 of 7

Rare variations revealed by exome sequencing were checked in 670 controls (KORA S4) by MALDI-TOF analysis. The variant allele was denoted as "2," the reference allele as "1."
[a] This number includes additional 554 Austrian control individuals investigated by a TaqMan genotyping assay. Segregation shows the number of affected pedigree A individuals who carry the variant allele.

376 German cases) and 1014 controls. For controls, we used a population-based cohort (KORA AGE) with a mean age of 76 years but excluded eight individuals known to be on medications for PD (Table 3). Exons 2 to 12 are located within a region that is duplicated 12 Mb upstream. Primers were designed to specifically amplify these exons (Table S4). The screening revealed six further rare coding SNVs in addition to p.Asp620Asn (Table 4). Including p.Asp620Asn, we identified four different nonsynonymous missense variants only present in cases, two only present in controls, and one present in cases and controls. Two of the variants unique to PD cases were predicted to be damaging by all three methods (c.1858G>A [p.Asp620Asn]; c.1570C>T [p.Arg524Trp]), and one was predicted by PolyPhen2 to be possibly damaging (c.723T>G, p.Ile241Met). The other variants were predicted to be benign by all methods. Family information was only available for the patient carrying the p.Arg524Trp variant. The only available family member was her mother, aged 74 years. She was found to also carry the variant and showed mild extrapyramidal signs, including intermittent resting tremor of the left fingers and mild postural tremor of both upper limbs, but no bradykinesia. However, a DAT SPECT examination showed normal striatal binding, excluding the possibility of an early stage of PD in this subject. Of note, the screening did not reveal any common nonsynonymous coding SNVs. Furthermore, common nonsynonymous coding SNVs were not found in the 72 control exomes from patients with other unrelated diseases, nor were any recorded in the dbSNP database (version 131).

Table 3. Summary of the Samples Used in This Study

Cohort	**Sample Size**	**Mean Age (SD)**	**Females/Males**
Austrian PD cases[a]	486	58.7 (11.3)	172/314
German PD cases[b]	376	71.1 (9.4)	119/257
KORA S4 controls[c]	680	54.7 (11.9)	280/400
KORA-AGE controls[d]	1014	76.0 (6.6)	508/505
Austrian controls[e]	554	46 (15.2)	254/300

Patients presenting with atypical or secondary (e.g., vascular) parkinsonian disorders as well as patients with known mutations were excluded.
[a] The Austrian cases were recruited at the Department of Neurology, Medizinische Universität Wien, Vienna, as well as in affiliated departments on a consecutive basis. A positive family history for PD was reported from 131 patients. A positive family history was defined by at least one other affected first- or second-degree related family member.
[b] The German PD population originated from the Paracelsus-Elena Klinik, Kassel, a hospital specializing in movement disorders.
[c] This control population was recruited from the KORA S4 survey, comprising individuals who were aged 25–74 years and were examined during 1999–2001.
[d] The KORA-AGE samples were collected in 2009 as a gender- and age-stratified subsample of the KORA S1–S4 studies comprising participants born before 1944. KORA S1–S4 surveys comprise four independent cross-sectional population-based studies in the region of Augsburg, Southern Germany, and were conducted in 5 year intervals. Patients for whom PD was suspected on the basis of questionnaire data were excluded.
[e] These control samples were recruited through the Department of Neurology, Medical University of Vienna, as subjects without known history of a neurological disorder and included, for example, blood donors or unrelated companions or spouses of patients.

VPS35 is a component of the retromer complex and is involved in retrograde transport from the endosomes back to the trans-Golgi network.[25] This multi-protein complex consists of the cargo-recognition VPS26-VPS29-VPS35 heterotrimer and a membrane-targeting heterodimer or homodimer of SNX1 and/or SNX2 (vps5).[25,26] All proteins involved are evolutionarily conserved and have been previously described in *Saccharomyces cerevisiae*. The best characterized cargo proteins of the retromer complex are the cation-independent mannose 6-phosphate receptor

Table 4. Rare *VPS35* Variants in Cases and Controls

ID Cases	KORA AGE Controls	Heterozygous Nucleotide Change	Amino Acid Change	Predicted Impact on Protein			Exon/ Intron	Genomic Position (hg19, chr16)	KORA S4 Controls		
				(i)	(ii)	(iii)			1/1	1/2	2/2
Nonsynonymous											
-	1	c.151G>A	p.Gly51Ser	+	+	+	3	46,716,,039			
90/05	-	c.171G>A	p.Met57Ile	+	+	+	3	46,716,019	670	0	0
-	1	c.245C>G	p.Thr82Arg	+	+	+	4	46,715,367			
806	-	c.723T>G	p.Ile241Met	±	+	+	7	46,711,308	667	0	0
[211]	-	c.1570C>T	p.Arg524Trp	−	−	−	13	46,702,919	671	0	0
[Families A-C]	-	c.1858G>A	p.Asp620Asn	−	−	−	15	46,696,364	669	0	0
243, 524	2	c.2320C>A	p.Leu774Met	+	+	+	17	46,694,455			
Synonymous											
53097	-	c.492A>G	p.Glu164Glu				5	46,714,597	671	0	0
-	1	c.954A>T	p.Gly315Gly				9	46,708,542			
53496	-	c.1881C>T	p.Ala627Ala				15	46,696,341	668	5	0
45, 117, 53626	1	c.2145A>G	p.Leu715Leu				16	46,695,696	666	2	0
53667	-	c.2241C>T	p.Ile747Ile				17	46,694,534	667	2	0
53063	-	c.2346A>G	p.Glu782Glu				17	46,694,429	671	0	0
-	1	c.2361G>A	p.Glu787Glu				17	46,694,414			
Noncoding											
2212	2	c.1-35C>T					5′UTR	46,723,080	667	2	0
-	2	c.1-29C>T					5′UTR	46,723,074			
95, 2206	3	c.3+24A>G					1	46,723,019	662	6	0
159, 528	1	c.102+33G>A					2	46,717,387	668	2	0
[157, 2023]	-	c.103-77T>C					3	46,716,164	668	0	0
-	1	c.199+9T>G					3	46,715,982			
213	-	c.506+6T>C					5	46,714,577	644	0	0
53093	-	c.720+18C>T					6	46,712,773			
-	1	c.914+38T>C					8	46,710,457			
52824	-	c.1161-87A>C					10	46,706,471			
52791	-	c.1161-70G>A					10	46,706,454	668	0	0
-	1	c.1368+16C>T					11	46,706,161			
[2028]	-	c.1369-11G>A					12	46,705,783	669	0	0
-	1	c.1525-17delT					12	46,702,985			
-	1	c.1647+14T>C					13	46,702,828			
320	-	c.2212-45T>C					16	46,694,608	670	0	0
[352]	-	c.2391+7A>G					3′UTR	46,694,377			
-	1	c.2391+8A>G					3′UTR	46,694,376			

Variants for 863 cases and 1014 KORA AGE controls were determined by dye-binding/high-resolution DNA melting curve analysis and confirmed by Sanger sequencing. The table lists the case ID and the number of detected variant alleles of the cases and KORA AGE samples, respectively. Genotypes of identified variants were further investigated by MALDI-TOF analysis in approximately 680 KORA S4 controls. For the KORA S4 samples, the variant allele was denoted as "2," the reference allele as "1." cDNA numbering is based on reference gene NM_018206.4 for *VPS35*, where +1 corresponds to the A of ATG start translation codon. Familial cases are given in square brackets. Three methods were used for predicting the impact of SNPs on the protein. (1) PolyPhen2, (2) SNAP, and (3) SIFT; "+" indicates a benign impact, "±" indicates a possibly damaging impact, and "−" indicates a damaging impact. We detected a further nonsynonymous variant (c.1093C>T [p.Arg365Cys], genomic position 46,708,293) in a patient carrying two *PARKIN* variants (c.exon3_4del and p.Arg275Trp). This variant was not present in 670 KORA S4 and 1014 KORA AGE controls. It is predicted to be possibly damaging by all three methods. This patient's brother is also affected by PD. He carries the 2 PARKIN variants but not the *VPS35* variant.

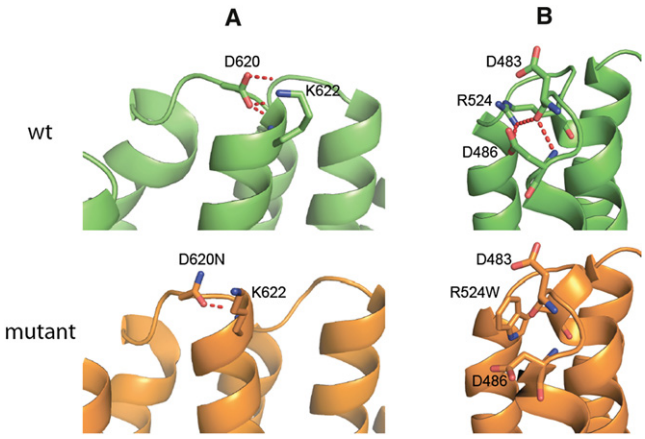


Figure 2. Hydrogen-Bonding Capacities for Wild-Type Asp620 and Arg524 and the Variants p.Asp620Asn and p.Arg524Trp
Hydrogen bonds (HB) are shown as red dashed lines. Asp60 and Arg524 are in green; p.Asp620Asn and p.Arg524Trp are in orange. (A) Asp620 forms a HB to Lys622 and shows an additional salt-bridge interaction. p.Asp620Asn forms fewer HBs, and no electrostatic interaction is possible.
(B) Arg524 forms a HB network with Asp483 and Asp486. This network is broken by the p.Arg524Trp substitution.

(CI-MPR)[27] and Vps10p in mammals and *Saccharomyces cerevisiae*, respectively; these proteins transport hydroxylases to the lysosomes or lysosomal vacuoles. Recently, additional cargo proteins and functions of VPS35 have been described.[28,29] Most interesting in our context is the involvement of the retromer into the retrograde transport of SORL1, a VPS10P-domain receptor protein that has been implicated in Alzheimer disease.[30,31] The crystal structure of the C-terminal part of VPS35 has been resolved.[32] The three variants p.Asp620Asn, p.Arg524Trp, and p.Leu774Met are located in this part of the protein, and we have investigated their impact on protein stability by using molecular dynamics (MD) simulations. We manually introduced the mutations to the crystal structure and modeled the side chains by using scwrl 4.0.[33] All MD simulations were performed via GROMACS 4.5,[34] with the all-atom force field AMBER03[35] and the water model TIP3P[36] as parameters. All three proteins are found on the edge of helices interacting with VPS29. Wild-type residue Asp620 forms frequent hydrogen bonds (HBs) to Lys622, but these bonds are less frequent in the p.Asp620Asn variant (Figure 2A). Similarly, Arg524 is involved in a triple HB network together with residues Asp483 and Asp486, but this network is broken by the introduction of p.Arg524Trp (Figure 2B). Both changes result in the loss of salt bridges and cause the protein to be locally more flexible, as shown by root-mean-square fluctuation (RMSF) profiles (Figure S2). In contrast to the effect predicted for p.Arg524Trp and p.Asp620Asn, the p.Leu774Met variant was not predicted to have a strong impact on protein stability.

In summary, we identified rare *VPS35* missense variants that are potentially pathogenic. One of these variants, p.Asp620Asn, cosegregates with late-onset PD in three unrelated families. The observation that the three families share only a small common haplotype across *VPS35*, the high conservation of VPS35, the predicted structural changes, and the protein's known involvement in lysosomal trafficking together provide strong support for the p.Asp620Asn variant's being causative for late-onset PD, although we identified only a single familial mutation. The penetrance of p.Asp620Asn is high but not complete and might be lower for the other variants. The proportion of PD caused by *VPS35* variants is expected to be low. Although exome sequencing provides perfect access to rare-variant detection, both large families and large collections of cases and controls remain a crucial resource for the identification of disease genes.

Supplemental Data

Supplemental Data include two figures and four tables and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments

We thank all patients and their families for participating in this study. We also thank C. Fischer and B. Schmick for technical assistance and S. Schmidegg, S. Hoedl, and M. Guger for clinical examination of family A. This work was supported by a grant from the German Ministry for Education and Research (01GR0804-4). The KORA study was financed by the Helmholtz Zentrum München, the German Federal Ministry of Education and Research, the State of Bavaria, the German National Genome Research Network (NGFNplus: 01GS0823), and the Munich Center of Health Sciences (MCHealth) as part of LMUinnovativ. M.N.O., S.C.R., and B.R. were supported by the Alexander von Humboldt Foundation.

Received: May 8, 2011
Revised: June 15, 2011
Accepted: June 21, 2011
Published online: July 14, 2011

Web Resources

The URLs for data presented herein are as follows:

ExonPrimer, http://ihg.helmholtz-muenchen.de/exonprimer.html
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org
UCSC Genome Browser, http://genome.ucsc.edu

References

1. Lang, A.E., and Lozano, A.M. (1998). Parkinson's disease. First of two parts. N. Engl. J. Med. *339*, 1044–1053.
2. Bonifati, V., Fabrizio, E., Vanacore, N., De Mari, M., and Meco, G. (1995). Familial Parkinson's disease: A clinical genetic analysis. Can. J. Neurol. Sci. *22*, 272–279.
3. Payami, H., Larsen, K., Bernard, S., and Nutt, J. (1994). Increased risk of Parkinson's disease in parents and siblings of patients. Ann. Neurol. *36*, 659–661.
4. Zimprich, A., Biskup, S., Leitner, P., Lichtner, P., Farrer, M., Lincoln, S., Kachergus, J., Hulihan, M., Uitti, R.J., Calne, D.B.,

et al. (2004). Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron *44*, 601–607.
5. Bonifati, V., Rizzu, P., van Baren, M.J., Schaap, O., Breedveld, G.J., Krieger, E., Dekker, M.C., Squitieri, F., Ibanez, P., Joosse, M., et al. (2003). Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science *299*, 256–259.
6. Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y., and Shimizu, N. (1998). Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature *392*, 605–608.
7. Paisán-Ruíz, C., Jain, S., Evans, E.W., Gilks, W.P., Simón, J., van der Brug, M., López de Munain, A., Aparicio, S., Gil, A.M., Khan, N., et al. (2004). Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron *44*, 595–600.
8. Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., et al. (1997). Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science *276*, 2045–2047.
9. Valente, E.M., Abou-Sleiman, P.M., Caputo, V., Muqit, M.M., Harvey, K., Gispert, S., Ali, Z., Del Turco, D., Bentivoglio, A.R., Healy, D.G., et al. (2004). Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science *304*, 1158–1160.
10. Johnson, J., Hague, S.M., Hanson, M., Gibson, A., Wilson, K.E., Evans, E.W., Singleton, A.A., McInerney-Leo, A., Nussbaum, R.L., Hernandez, D.G., et al. (2004). SNCA multiplication is not a common cause of Parkinson disease or dementia with Lewy bodies. Neurology *63*, 554–556.
11. Gilks, W.P., Abou-Sleiman, P.M., Gandhi, S., Jain, S., Singleton, A., Lees, A.J., Shaw, K., Bhatia, K.P., Bonifati, V., Quinn, N.P., et al. (2005). A common LRRK2 mutation in idiopathic Parkinson's disease. Lancet *365*, 415–416.
12. Nichols, W.C., Pankratz, N., Hernandez, D., Paisán-Ruíz, C., Jain, S., Halter, C.A., Michaels, V.E., Reed, T., Rudolph, A., Shults, C.W., et al; Parkinson Study Group-PROGENI investigators. (2005). Genetic screening for a single common LRRK2 mutation in familial Parkinson's disease. Lancet *365*, 410–412.
13. Di Fonzo, A., Rohé, C.F., Ferreira, J., Chien, H.F., Vacca, L., Stocchi, F., Guedes, L., Fabrizio, E., Manfredi, M., Vanacore, N., et al; Italian Parkinson Genetics Network. (2005). A frequent LRRK2 gene mutation associated with autosomal dominant Parkinson's disease. Lancet *365*, 412–415.
14. Sidransky, E., Nalls, M.A., Aasly, J.O., Aharon-Peretz, J., Annesi, G., Barbosa, E.R., Bar-Shira, A., Berg, D., Bras, J., Brice, A., et al. (2009). Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N. Engl. J. Med. *361*, 1651–1661.
15. Aharon-Peretz, J., Rosenbaum, H., and Gershoni-Baruch, R. (2004). Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. N. Engl. J. Med. *351*, 1972–1977.
16. Simón-Sánchez, J., Schulte, C., Bras, J.M., Sharma, M., Gibbs, J.R., Berg, D., Paisan-Ruiz, C., Lichtner, P., Scholz, S.W., Hernandez, D.G., et al. (2009). Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat. Genet. *41*, 1308–1312.
17. Satake, W., Nakabayashi, Y., Mizuta, I., Hirota, Y., Ito, C., Kubo, M., Kawaguchi, T., Tsunoda, T., Watanabe, M., Takeda, A., et al. (2009). Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nat. Genet. *41*, 1303–1307.
18. Nalls, M.A., Plagnol, V., Hernandez, D.G., Sharma, M., Sheerin, U.M., Saad, M., Simón-Sánchez, J., Schulte, C., Lesage, S., Sveinbjörnsdóttir, S., et al; International Parkinson Disease Genomics Consortium. (2011). Imputation of sequence variants for identification of genetic risks for Parkinson's disease: A meta-analysis of genome-wide association studies. Lancet *377*, 641–649.
19. Hughes, A.J., Daniel, S.E., Kilford, L., and Lees, A.J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: A clinico-pathological study of 100 cases. J. Neurol. Neurosurg. Psychiatry *55*, 181–184.
20. Wichmann, H.E., Gieger, C., and Illig, T. (2005). KORA-gen–resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen *67* (*Suppl. 1*), 26–30.
21. Ramensky, V., Bork, P., and Sunyaev, S. (2002). Human non-synonymous SNPs: server and survey. Nucleic Acids Res. *30*, 3894–3900.
22. Bromberg, Y., and Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res. *35*, 3823–3835.
23. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.
24. Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. (2002). Merlin—Rapid analysis of dense genetic maps using sparse gene flow trees. Nat. Genet. *30*, 97–101.
25. Bonifacino, J.S., and Rojas, R. (2006). Retrograde transport from endosomes to the trans-Golgi network. Nat. Rev. Mol. Cell Biol. *7*, 568–579.
26. Rojas, R., Kametaka, S., Haft, C.R., and Bonifacino, J.S. (2007). Interchangeable but essential functions of SNX1 and SNX2 in the association of retromer with endosomes and the trafficking of mannose 6-phosphate receptors. Mol. Cell. Biol. *27*, 1112–1124.
27. Damen, E., Krieger, E., Nielsen, J.E., Eygensteyn, J., and van Leeuwen, J.E. (2006). The human Vps29 retromer component is a metallo-phosphoesterase for a cation-independent mannose 6-phosphate receptor substrate peptide. Biochem. J. *398*, 399–409.
28. Braschi, E., Goyon, V., Zunino, R., Mohanty, A., Xu, L., and McBride, H.M. (2010). Vps35 mediates vesicle transport between the mitochondria and peroxisomes. Curr. Biol. *20*, 1310–1315.
29. Korolchuk, V.I., Schütz, M.M., Gómez-Llorente, C., Rocha, J., Lansu, N.R., Collins, S.M., Wairkar, Y.P., Robinson, I.M., and O'Kane, C.J. (2007). Drosophila Vps35 function is necessary for normal endocytic trafficking and actin cytoskeleton organisation. J. Cell Sci. *120*, 4367–4376.
30. Willnow, T.E., Petersen, C.M., and Nykjaer, A. (2008). VPS10P-domain receptors—Regulators of neuronal viability and function. Nat. Rev. Neurosci. *9*, 899–909.
31. Rogaeva, E., Meng, Y., Lee, J.H., Gu, Y., Kawarai, T., Zou, F., Katayama, T., Baldwin, C.T., Cheng, R., Hasegawa, H., et al. (2007). The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nat. Genet. *39*, 168–177.
32. Hierro, A., Rojas, A.L., Rojas, R., Murthy, N., Effantin, G., Kajava, A.V., Steven, A.C., Bonifacino, J.S., and Hurley, J.H.

(2007). Functional architecture of the retromer cargo-recognition complex. Nature *449*, 1063–1067.

33. Krivov, G.G., Shapovalov, M.V., and Dunbrack, R.L., Jr. (2009). Improved prediction of protein side-chain conformations with SCWRL4. Proteins *77*, 778–795.
34. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory Comput. *4*, 435–447.
35. Duan, Y., Wu, C., Chowdhury, S., Lee, M.C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., et al. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J. Comput. Chem. *24*, 1999–2012.
36. Mahoney, M.W. (2000). A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. J. Chem. Phys. *112*, 8910–8922.

Neurogenetics (2014) 15:49–57
DOI 10.1007/s10048-013-0383-8

ORIGINAL ARTICLE

Rare variants in *LRRK1* and Parkinson's disease

Eva C. Schulte · Daniel C. Ellwanger · Sybille Dihanich · Claudia Manzoni · Katrin Stangl · Barbara Schormair · Elisabeth Graf · Sebastian Eck · Brit Mollenhauer · Dietrich Haubenberger · Walter Pirker · Alexander Zimprich · Thomas Brücke · Peter Lichtner · Annette Peters · Christian Gieger · Claudia Trenkwalder · Hans-Werner Mewes · Thomas Meitinger · Patrick A. Lewis · Hans H. Klünemann · Juliane Winkelmann

Received: 28 July 2013 / Accepted: 15 October 2013 / Published online: 16 November 2013

Abstract Approximately 20 % of individuals with Parkinson's disease (PD) report a positive family history. Yet, a large portion of causal and disease-modifying variants is still unknown. We used exome sequencing in two affected individuals from a family with late-onset PD to identify 15 potentially causal variants. Segregation analysis and frequency assessment in 862 PD cases and 1,014 ethnically matched controls highlighted variants in *EEF1D* and *LRRK1* as the best candidates. Mutation screening of the coding regions of these genes in 862 cases and 1,014 controls revealed several novel

Electronic supplementary material The online version of this article (doi:10.1007/s10048-013-0383-8) contains supplementary material, which is available to authorized users.

E. C. Schulte · J. Winkelmann
Klinik und Poliklinik für Neurologie, Klinikum rechts der Isar, Technische Universität München, Ismaningerstr. 22, 81675 Munich, Germany

E. C. Schulte · B. Schormair · E. Graf · S. Eck · P. Lichtner · T. Meitinger · J. Winkelmann
Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany

D. C. Ellwanger · H.-W. Mewes
Department of Genome-Oriented Bioinformatics, Technische Universität München, Freising, Germany

S. Dihanich · C. Manzoni · P. A. Lewis
Department of Molecular Neuroscience, UCL Institute of Neurology, University College London, London, UK

K. Stangl · H. H. Klünemann
Klinik und Poliklinik für Psychiatrie, Psychosomatik und Psychotherapie, Universität Regensburg, Regensburg, Germany

B. Schormair · P. Lichtner · T. Meitinger · J. Winkelmann
Institut für Humangenetik, Technische Universität München, Munich, Germany

B. Mollenhauer · C. Trenkwalder
Paracelsus Elena Klinik, Kassel, Germany

B. Mollenhauer
Neurochirurgische Klinik, Georg August Universität Göttingen, Göttingen, Germany

D. Haubenberger · W. Pirker · A. Zimprich
Department of Neurology, Medical University of Vienna, Vienna, Austria

T. Brücke
Neurologische Klinik, Wilhelminenspital, Vienna, Austria

A. Peters
Institut für Epidemiologie II, Helmholtz Zentrum München, Munich, Germany

C. Gieger
Institut für Genetische Epidemiologie, Helmholtz Zentrum München, Munich, Germany

C. Trenkwalder
Neurologische Klinik, Georg August Universität Göttingen, Göttingen, Germany

T. Meitinger · J. Winkelmann
Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

P. A. Lewis
School of Pharmacy, University of Reading, Whiteknights, Reading, UK

J. Winkelmann (✉)
Department of Neurology and Neurosciences, Stanford University, Palo Alto, USA
e-mail: winkelmann@lrz.tum.de

non-synonymous variants in both genes in cases and controls. An in silico multi-model bioinformatics analysis was used to prioritize identified variants in *LRRK1* for functional follow-up. However, protein expression, subcellular localization, and cell viability were not affected by the identified variants. Although it has yet to be proven conclusively that variants in *LRRK1* are indeed causative of PD, our data strengthen a possible role for *LRRK1* in addition to *LRRK2* in the genetic underpinnings of PD but, at the same time, highlight the difficulties encountered in the study of rare variants identified by next-generation sequencing in diseases with autosomal dominant or complex patterns of inheritance.

Keywords Parkinson's disease · *LRRK1* · *EEF1D* · Exome sequencing

Introduction

Characterized by resting tremor, bradykinesia, rigidity, and postural instability, Parkinson's disease (PD) is a prominent neurodegenerative disorder. Genetic factors contribute to the risk of PD—both sporadic and familial. Although up to 20 % of PD cases are believed to be familial [1, 2], thus far, rare genetic variants in only a few genes have been unequivocally shown to underlie these familial forms. They include *PARK2*/*PARKIN*, *PINK1*, *PARK7*/*DJ-1*, *SNCA*, and *LRRK2* [3–8]. While all of these were identified by classical linkage analysis in large, multi-generation families, recently, next-generation sequencing has enabled the identification of disease-causing variants in smaller families and—what is especially important with regard to the investigation of neurodegenerative conditions with an onset late in life—without the need of genotypic information from more than one generation of affected individuals. Recently, exome sequencing was used to identify *VPS35* as an additional gene involved in late-onset familial PD [9, 10]. Still, to date, the identified genes only explain a small portion of the genetic burden in familial PD. It is likely that genetic factors involved in bringing about the PD phenotype comprise both genetic variants of strong effect as well as variants of weaker effect which contribute to disease risk or phenotypic modification. A thorough understanding of the entire spectrum of genetic alterations implicated in the disease is necessary to better understand disease pathogenesis and to provide more specific treatment options in the future.

Here, we describe whole exome sequencing in a German family with autosomal dominant late-onset PD in whom known PD-linked mutations has previously been excluded in an attempt to pinpoint the disease-causing genetic variant. Two variants in leucine-rich repeat kinase 1 (*LRRK1*) and eukaryotic translation elongation factor 1 delta (*EEF1D*) emerged as the best candidate variants.

Materials and methods

Participants

The family was evaluated by neurologists specializing in movement disorders. All family members received a detailed neurologic exam. Information on deceased family members was gathered from medical and family records. Cases and controls used in genotyping and variant screening have been reported previously [9, 11] and are described in more detail in the supplement. Ethics review board approval and participants' written informed consent were obtained prior to the initiation of the study.

Analysis of copy number variation

Genome-wide copy number variant (CNV) analysis was carried out using Affymetrix Whole-Genome 2.7 M Array in conjunction with the Chromosome Analysis Suite with a confidence index of 85, a minimum homozygous region size of 10 kb and a minimum probe count of 5.

Exome sequencing

Exome sequencing was performed on a Genome Analyzer IIx (Illumina) after in-solution enrichment of exonic sequences (SureSelect Human All Exon 38 Mb kit, Agilent). For both samples, two lanes of a flow cell were sequenced, each as 54-bp paired-end runs. Read alignment was carried out with BWA (version 0.5.8) to the human genome assembly hg19. Single nucleotide variants (SNVs) and small insertions and deletions (indels) were detected with SAMtools (version 0.1.7). Prior to exome sequencing, presumably causal mutations in known Parkinson's disease genes (*SNCA*, *PARK2*, *DJ1*, *PINK1*, and *LRRK2* (p.G2019S only)) had been excluded. Moreover, no known PD-linked variants were identified in either V:8 or V:17 by exome sequencing.

Genotyping

All 15 candidate variants were genotyped in 862 cases (376 of German (age 71.1±9.4 years, 31.6 % female) and 486 of Austrian (age 58.7±11.3 years, 35.4 % female) origin) and 1,014 population-based controls pertaining to the KORA-AGE cohort (age 76±6.6 years, 50.1 % female) using MALDI-TOF mass spectrometry on the Sequenom® platform. Association was tested using the allelic test in PLINK.

Variant screening

We used Idaho®'s LightScanner high-resolution melting curve analysis to screen the eight coding exons of *EEF1D* for

variants in the same set of 862 cases and 1,014 controls. For technical reasons, a part of exon 3 of *EEF1D* could not be evaluated. For *LRRK1*, the ras of complex proteins (ROC, p.631 to 826), the C-terminal of ROC (COR, p.827 to 1241), and the kinase (p.1242 to 1525) domains as determined by an InterproScan sequence search or extracted from the literature [12] were screened. In the case of altered melting patterns suggestive of variants, Sanger sequencing ensued. Significance was judged using the χ^2 test.

Bioinformatic prioritization of variants

We collected a set of reference SNVs known to impair LRRK1 function. After computing a multiple sequence alignment using ClustalW based on *LRRK1*/*LRRK2* pairs in 18 organisms, we introduced mutations into *LRRK1* which mimic non-synonymous *LRRK2* mutations related to PD (rs33939927 (p.Arg1441Gly), rs35801418 (p.Tyr1699Cys), rs34637584 (p.Gly2019Ser), rs35870237 (p.Ile2020Thr)), and added *LRRK1* variants with a reported functional impact [12] to the set of reference SNVs. An in silico approach was applied to determine the disease potential of reference SNVs and the novel, non-synonymous *LRRK1* variants. To reduce the error rates of single models in predicting the functional effect of a given variant on the protein, we implemented a multi-model ensemble combining prediction results of six publically available prediction algorithms into a combined Pscore (Fig. 3a, online methods). Additionally, a Dscore was computed, scoring the severity of structural changes between the wild type and the variant peptide based on the mean square deviation (online methods). By combining the Pscore and the Dscore, we computed a single overall mutation score (Mscore), rating the disease potential of an SNV between 0 (harmless polymorphism) and 1 (disease mutation) (online methods). SNVs were then ranked by their Mscore, and hierarchical clustering was conducted by Ward's minimum variance agglomeration method and Euclidean distance matrix and analyzed in R; p values were calculated by multi-scale bootstrap resampling [13]. Also see supplement.

Cellular analyses

Cellular analyses were carried out as previously described [14]. For a detailed description, see supplement.

Results

Pedigree and clinical phenotype

We describe a five-generation family from Southern Germany in which six members were affected by PD and the pattern of inheritance seems to be autosomal dominant with reduced penetrance (Fig. 1). Clinical assessment revealed a tremor-dominant, levodopa-responsive Parkinson's syndrome with an age of onset at 56.7±1.15 years in all living affected individuals (V:8, V:9, and V:17, Online Resources Tabl 1). Further, all three affected individuals also showed positive Babinski signs and suffered episodes of depression. Mild to moderate cognitive impairment especially with regard to visuoconstruction, memory, and attention was present in all individuals. Dopamine transporter SPECT (DAT-SPECT) performed in two affected individuals (V:8 and V:17) revealed reduced tracer uptake in the putamen and asymmetrically in the caudate nucleus, in line with a diagnosis of PD.

The affected parent and aunt (IV:5 and IV:7) of the proband died before initiation of the study. An additional cousin, V:1, had Parkinson's syndrome but also suffered from multiple sclerosis. She also died before initiation of the study. Lastly, a second cousin removed by four generations is also known to suffer from late-onset PD. The prevalence of PD in the general population is approximately 1 % [15]. Accordingly, we expected to find at least one phenocopy in this extended pedigree of 114 individuals. Since no additional family members on her side of the family showed signs of PD and since she shared none of the candidate variants common to the other three affected individuals examined, we concluded that it is unlikely that PD in her case is due to the same genetic variant as in the other affected individuals.

Identification of candidate variants by exome sequencing and segregation analysis

A genome-wide CNV scan revealed no structural variation≥ 10 kb common to two affected members of the family (V:8 and V:17, Fig. 1). Exome sequencing was performed for the same individuals. This generated 6.57 gigabases (Gb) of alignable sequence for V:8 (average coverage=70.93, >8× coverage=90.65 %) and 6.67 Gb for V:17 (average coverage=76.29, >8× coverage=92.23 %). All detected variants shared by the two affected individuals (16,283 variants) were filtered against variants annotated in dbSNP132 as well as in-house exomes (n=1076) of individuals with unrelated diseases and variants with a minor allele frequency (MAF)≥0.01 were excluded from the follow-up, leaving 71 coding variants. Of these, 36 variants were predicted to alter the amino acid sequence (i.e., missense, nonsense, stop-loss, splice site, or frameshift variants and indels) and were genotyped in a third affected individual (V:9) (Online Resources Fig 1). Fifteen variants in 15 genes were present in all three affected individuals and were pursued further by Sanger sequencing-based testing for segregation in 32 members of the family belonging to generation V. Under the assumption that a given variant would be causal for PD, penetrances ranged between 30 and 50 %, with variants in

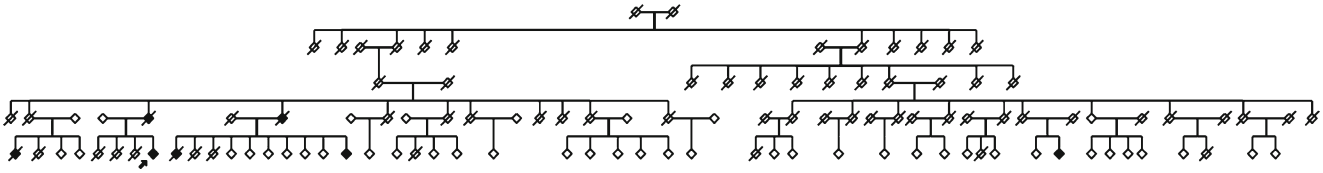


Fig. 1 Pedigree of family used for exome sequencing. *Open symbols* indicate unaffected family members; affected individuals are denoted by *closed symbols*. An *arrow* denotes the proband. Sex was obscured and birth order was altered to protect privacy. A *diagonal line* indicates a deceased individual

LRRK1, *EEF1D*, and *ARHGAP39* reaching the highest predicted penetrances (Table 1).

Frequency assessment of candidate variants in a case/control cohort

We genotyped the remaining 15 variants in a case/control sample, consisting of 862 individuals with PD and 1,014 KORA-AGE general population controls (Table 1). Two assays (*UGT1A9* p.Val167Ala, *TUBB6* p.Thr275Ala) did not meet quality control thresholds and were excluded from the analysis. The remaining 13 variants were, overall, very rare. Six (*EEF1D* p.Ala549Val, *MUC17* p.Gln4310X, *CCDC60* p.Arg155His, *NAAA* p.Arg211Trp, *PTPRN2* p.Glu317Lys, and *GLP2R* p.Ile61Met) were validated in the proband but were otherwise not found again in the 1,876 individuals tested. *FCGBP* p.Glu4657fs was present in one additional PD patient but not in controls, and *ZNF438* p.Thr454Ile was found in the proband and one control. Four additional variants annotated in dbSNP132 were identified at similar frequencies in cases and controls (*ARHGAP39* p.Arg667Gln and *MFSD3* p.Met311Thr) or were more common in controls than in cases (*AQP4* p.Met202Thr and *BRCA2* p.The1524Val). *LRRK1* p.Arg1261Gln was found in eight controls and in four cases of our case/control sample (MAF 0.23/0.40 %). However, in four other control samples (680 additional KORA general population controls (0.07 %), 1,076 in-house exomes (0.05 %), 1,000 genomes (0.00 %), and NHLBI-ESP exomes (0.09 %)), MAFs were significantly lower, and the variant was, therefore, also analyzed further.

Mutational screening of *EEF1D* and *LRRK1* in case/control cohort

While no single clear candidate for a causal variant emerged, two genes—*EEF1D* and *LRRK1*—were interesting with regard to functional considerations and predicted penetrance for PD in the family. The translation machinery has recently been implicated in PD pathogenesis [16, 17]. Also, the *EEF1D* p.Ala549Val variant was not found again in 3,064 individuals (genotyping cohort plus in-house exomes) and was also not annotated in the 1000 Genomes database. *LRRK1*, the paralog of the well-established PD gene *LRRK2*, has been shown to regulate endosomal protein transport, thus linking it to the lysosomal pathway [18] which may be compromised in PD [19, 20]. Formation of heterodimers between LRRK1 and LRRK2 has also been reported [21, 22]. We screened the coding regions of these genes in 862 Austrian and German PD cases and 1,014 controls searching for additional variants. This cohort comprised the same individuals used for the above frequency assessment of exome variants. We identified seven (six non-synonymous, one del) novel variants predicted to change the amino acid sequence of *EEF1D*. These were rare and occurred with similar frequencies in cases (five individuals with a variant) and controls (four individuals with a variant) ($p > 0.5$, χ^2 test, Online Resources Tab 2). Variants did not cluster in a specific part of the gene (Fig. 2). The ROC, COR, and kinase domains of *LRRK1* harbored a total of 20 novel amino acid sequence-changing variants (19 non-synonymous, 1 del) and 2 previously reported non-synonymous variants (rs56003881, rs41531245). Variants were found at similar frequencies in both groups (30 in cases, 31 in controls) ($p > 0.5$, χ^2 test, Online Resources Tab 2). While small numbers preclude quantitative analyses, it is noteworthy that within the first 20 bp of the kinase domain, variants were present in both cases and controls, while beyond p.1262, all non-synonymous variants identified in the kinase domain occurred in cases only (Fig. 2). None of the individuals harboring *LRRK1* variants were also positive for known *LRRK2* variants p.Arg1441Cys, p.Tyr1699Cys, p.Gly2019Ser, or p.Ile2020Thr.

Prioritization of *LRRK1* variants using a novel bioinformatics algorithm

Since heterodimer formation between *LRRK1* and *LRRK2* has been described [21, 22], we decided to further assess the identified variants in *LRRK1*. To this end, we used a novel bioinformatics algorithm based on a multi-model ensemble of prediction algorithms and structural analysis to select variants in *LRRK1* for functional follow-up. Mutation scores were calculated for the 19 novel, non-synonymous *LRRK1* variants identified in both cases and controls, the *LRRK1* variants (p.Lys746Glu, p.Phe1022Cys, p.Gly1411Arg and p.Ile1412Thr) corresponding to four known pathogenic *LRRK2* mutations (p.Arg1441Gly, p.Tyr1699Cys,

Table 1 Fifteen rare, non-synonymous variants shared by individuals V:8, V:9, and V:17 of family PARK_0001

Genomic position (hg19)	Gene	Alleles	In-house exomes	Genotyping		dbSNP132	1,000 Genomes	NHLBI-ESP	Transcript	Variation		Penetrance for PD (%)
				Cases	Controls		AF/DP	(EA only, hetero/ total individuals)		Nucleotide	Amino acid	
			$n=1076$	$n=862$	$n=1,014$							$n=32$
chr15:101593219	*LRRK1*	1	1	4	8		Not found	8/4,250	NM_024652.3	c.3782G>A	p.Arg1261Gln	50.00
chr8:144662740	*EEF1D*	1	0	1	0		Not found	Not found	NM_032378.4	c.1646G>A	p.Ala549Val	50.00
chr8:145771154	*ARHGAP39*	1	3	2	2	rs11994207	0.004:1,747	20/4,279	NM_025251.1	c.2000C>T	p.Arg667Gln	50.00
chr7:100694947	*MUC17*	1	0	1	0		Not found	Not found	NM_001040105.1	c.12928C>T	p.Gln4310X	42.85
chr8:145736082	*MFSD3*	1	3	4	2	rs35905340	0.003:1,950	19/4,280	NM_138431.1	c.932 T>C	p.Met311Thr	42.85
chr12:119926578	*CCDC60*	1	0	1	0		Not found	Not found	NM_178499.3	c.464G>A	p.Arg155His	42.85
chr19:40366263	*FCGBP*	1	1	2	0		Not found	Not found	NM_003890.2	c.13971_13971delC	p.Glu4657fs	42.85
chr4:76846923	*NAAA*	1	0	1	0		Not found	Not found	NM_001042402.1	c.631G>A	p.Arg211Trp	37.50
chr7:157931118	*PTPRN2*	1	1	1	0		Not found	3/4,297	NM_002847.3	c.949C>T	p.Glu317Lys	37.50
chr10:31137973	*ZNF438*	1	0	1	1		Not found	Not found	NM_001143769.1	c.1361G>A	p.Thr454Ile	37.50
chr2:234581080	*UGT1A9*	1	1	n/a	n/a		Not found	13/4,287	NM_021027.2	c.500 T>C	p.Val167Ala	33.33
chr18:24440758	*AQP4*	1	2	1	7	rs72557975	0.003:2,587	9/4,291	NM_004028.3	c.605A>G	p.Met202Thr	33.33
chr13:32913062	*BRCA2*	1	1	2	6	rs56386506	Not found	1/4,299	NM_000059.3	c.4570 T>G	p.Phe1524Val	30.00
chr17:9729563	*GLP2R*	1	0	1	0		Not found	Not found	NM_004246.1	c.183C>G	p.Ile61Met	30.00
chr18:12325722	*TUBB6*	1	0	n/a	n/a		Not found	16/4,284	NM_032525.1	c.823A>G	p.Thr275Ala	30.00

The rare variants common to all three affected individuals were genotyped in 862 cases and 1,014 controls. Penetrance with regard to the PD phenotype was assessed in 32 family members belonging to the same generation as the affected individuals

AF allele frequency, *DP* sequencing depth (number of reads), *EA* European American, *hetero* heterozygotes

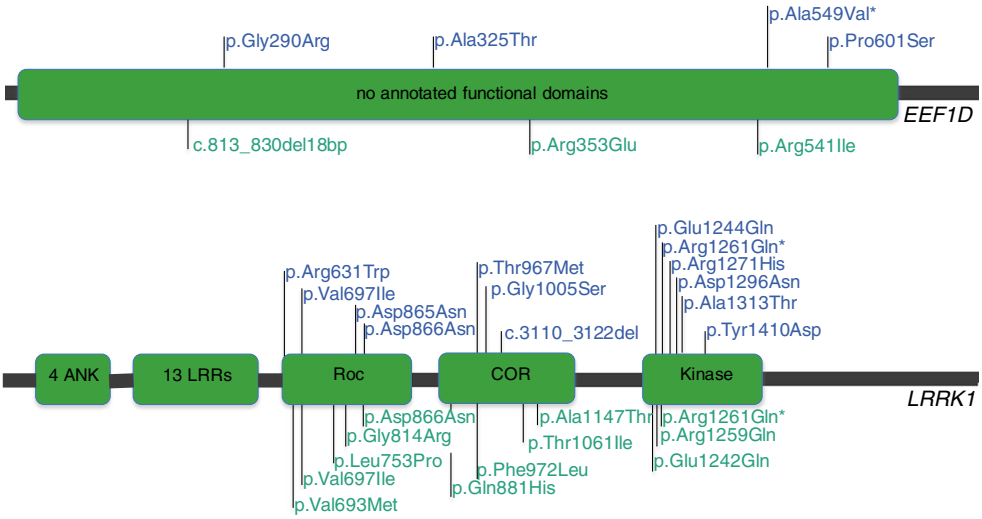


Fig. 2 Location of *EEF1D* and *LRRK1* variants identified in variant screening in relation to known functional domains. An *asterisk* denotes the variant identified by exome sequencing Variants printed in blue and annotated above the gene were found in cases, variants in green and below the gene were found in controls

p.Gly2019Ser, and p.Ile2020Thr) and three artificial variants known to abolish *LRRK1* GTP-binding (p.Lys651Ala) and kinase activity (p.Lys746Gly and p.Lys1270Trp) (Fig. 3a) [12]. Hierarchical clustering showed that three of the novel variants (p.Arg631Trp, p.Arg1271His and p.Tyr1410Asp)—present only in PD cases—clustered with the *LRRK1* equivalents of *LRRK2* p.Arg1441Gly, p.Tyr1699Cys, p.Gly2019Ser, and p.Ile2020Thr as well as the kinase- and GTP-binding dead amino acid substitutions (Fig. 3b). Accordingly, these three variants in addition to the initial variant identified by exome sequencing (p.Arg1261Gln) were selected for functional follow-up.

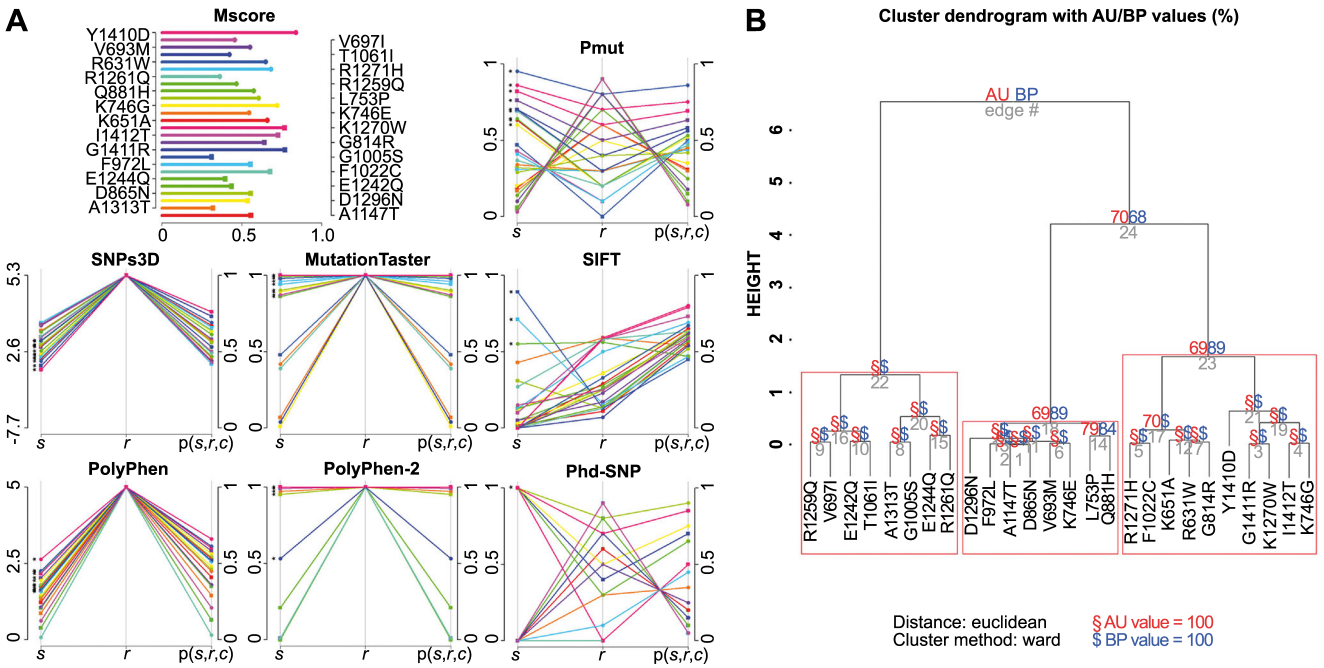


Fig. 3 Prediction of pathogenic potential of newly identified variants. **a** For each variant (*colored lines*) the predicted score s of an individual algorithm, its reliability r, and the transformed score $p(s, r, c)$ are shown. Variants holding a predicted disease-causing potential (class=1) were respectively marked with an *asterisk*. The diverse results among each single algorithm motivated the calculation of one combined score (*Pscore*), which was adjusted by additional structural analyses (*Dscore*) resulting in a mutation score (*Mscore*). The highest scoring variant is p.Tyr1410Asp (Mscore=0.839), a variant only present in PD cases, followed by the *LRRK2* equivalent of Gly2019Ser (Mscore=0.771), the loss of autophosphorylation mutation Lys1270Trp (Mscore=0.768), and two variants abolishing kinase activity: Ile1412Thr (Mscore=0.728) and Lys746Gly (Mscore=0.723). **b** Hierarchical clustering with Ward's minimum variance agglomeration method and Euclidean distance matrix shows that three of the novel variants which were only found in individuals with PD (p.Arg631Trp, p.Arg1271His, and p.Tyr1410Asp) cluster with the *LRRK1* equivalents of *LRRK2* p.Arg1441Gly, p.Tyr1699Cys, p.Gly2019Ser, and p.Ile2020Thr as well as the *LRRK1* kinase- and GTP-binding dead amino acid substitutions

Functional assessment of *LRRK1* variants

In SHSY5Y neuroblastoma cells, levels of protein expression as assessed by Western blot were not changed by any of the four newly identified variants or the artificial variants ablating GTP-binding (p.Lys651Ala) or kinase activity (p.Lys1270Trp) (Fig. 4a). Likewise, the presence of these variants was not associated with significant toxicity as measured by MTT assay (Fig. 4b) and did not alter cytoplasmic localization of myc-tagged LRRK1 (Fig. 4c). Like others [23, 24], we could not detect LRRK1 kinase activity above the background and could not determine whether activity was altered by the variants.

Discussion

In an unbiased, whole exome approach, we identified a variant in *LRRK1* (p.Arg1261Gln) as a candidate for a potentially causal variant in familial PD. Although this finding is intriguing and functionally plausible, we are unable to conclude that this is indeed the cause of PD in our family. For one, the variant was found in both cases and controls in our larger case/control sample. Yet, the actual variant frequency in controls appears to be lower than that found in the KORA-AGE cohort (8 in 1,014 KORA-AGE vs. 0 in 1,000 genomes, 1 in 1,076 in-house exomes, and 1 in additional 680 KORA controls), and it could be possible that KORA-

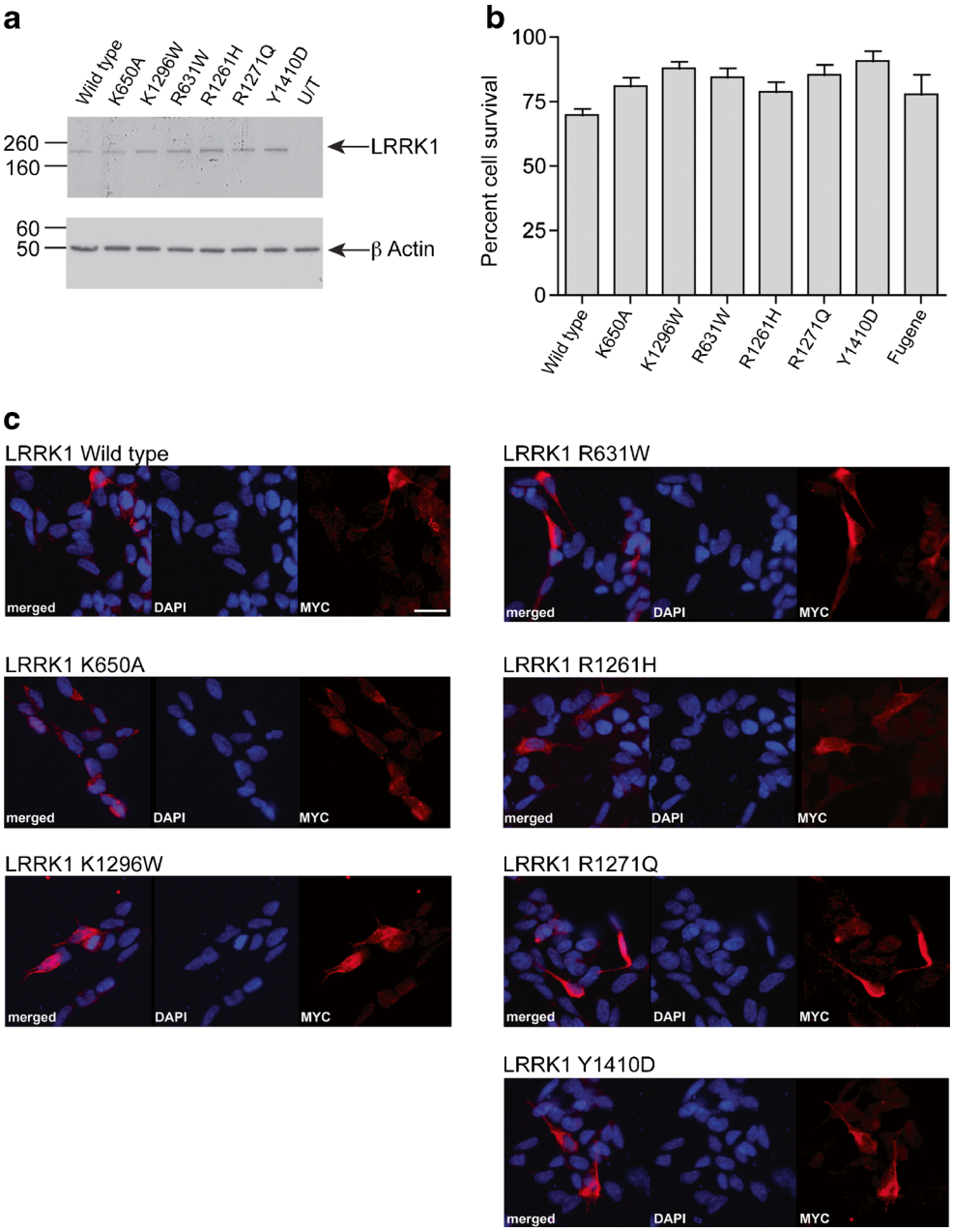


Fig. 4 Cellular expression of LRRK1 and mutant variants. **a** Western blot analysis of myc-tagged LRRK1 expression in SHSY5Y cells with beta actin-loading control. **b** Analysis of LRRK1 toxicity as measured by MTT assay in SHSY5Y cells. No significant toxicity was associated with wild-type LRRK1, artificial mutations in LRRK1, or disease-associated coding changes. Data is expressed as percentage of untransfected control cells, mean, and standard error measurement displayed. **c** Immunocytochemistry analysis of myc-tagged LRRK1 constructs. Staining for myc is shown separately and merged. All tagged constructs displayed a diffuse cytoplasmic staining pattern. *Scale bar* = 20 μm

AGE is enriched for the *LRRK1* p.Arg1261Gln variant due to a founder effect or that is also present in controls because its PD-related nature depends on a specific genetic context. Secondly, the other 14 identified rare variants shared by all three affected individuals also represent potential candidates. Especially, *EEF1D* p.Ala549Val, which was not found again in any individual genotyped (n=3064, case/control sample and in-house exomes) or the 1,000 genomes or the NHLBI-ESP exomes, represents another good contender. In general, these findings draw attention to the fact that in many cases, very large populations will need to be evaluated to conclusively judge the disease-related nature of a rare variant such as those identified by exome sequencing. Most recent studies show that while the power to detect associations for genes harboring rare variants varies widely across genes, only <5 % of genes achieved 80 % power even assuming high odds ratios (OR) of 5 when tested in 400 cases and 400 controls [25]. Ultimately, it is also possible that the truly causal variant was not picked up in this study because it lies outside the targeted regions of the exome.

The fact that both the *LRRK1* and *EEF1D* variants were also found in three unaffected members of the family each per se does not contradict potential causality as it is known from other autosomal dominant forms of PD that even among members of a single family, penetrance of known PD mutations can vary widely. Of individuals who harbor the *LRRK2* p.Gly2019Ser mutation, for example, only 28 % will develop PD by the age of 59 years [26]. Thus, predicted penetrances of the variants identified in our family are in line with what is reported in the literature for other forms of autosomal dominant PD.

Both LRRK1 and LRRK2 belong to the Roco family of proteins. These proteins are likely to perform a number of different functions as they are not only characterized by a conserved Ras-like GTPase domain called ROC and a characteristic COR domain of unknown function but also harbor kinase and protein–protein interaction domains [27]. While a contribution of mutations in *LRRK2* to disease development in PD seems firmly established, the role of *LRRK2* paralog *LRRK1* is unclear. It is known that LRRK1 and LRRK2 form heterodimers in HEK293T cells [21, 22] and that both proteins are expressed in similar tissues. Accordingly, a hypothetical role for LRRK1 in addition to LRRK2 is plausible.

The precise role of LRRK2 in PD pathogenesis, however, has not been fully established. Accordingly, even if one were to assume a similar role of LRRK1 in disease development, exactly which function of the protein would be involved in the disease is uncertain. Therefore, the lack of a functional effect on protein expression levels, subcellular localization, and cell viability of the *LRRK1* variants we identified does not equate to a definitely missing role of *LRRK1* in PD. Interestingly, *Lrrk1* has also been implicated in a quantitative trait locus for dopaminergic amacrine cell number in the murine retina [28]. Further, the recent link between LRRK1 and endosomal protein trafficking [18, 29] is also very intriguing in light of the fact that one of the postulated pathomechanisms for LRRK2 in PD involves aberrant lysosomal function or localization [20, 30, 31].

However, studies addressing the role of both common and rare genetic variants in *LRRK1* with regard to PD do not seem to substantiate the conception of *LRRK1* as a “PD gene” [21, 32–34]. While none of these studies found a common or rare variant clearly linked to PD, nonetheless, across three studies ([32, 34] and our study), the p.Thr967Met variant has only been identified in seven out of 1,552 cases but not in any of 1,535 controls ($p_{\text{nominal}} \leq 0.01$, χ^2 test; not significant after correction for multiple testing, OR=14.90 (95 % confidence interval=0.85 to 261.18)). Yet, in a family with multiple individuals with PD, the variant did not segregate with the phenotype [34]. Evidence also suggests that variants in *LRRK1* are able to modify the PD phenotype. Tunisian individuals with *LRRK2* p.Gly2019Ser showed a trend towards a 6-year earlier age of onset when they also carried *LRRK1* p.Leu416Met [20]. In line with this, it has been demonstrated for other genetic disorders that genetic variants at related loci can both drive and modify a given phenotype depending on the variant and the genetic context [35]. At the moment, both functional and genetic data addressing a role of *LRRK1* as a PD gene are inconclusive. Nonetheless, it is interesting that in our unbiased whole exome approach, one of the top candidate variants for a genetic factor underlying or contributing to the PD phenotype in our family is a non-synonymous variant in the kinase domain of *LRRK1* and that other individuals suffering from PD harbor *LRRK1* variants (p.Tyr1410Asp) only with one amino acid away from the location which is equivalent to the prominent p.Gly2019Ser mutation of *LRRK2*.

In summary, all variants shared by the three affected individuals in our family represent potential causal or modifying alleles in PD. As is the case for all rare and very rare variants, establishing definitive causality is difficult, and only the identification of additional PD families harboring these variants or their analysis in sufficiently powered case/control studies will tell whether these variants do indeed hold a role in bringing about PD.

Acknowledgments We are gratefully indebted to Katja Junghans, Susanne Lindhof, Jelena Golic, Sybille Frischholz, and Regina Feldmann at the Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany, for their expert technical assistance in performing Sequenom genotyping and Light Scanner analyses. This study was funded by in-house institutional funding from Technische Universität München and Helmholtz Zentrum München, Munich, Germany. Recruitment of case and control cohorts was supported by institutional (HelmholtzZentrumMünchen, Munich, Germany) and government funding from the German Bundesministerium für Bildung und Forschung (03.2007-02.2011 FKZ 01ET0713). PAL is funded by a Parkinson's UK research fellowship

(F1002). The work performed by PAL, SD, and CM was also funded in part by grants from the Michael J. Fox Foundation LRRK2 consortium and by the Wellcome Trust/MRC Joint Call in Neurodegeneration award (WT089698) to the UK Parkinson's Disease Consortium (UKPDC) whose members are from the UCL Institute of Neurology, the University of Sheffield, and the MRC Protein Phosphorylation Unit at the University of Dundee.

References

1. Payami H, Larsen K, Bernard S, Nutt J (1994) Increased risk of Parkinson's disease in parents and siblings of patients. Ann Neurol 36:659–661
2. Bonifati V, Fabrizio E, Vanacore N, De Mari M, Meco G (1995) Familial Parkinson's disease: a clinical genetic analysis. Can J Neurol Sci 22:272–279
3. Zimprich A, Biskup S, Leitner P et al (2004) Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 44:601–607
4. Bonifati V, Rizzu P, van Baren MJ et al (2003) Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science 299:256–259
5. Kitada T, Asakawa S, Hattori N et al (1998) Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 392: 605–608
6. Paisan-Ruiz C, Jain S, Evans EW et al (2004) Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron 44:595–600
7. Polymeropoulos MH, Lavedan C, Leroy E et al (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 276:2045–2047
8. Valente EM, Abou-Sleiman PM, Caputo V et al (2004) Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science 304:1158–1160
9. Zimprich A, Benet-Pages A, Struhal W et al (2011) A Mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. Am J Hum Genet 89:168–175
10. Vilarino-Güell C, Wider C, Ross OA et al (2011) VPS35 mutations in Parkinson disease. Am J Hum Genet 89:162–167
11. Wichmann HE, Gieger C, Illig T (2004) KORAgen—resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 67(Suppl 1):26–30
12. Korr D, Toschi L, Donner P, Pohlenz HD, Kreft B, Weiss B (2006) LRRK1 protein kinase activity is stimulated upon binding of GTP to its Roc domain. Cell Signal 18:910–920
13. Shimodaira H (2004) Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. Ann Statistics 32:2616–2641
14. Greggio E, Lewis PA, van der Brug MP et al (2007) Mutations in LRRK2/dardarin associated with Parkinson disease are more toxic than equivalent mutations in the homologous kinase LRRK1. J Neurochem 102:93–102
15. Elbaz A, Bower JH, Maraganore DM et al (2002) Risk tables for parkinsonism and Parkinson's disease. J Clin Epidemiol 55:25–31
16. Chartier-Harlin MC, Dachsel JC, Vilarino-Guell C et al (2011) Translation initiator EIF4G1 mutations in familial Parkinson disease. Am J Hum Genet 89:398–406
17. Gehrke S, Imai Y, Sokol N, Lu B (2010) Pathogenic LRRK2 negatively regulates microRNA-mediated translational repression. Nature 466:637–641
18. Hanafusa H, Ishikawa K, Kedashiro S et al (2011) Leucine-rich repeat kinase LRRK1 regulates endosomal trafficking of the EGF receptor. Nat Comm 2:158
19. Sidransky E, Nalls MA, Aasly JO et al (2009) Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N Engl J Med 361:1651–1661
20. Tong Y, Yamaguchi H, Giaime E et al (2010) Loss of leucine-rich repeat kinase 2 causes impairment of protein degradation pathways, accumulation of alpha-synuclein, and apoptotic cell death in aged mice. Proc Natl Acad Sci U S A 107:9879–9884
21. Dachsel JC, Nishioka K, Vilarino-Güell C et al (2010) Heterodimerization of Lrrk1-Lrrk2: implications for *LRRK2*-associated Parkinson disease. Mech Age Dev 131:210–214
22. Klein CL, Rovelli G, Springer W, Schall C, Gasser T, Kahle PJ (2009) Homo- and heterodimerization of ROCO kinases: LRRK2 kinase inhibition by the LRRK2 ROCO fragment. J Neurochem 111: 703–715
23. Deng X, Dzamko N, Prescott A et al (2011) Characterization of a selective inhibitor of the Parkinson's disease kinase LRRK2. Nat Chem Biol 7:203–205
24. Civiero L, Vancraenenbroeck R, Belluzzi E et al (2012) Biochemical characterization of highly purified leucine-rich repeat kinases 1 and 2 demonstrates formation of homodimers. PLoS One 7:e43472
25. Tennessen JA, Bigham AW, O'Connor T et al (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337:64–69
26. Healy DG, Falchi M, O'Sullivan SS et al (2008) Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case–control study. Lancet Neurol 7:583–590
27. Marin I, van Egmond WN, van Haastert PJM (2008) The Roco protein family: a functional perspective. FASEB 22:3103–3110
28. Whitney IE, Raven MA, Ciobanu DC, Williams RW, Reese BE (2009) Multiple genes on chromosome 7 regulate dopaminergic amacrine cell number in the mouse retina. Invest Ophthal Vis Sci 50:1996–2003
29. Ishikawa K, Nara A, Matsumoto K, Hanafusa H (2012) EGFR-dependent phosphorylation of leucine-rich repeat kinase LRRK1 is important for proper endosomal trafficking of EGFR. Mol Biol Cell 23:1294–1306
30. Gómez-Suaga P, Luzón-Toro B, Churamani D et al (2012) Leucine-rich repeat kinase 2 regulates autophagy through a calcium-dependent pathway involving NAADP. Hum Mol Genet 21:511–525
31. Dodson MW, Zhang T, Jiang C, Chen S, Guo M (2012) Roles of the *Drosophila LRRK2* homolog in Rab7-dependent lysosomal positioning. Hum Mol Genet 21:1350–1363
32. Haugarvoll K, Toft M, Ross OA, White LR, Aasley JO, Farrer MJ (2007) Variants in the LRRK1 gene and susceptibility to Parkinson's disease in Norway. Neurosci Lett 416:299–301
33. Chung SJ, Armasu SM, Biernacka JM et al (2011) Common variants in PARK loci and related genes and Parkinson's disease. Mov Disord 26:280–288
34. Taylor JP, Hulihan MM, Kachergus JM et al (2007) Leucine-rich repeat kinase 1: a paralog of LRRK2 and a candidate gene for Parkinson's disease. Neurogenetics 8:95–102
35. Davis EE, Zhang Q, Liu Q et al (2011) TTC21B contributes both causal and modifying alleles across the ciliopathy spectrum. Nat Genet 43:189–196

OPEN ACCESS Freely available online

PLOS ONE

Rare Variants in *PLXNA4* and Parkinson's Disease

Eva C. Schulte[1,2], Immanuel Stahl[1,2], Darina Czamara[3,4], Daniel C. Ellwanger[5], Sebastian Eck[2], Elisabeth Graf[2], Brit Mollenhauer[6,7], Alexander Zimprich[8], Peter Lichtner[2,9], Dietrich Haubenberger[8], Walter Pirker[8], Thomas Brücke[10], Benjamin Bereznai[11], Maria J. Molnar[11], Annette Peters[12], Christian Gieger[13], Bertram Müller-Myhsok[3,4], Claudia Trenkwalder[6,7], Juliane Winkelmann[1,2,4,9,14]*

1 Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität, München, Munich, Germany, **2** Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany, **3** Max-Planck Institut für Psychiatrie, Munich, Germany, **4** Munich Cluster for Systems Neurology (SyNergy), Munich, Germany, **5** Chair for Genome-Oriented Bioinformatics, Technische Universität München, Life and Food Science Center Weihenstephan, Freising-Weihenstephan, Germany, **6** Paracelsus Elena Klinik, Kassel, Germany, **7** Neurochirurgische Klinik, Georg August Universität, Göttingen, Germany, **8** Department of Neurology, Medical University of Vienna, Vienna, Austria, **9** Institut für Humangenetik, Technische Universität München, Munich, Germany, **10** Department of Neurology, Wilhelminenspital, Vienna, Austria, **11** Center for Molecular Neurology, Department of Neurology, Semmelweis University, Budapest, Hungary, **12** Institute for Epidemiology II, Helmholtz Zentrum München, Munich, Germany, **13** Institute for Genetic Epidemiology, Helmholtz Zentrum München, Munich, Germany, **14** Department of Neurology and Neurosciences, Stanford University, Palo Alto, California, United States of America

Abstract

Approximately 20% of individuals with Parkinson's disease (PD) report a positive family history. Yet, a large portion of causal and disease-modifying variants is still unknown. We used exome sequencing in two affected individuals from a family with late-onset familial PD followed by frequency assessment in 975 PD cases and 1014 ethnically-matched controls and linkage analysis to identify potentially causal variants. Based on the predicted penetrance and the frequencies, a variant in *PLXNA4* proved to be the best candidate and *PLXNA4* was screened for additional variants in 862 PD cases and 940 controls, revealing an excess of rare non-synonymous coding variants in *PLXNA4* in individuals with PD. Although we cannot conclude that the variant in *PLXNA4* is indeed the causative variant, these findings are interesting in the light of a surfacing role of axonal guidance mechanisms in neurodegenerative disorders but, at the same time, highlight the difficulties encountered in the study of rare variants identified by next-generation sequencing in diseases with autosomal dominant or complex patterns of inheritance.

Citation: Schulte EC, Stahl I, Czamara D, Ellwanger DC, Eck S, et al. (2013) Rare Variants in *PLXNA4* and Parkinson's Disease. PLoS ONE 8(11): e79145. doi:10.1371/journal.pone.0079145

Editor: Christian Wider, Centre Hospitalier Universitaire Vaudois (CHUV), Switzerland

Received July 31, 2013; **Accepted** September 18, 2013; **Published** November 11, 2013

Funding: The study was funded in its majority through institutional funding from Technische Universität München and Helmholtz Zentrum München, Munich, Germany. Dr. DH received an NINDS Intramural Competitive Fellowship and research support from the Austrian Science Fund (Erwin Schroedinger Fellowship, project#J2783-B09) and the NINDS Intramural Research Program. Dr. BB and Dr. MJM received support from the Hungarian National Innovation Office (TAMOP-4-2-1/B-03/11 KMR-2010-001). Recruitment and management of the KORA cohort is funded by both institutional (Helmholtz Zentrum München, Munich, Germany) and government funding from the German Bundesministerium für Bildung und Forschung (03.2007-02.2011 FKZ 01ET0713). DC was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation) within the framework of the Munich Cluster for Systems Neurology (EXC 1010 SyNergy). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have read the journal's policy and have the following conflicts of interest to declare: Dr. ECS, Mr. IM, Dr. DC, Mr. DCE, Mr. SE and Mrs. EG report no disclosures. Dr. BM received travel compensation from Novartis and Boehringer-Ingelheim, lecturing fees from Orion and Glaxo-Smith-Kline, grant support from GE Healthcare, Boehringer-Ingelheim, Desitin, TEVA-Pharma, and serves as a consultant to Bayer-Schering Pharma and the Michael J. Fox Foundation for Parkinson's Disease Research. Dr. DH received speaker honoraria from Ipsen Pharma. Dr. WP received speaker honoraria from AOP Orphan Pharma, Medtronic Inc., Novartis, Boehringer-Ingelheim, Abbott Pharm and UCB as well as travel compensation from Ipsen Pharma, Boehringer-Ingelheim, and Medtronic Inc. Dr. BB declares no conflicts of interest. Dr. MJM serves/has served on scientific advisory boards for Genzyme Europe B.V., received speaker honoria from Roche and serves as the Editor-in-Chief of the Hungarian edition of Neurology. Dr. AZ, Dr. TB, Dr. PL, Dr. AP, Dr. CG and Dr. BMM report no disclosures. Dr. CT serves on scientific advisory boards for Boehringer Ingelheim and UCB, has received speaker honoraria from Boehringer Ingelheim, UCB, and Mundipharma as well as travel compensation from UCB, Boehringer-Ingelheim, and Mundipharma. Dr. JW serves on a scientific advisory board for UCB, has received speaker honoraria from UCB and Vifor Pharma. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: winkelmann@lrz.tum.de

Introduction

Characterized by resting-tremor, bradykinesia, rigidity, and postural instability, Parkinson's disease (PD) is one of the most prominent neurodegenerative disorders. Genetic factors contribute significantly to the risk of developing PD–both sporadic and familial. Although up to 20% of PD cases are believed to be familial [1,2], thus far, variants in only a few genes have been unequivocally shown to underlie familial PD. These include *PARK2*, *PINK1*, *PARK7*, *SNCA*, and *LRRK2* [3–8]. While all of these genes were identified by classical linkage analysis in large, multi-generation families, recently, next-generation sequencing has enabled the identification of disease-causing variants in smaller families and with an onset later in life without the need of genotypic information from more than one generation of affected individuals. By exome sequencing, *VPS35* was identified as a gene involved in late-onset familial PD [9,10]. Still, to date, the identified genes only explain a small portion of the genetic "burden" in PD. However, a thorough understanding of the genetic alterations implicated in disease development is necessary to better comprehend disease pathogenesis and to provide more specific and, thus, more effective treatment options in the future.

Here, we describe exome sequencing of a German family with autosomal dominant late-onset PD in an attempt to pinpoint the disease-causing genetic variant.

Methods

Ethics Statement

Ethics review board approval was obtained from the ethics review board at Klinikum rechts der Isar, Technische Universität München, and Bayerische Landesärztekammer, both Munich, Germany, Hessische Landesärztekammer, Frankfurt, Germany, the ethics review board at Medical University Vienna, Vienna, Austria, and the ethics review board at Semmelweis University, Budapest, Hungary. Participants' written informed consent was obtained.

Participants

All living family members received a detailed neurologic exam by neurologists specializing in movement disorders. Cases and controls used in genotyping and variant screening have been reported previously [10,11] and are described in more detail in the supplement.

Exome Sequencing

Exome sequencing was performed with DNA isolated from lymphozytes of IV:11 and IV:18 on a Genome Analyzer IIx system (Illumina) after in-solution enrichment of exonic sequences (SureSelect Human All Exon 38 Mb kit for IV:11 and 50 Mb kit for IV:18, Agilent) as 76 bp paired-end runs. Read alignment was carried out with BWA (version 0.5.8). Single-nucleotide variants and small insertions and deletions (indels) were detected with SAMtools (version 0.1.7). Raw sequencing data are available upon request.

Genotyping

All ten candidate variants tested for segregation by Sanger sequencing were genotyped in 975 cases and 1014 population-based controls pertaining to the KORA-AGE cohort using MALDI-TOF masspectrometry on the Sequenom® platform. Demographic data are given in the supplement. Association was tested by allelic statistics as implemented in PLINK.

Linkage Analysis

We genotyped six family members (IV:11, IV:14, IV:16, IV:18, IV:20 and IV:21) with oligonucleotide SNP arrays (500 K, Illumina). Parametric linkage analysis was performed using a subset of 12,875 SNPs using MERLIN and an autosomal dominant model with incomplete penetrance of 70%.

Variant Screening

We used Idaho®'s LightScanner high-resolution melting curve analysis to screen the coding regions and exon/intron boundaries of *PLXNA4* for variants. 862 cases and 940 population-based controls pertaining to the KORA-AGE cohort were included in the screening. Demographic data are given in the supplement. In the case of an altered melting pattern, Sanger sequencing ensued to identify the underlying variant. Group comparisons between cases and controls were performed for each gene and each variant separately using Fisher's Exact and χ^2 tests as appropriate.

Cell Viability and Immunocytochemistry

Cultured primary fibroblasts from IV:11 and an offspring were stained using a live/dead staining (Invitrogen) and analyzed by FACS and stained with anti-PLXNA4 (1:100, Sigma) and analyzed by fluorescence microscopy. Details are given in the supplement.

Construction of a Qualitative Systems Biological Model

To investigate the role of *PLXNA4* in the PD biological system, we applied an integrative modeling approach to construct a qualitative multifactorial interaction network linking *PLXNA4* and genetic factors associated with PD. An interactome with known and predicted interactions of *PLXNA4* and its direct neighbors was prepared based on four commonly used databases and integrated to known PD pathways from KEGG and CIDeR as well as a manual literature search. For a detailed description see supplement.

Results

Pedigree and Clinical Phenotype

We describe a five-generation family from Central Germany in which four members were affected by PD and the pattern of inheritance seems to be autosomal dominant with reduced penetrance (Figure 1). Clinical assessment revealed tremor-dominant, levodopa-responsive parkinsonism with an age of onset at 60 and 67 years of age in the two affected individuals examined (Table S1 in File S1). Both individuals also reported subjective cognitive impairment. Restless legs syndrome was present in IV:11 as well as one of her children. Transcranial ultrasound showed bilateral hyperechogenicity of the substantia nigra in IV:18 but was not performed in IV:11. MRI was in line with a diagnosis of PD in both. The affected parent (III:7) and aunt (III:5) of IV:11 were deceased before initiation of the study, so that no detailed phenotype information is available. Moreover, another aunt (III:2) on the same side of the family was reported to have suffered from an unclassified form of dementia.

Identification of Candidate Variants by Exome Sequencing and Frequency Assessment of Candidate Variants in a Case/Control Cohort

Exome sequencing was performed using DNA from two second cousins (IV:11 and IV:18, Figure 1A). This generated 11.68 gigabases (Gb) of alignable sequence for IV:11 (average coverage = 108.46, base pairs with >8 reads = 93.67%) and 15.02 Gb for patient IV:18 (average coverage = 154.13, base pairs with >8 reads = 94.74%). All 28,803 detected variants shared by the two affected individuals were filtered against in-house exomes (n = 1739) of individuals with unrelated diseases. Here, variants were allowed to be present in ≤1% of exomes. Moreover, synonymous and non-coding variants as well as all variants annotated in dbSNP135 with a minor allele frequency (MAF) ≥0.01 were excluded from the follow-up (Figure S1). No known variants believed to play a causative role in PD were found in either IV:11 or IV:18.

All ten remaining missense, nonsense, stoploss, splice site or frameshift variants and indels were genotyped in 975 cases and 1014 population-based controls (Table 1). The variants were, overall, very rare. Two (*PLXNA4* p.Ser657Asn and *OGN* p.Leu124fs) were validated in the individual in whom they were first identified but were otherwise not found again in the 1989 individuals tested. *CPNE1* p.Ser1831Thr was present in the index case as well as one additional control individual and *GOLGA4* p.Gln425Arg was identified in one additional PD patient. The other six variants (*RBM28* p.Asp300Gly, *IMPDH1* p.His296Arg, *ARPP21* p.Ala576Thr, *PHF2* p.Ser840Asn, *SLC22A13* p.Arg16His and *SPANXE* p.Leu42Ile) were not as rare (MAF≥0.03%) and

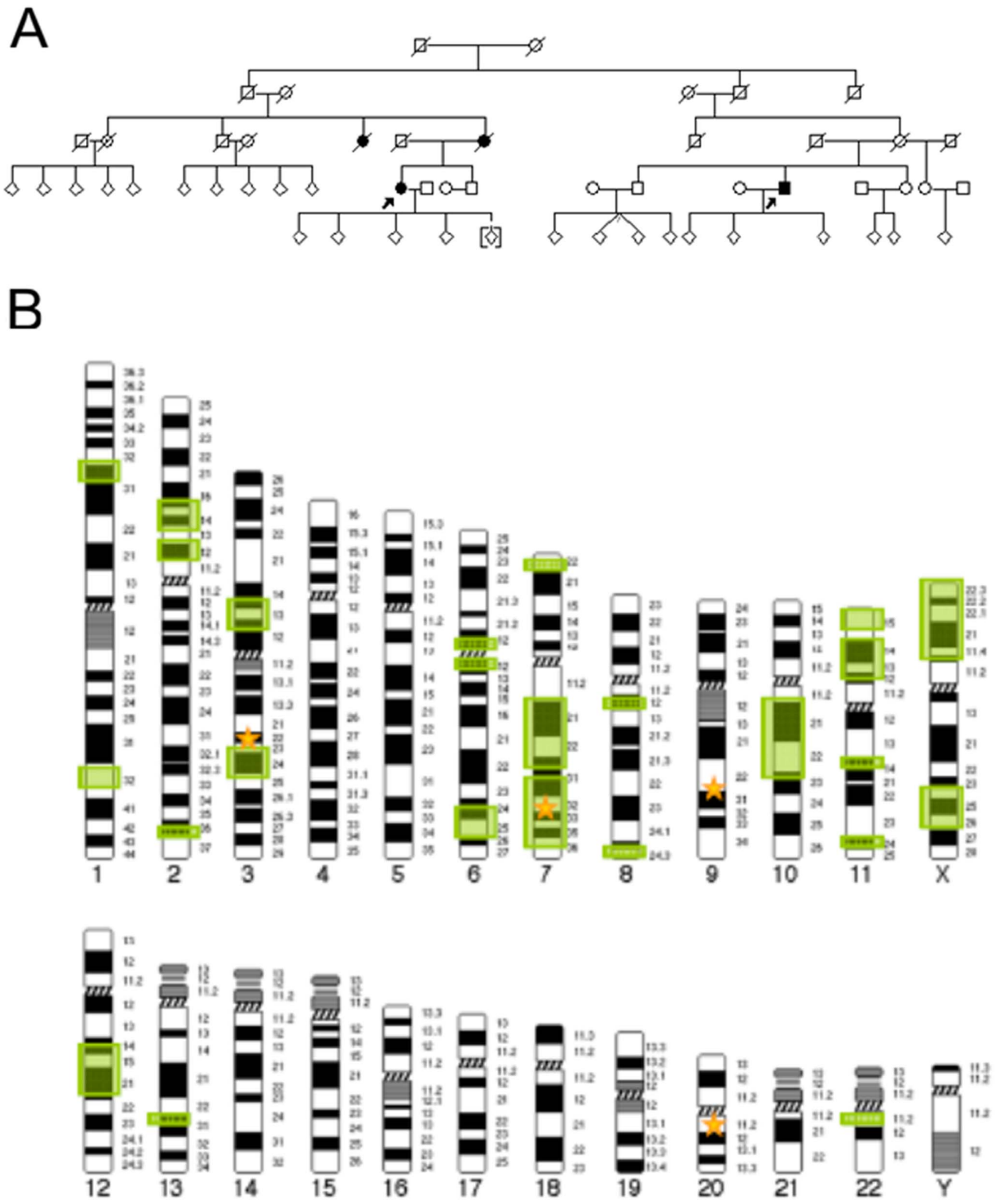


Figure 1. Pedigree and Linkage Analysis. (A) Pedigree of family used for exome sequencing. Open symbols indicate unaffected family members, affected individuals are denoted by closed symbols. An arrow denotes the individuals whose exomes were sequenced. Sex was obscured and birth order was altered to protect privacy. A diagonal line indicates a deceased individual. (B) 25 genomic regions on 12 chromosomes with logarithm of the odds (LOD) score≥0.5 were identified by linkage analysis. Green boxes represent genomic regions with LOD≥0.5, yellow stars represent the location of the four candidate genes remaining after frequency assessment (*GOLGA4*-chr3, *PLXNA4*-chr7, *OGN*-chr9, *CPNE1*-chr20). *PLXNA4* on chromosome 7 represents the only of the four genes overlapping a genomic region with LOD≥0.5.
doi:10.1371/journal.pone.0079145.g001

Table 1. Ten Rare, Non-synonymous Variants Shared by Individuals IV:11 and IV:18 of Family PARK_0005.

Genomicposition (hg19)	gene	number of alleles	*in house* exomes (n = 1739)	genotyping		dbSNP135	1000genomes	NHLBI-ESP [39] (EA only)	transcript	variation		penetrance for PD in % (n = 6)	PolyPhen2
				cases (n = 975)	controls (n = 1014)					nucleotide	amino acid		
chr3:35780947	*ARPP21*	1	12	9	11	rs151173813	0.0031:2218	A = 37/G = 8563	NM_001267617.1	c.1726G>A	p.Ala576Thr	N/A	benign
chr3:37365968	*GOLGA4*	1	4	2	0	rs139536585	not found	G = 8/A = 8592	NM_001172713.1	c.1274A>G	p.Gln425Arg	66.67%	benign
chr7:127950857	*RBM28*	1	1	6	5	rs148028531	0.0007:14795	C = 20/T = 8580	NM_018077.2	c.2273T>C	p.Asp758Gly	40.00%	poss. damaging
chr7:128037009	*IMPDH1*	1	7	6	5	rs61751223	0.0052:2280	C = 23/T = 8577	NM_000883.3	c.887T>C	p.His296Arg	40.00%	benign
chr7:131910932	*PLXNA4*	1	0	1	0	novel	not found	not found	NM_020911.1	c.1970C>T	p.Ser657Asn	40.00%	prob. damaging
chr9:95155422–95155423	*OGN*	1	0	1	0	novel	not found	not found	NM_014057.3	c.372_373 delAA	p. Leu124fs	50.00%	frameshift
chr20:34219872	*CPNE1*	1	0	1	1	novel	not found	not found	NM_003915.5	c.547A>T	p.Ser183Thr	66.67%	benign
chr9:96436037	*PHF2*	1		15	26	rs41276200	0.002:2389	A = 120/G = 8480	NM_005392.3	c.2519G>A	p.Ser840Asn	N/A	benign
chr3:38307398	*SLC22A13*	1		20	23	rs72542450	0.01:1459	A = 84/G = 8516	NM_004256.3	c.47G>A	p.Arg16His	N/A	benign
chrX:140785792	*SPANXE*	1,2		14/7	9/15	novel	not found	not in database	NM_145665.1	c.124G>T	p.Leu42Ile	N/A	not scored

The rare variants common to the two affected individuals were genotyped in 975 cases and 1014 controls. Penetrance with regard to the PD phenotype was assessed in 6 family members belonging to the same generation as the affected individuals. EA = European American.

doi:10.1371/journal.pone.0079145.t001

found at similar frequencies in both cases and controls and were, therefore, regarded to be unlikely candidates (Table 1).

Segregation Analysis and Genotyping of Additional *PLXNA4* Variants

The remaining four variants shared by the two affected individuals (Table 1) were pursued further by Sanger-sequencing-based testing for segregation in 6 family members belonging to generation IV. Under the assumption that a given variant would be causal for PD, penetrance ranged between 40.0 and 66.6% in 6 individuals belonging to generation IV. Moreover, on careful scrutiny of the exome data, both index patients were found to harbor one additional, variant of *PLXNA4* (p.Phe40Leu (rs145024048, 111/8489 in NHLBI-ESP exomes) for IV:11 and p.Arg302His (rs143813209, 3/8597 in NHLBI-ESP exomes) for IV:18). These two variants were also genotyped in 15 additional members of the family. *PLXNA4* p.Phe40Leu was found in 5 additional individuals and p.Arg302His was found in 7 additional family members. Importantly and contrary to the exome sequencing data, by Sanger sequencing, IV:11 was also found to harbor the *PLXNA4* p.Arg302His variant. The combination of the *PLXNA4* index variant and p.Phe40Leu was present only in IV:11, while the index variant and *PLXNA4* p.Arg302His were found in a total of 7 individuals belonging to the pedigree. None of the three additional candidate genes harbored additional non-synonymous coding variants in either IV:11 or IV:18.

Linkage Analysis

In order to further prioritize genes for follow-up, we performed parametric linkage analysis. In doing so, we identified 25 genomic regions with a suggestive linkage signal (LOD≥0.5) (Figure 1B). Only one of these regions, located on chromosome 7 (chr7:106,254,234 to 134,663,671; maximum two-point LOD score = 0.76), contained one of the four candidate genes identified during exome sequencing, lending further support to the potential causality of variants in *PLXNA4*.

Mutational Screening of *PLXNA4* in Case/Control Cohort

Linkage analysis highlighted the variant in *PLXNA4* as a potentially causal or modifying variant for the PD phenotype in our family. Also, the affected amino acid in *PLXNA4* is highly conserved in all vertebrates and two of three commonly used prediction algorithms [12–14] predicted it to be "damaging". Accordingly, we screened the 32 coding exons as well as the exon/intron boundaries of *PLXNA4* in 862 Austrian and German cases and 940 controls in order to assess a fuller spectrum of rare genetic variation found. For the most part, this cohort comprised the same individuals used for the above frequency assessment. In *PLXNA4*, a total of 38 novel (37 non-synonymous, 1 deletion) and 6 known variants (rs143813209, rs113830939, rs112682233, rs62622406, rs117458710 and rs73155258, all non-synonymous) resulting in a change in the amino acid sequence were identified (Table S2 in File S1). The large majority (86.21%) of variants were very rare, with MAF≤0.2% in controls. Overall, a similar number of cases (n = 107) and controls (n = 117) harbored at least one variant predicted to result in a changed amino acid sequence ($p>0.05$, χ^2 test). The same held true when only variants with MAF≤1.0% (46 cases vs. 52 controls, $p>0.05$, χ^2 test) were evaluated. Very rare variants with MAF≤0.2%, however, were more common in cases (n = 33) than controls (n = 18) ($p<0.02$, χ^2 test). Three cases but no controls were compound heterozygous for a non-synonymous variant in *PLXNA4*. Variants were located throughout the entire gene (Figure 2A).

Of the individuals harboring a rare non-synonymous variant in *PLXNA4*, information regarding family history was available for 17 individuals: 3 reported a first or second degree relative with PD and a positive history of essential tremor was present in the mother and a maternal uncle in one additional individual. The only brother of the individual harboring the *PLXNA4* p.Arg302Cys amino acid change was also found to have PD and to harbor this variant. However, the family was too small for formal segregation analysis.

When analyzed by means of three commonly used prediction algorithms (PolyPhen2, MutationTaster, SIFT) [12–14], the number of non-synonymous single nucleoide variants (SNVs) classified as functionally "damaging" (SNVs classified as "probably damaging" by PolyPhen2, "disease causing" by MutationTaster and "damaging" by SIFT) was greater in cases than in controls. This was especially prominent and statistically significant for PolyPhen2 when only very rare variants with MAF≤0.2% in controls were analyzed (PolyPhen2:19 variants in cases vs. 9 variants in controls, $p = 0.033$, χ^2 test; MutationTaster: 26 in cases vs. 14 in controls, $p = 0.028$, χ^2 test; SIFT/PROVEAN: 10 in cases vs. 2 in controls, $p = 0.018$, Fisher's Exact test) (Figure 2B). Deletions, which were only found in cases, cannot be assessed by PolyPhen2 and were, therefore, omitted from the analysis using this algorithm.

Functional Assessment of *PLXNA4* p.Ser657Asn in Fibroblasts

In fibroblast cell lines generated from both the index patient and an offspring who does not harbor the *PLXNA4* p.Ser657Asn variant (other variants not given to protect privacy) cell viability was similar (Figure 3A). Based on the results from the above mutation screening as well as the fact that *PLXNA4* is known to be expressed in the brain [15] and a role for axonal guidance factors similar to *PLXNA4* already postulated in PD [16], we further analyzed subcellular localization of the protein in the two cell lines but could not detect a difference (Figure 3B).

Modeling a Potential Role of *PLXNA4* in the PD Network

Beyond a proposed general role of axonal guidance pathways in the development of neurodegeneration [16,17], it is interesting to note that *PLXNA4* can be place into a network containing several firmly established PD genes (*SNCA, PARK2, DJ-1, LRRK2*), although both known and less reliable projected interactions have to be utilized (Figure 4).

Discussion

In an unbiased, whole-exome approach, we identified a variant in *PLXNA4* (p.Ser657Asn) as a candidate for a potentially causal variant in familial PD. Although this finding is intriguing and functionally plausible, we cannot conclude that this variant in *PLXNA4* is indeed the cause of PD in our family. Also, it is interesting that both affected individuals were found to harbor two or three non-synonymous variants in *PLXNA4*, thus, highlighting the possibility that a "multi-hit" model within the same gene or pathway could play a role with regard to phenotype expressivity.

Three of the final four variants (*PLXNA4* p.Ser657Asn, *OGN* p.L124fs and *CPNE1* p.Ser183Thr) are extremely rare and were only found in other family members but not in approximately 8,978 other individuals of European descent (genotyping sample (n = 1989), in-house exomes (n = 1739), 1000genomes (n = 1000) and NHLBI-ESP exomes (n = 4250)). This is interesting in light of the fact that–with regard to drug target genes–it was recently shown that the rarer a given variant the more likely it is

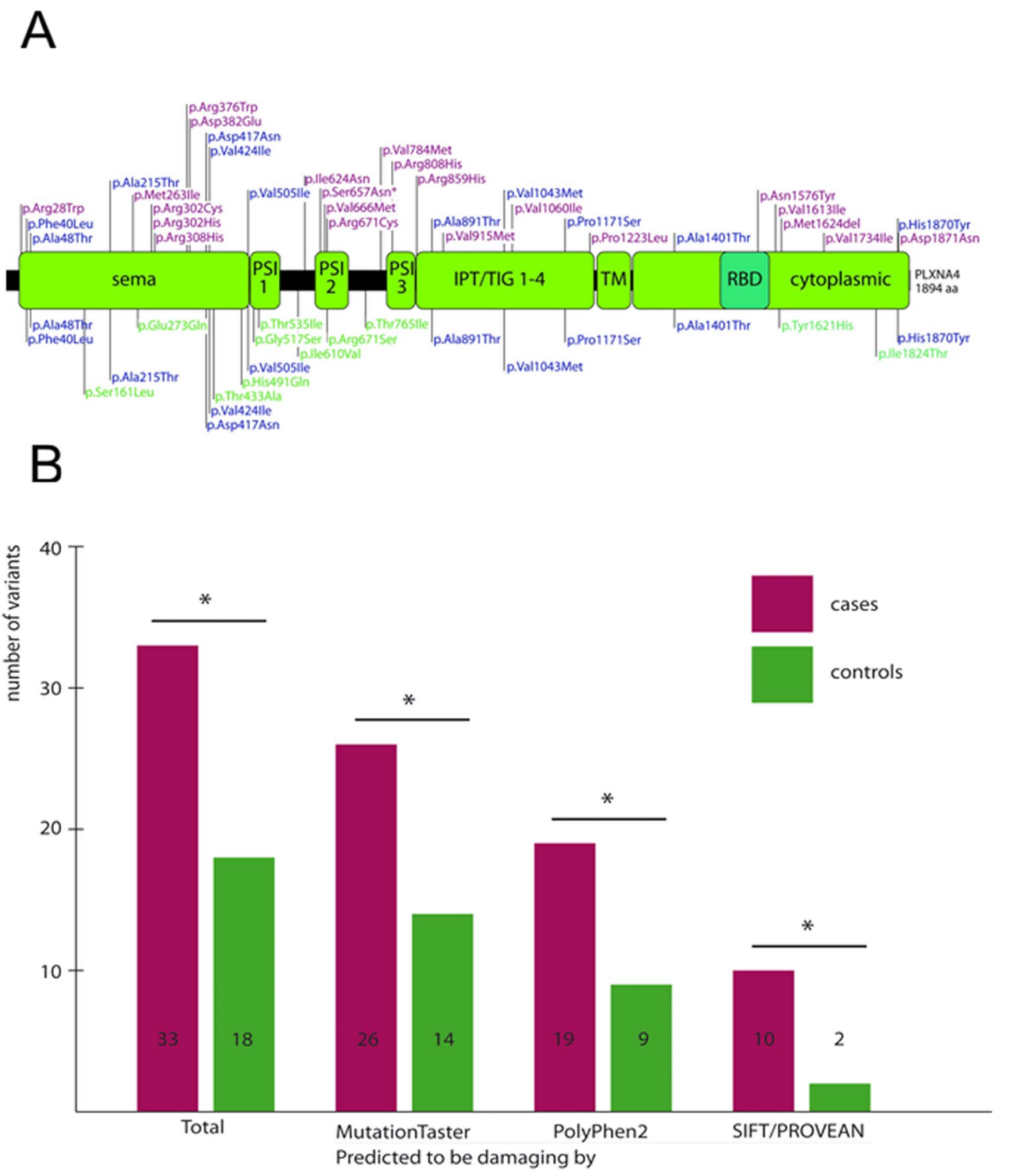


Figure 2. Mutation Screening of *PLXNA4* in PD case/control cohort. (A) Location of *PLXNA4* variants identified in variant screening in relation to known functional domains. An asterisk denotes the variant identified by exome sequencing. blue = variants found in both cases and controls, green = variants found only in cases, purple = variants found only in controls. (B) Analysis of *PLXNA4* variants using SIFT/PROVEAN, PolyPhen2 and MutationTaster reveals an excess of rare non-synonymous variants predicted to be damaging. Insertions and deletions cannot be assessed by PolyPhen2 all and were, therefore, omitted from the analysis using this algorithm.
doi:10.1371/journal.pone.0079145.g002

functionally relevant [18]. Yet, on the other hand, this rarity also means that from a genetic standpoint, at the moment, one can neither confirm nor exclude the possibility of a causal or modifying role in the PD phenotype. Further, even taken together additional evidence highlighting *PLXNA4* p.Ser657Asn (suggestive linkage signal, high conservation and predicted pathogenicity, excess of

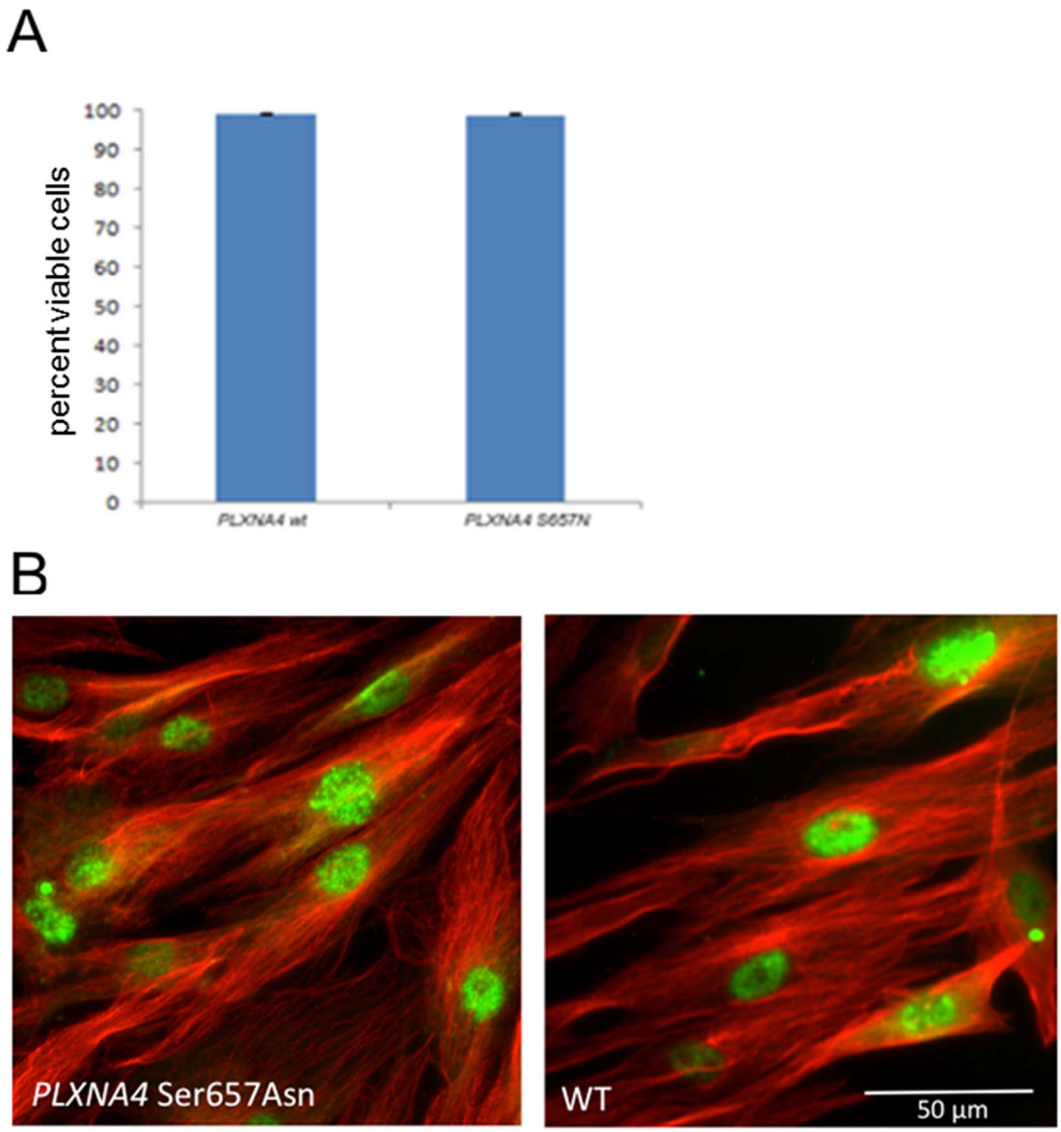


Figure 3. Assessment of cell viability and subcellular protein localization in fibroblasts. (A) The presence of *PLXNA4* p.Ser657Asn do not affect cell viability as assay by live-dead staining and FACS. (B) Immunohistochemistry shows similar subcellular localization of *PLXNA4* (anti-PLXNA4, Sigma, 1:500) in fibroblasts with and without the p.Ser657Asn amino acid substitution (scale bar = 50 µm).
doi:10.1371/journal.pone.0079145.g003

very rare coding variants in cases and functional considerations) can be viewed as suggestive at best and by no means exclude the possibility of other causative or modifying genetic factors that play a role in the PD phenotype in our family.

In general, these findings highlight the fact that in many cases very large populations will be needed to conclusively judge the disease-related nature of a rare variant. Recent studies show that while the power to detect associations for genes harboring rare variants varies widely across genes, only <5% of genes achieved 80% power even assuming high odds ratios (OR) of 5 and when tested in 400 cases and 400 controls. In the same scenario, no gene out of 12,000 genes tested achieved 80% power when assuming an OR of 1.5 [19]. Statistical evaluation is further complicated by the fact that it is not unreasonable to assume that many genes will habor both variants that are protective and predisposing with regard to a given phenotype, as was recently shown for the *APP* locus in Alzheimer's disease [20], which with the statistical analysis tools available today will always lead to an underestimation of the genetic contribution of rare variants at a given locus to a phenotype's heritability [21].

Ultimately, it is also possible that the truly causal variant was not picked up in this study because it lies outside the targeted regions of the exome. Here, the use of two enrichment kits of different sizes and different exome target definitions represents a specific weakness of the study. Also, we cannot exclude that IV:18 represents a phenocopy and that the underlying cause of PD in his case is different from that of the other affected individuals in the family. If this were the case, a much larger number of candidate variants than those assessed here could contribute to bringing about the PD phenotype in the examined family.

Moreover, copy number variants, another important player in the full spectrum of genetic variation, could, at the time of study, not yet confidently be assessed in exome sequencing data and were, therefore, not evaluated in our study. Lastly, while suggestive non-significant LOD scores have been used to prioritize variants identified in exome [22] or whole genome [23] sequencing they

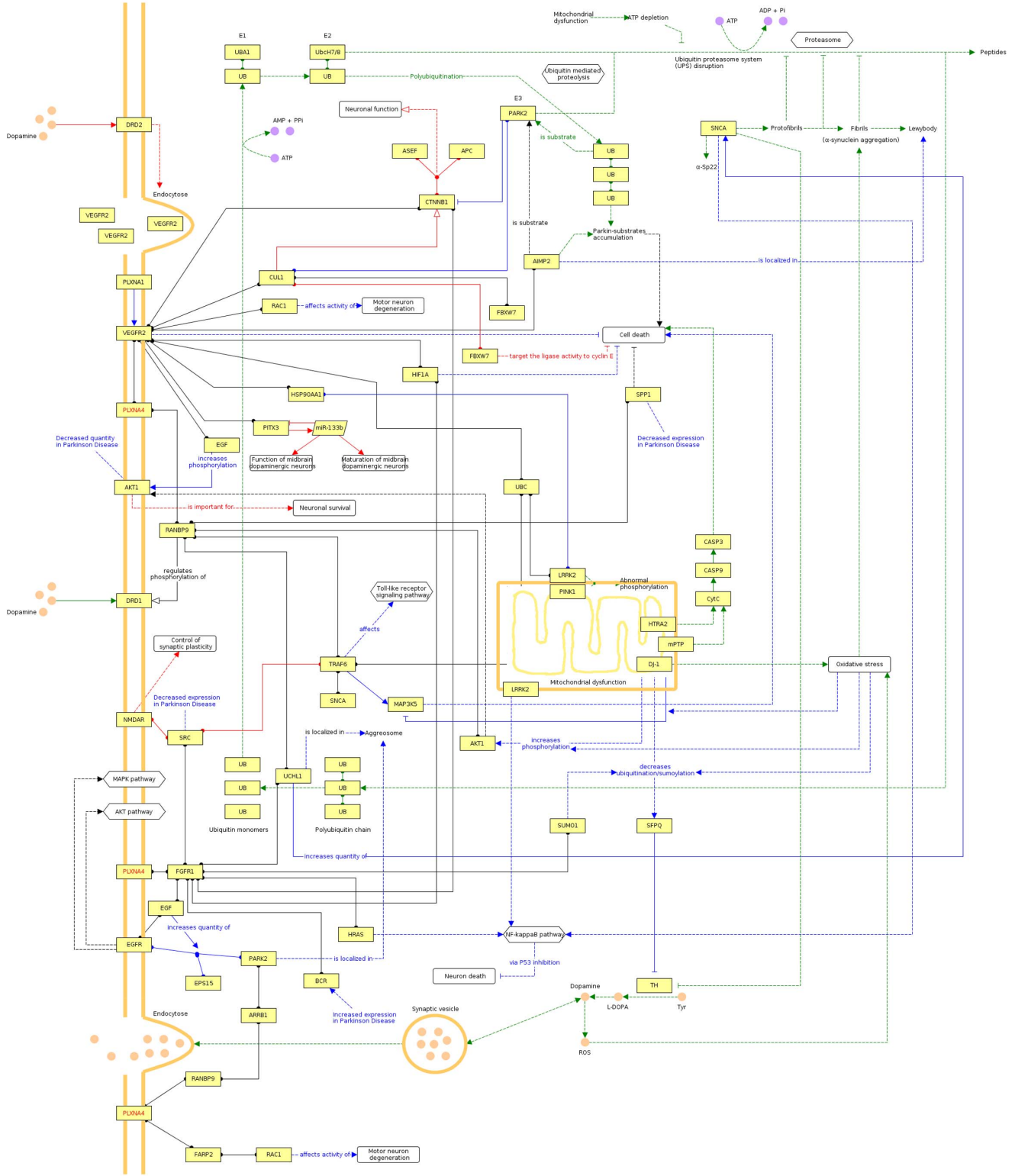


Figure 4. Qualitative multifactorial interaction network of *PLXNA4* and genetic factors with known and hypothetical relevance to PD. Edges obtained from CIDeR are highlighted in blue, PD-specific pathways from KEGG are given in green, red edges denote annotations from OMIM and edges extracted from literature, protein-protein interaction databases or high-confidence predictions are colored black. Undirected protein-protein interactions hold circular ends, directed molecular relations are marked by arcs, whereas general regulations have arrows with no filling, activations have filled arrows and inhibitions have blunted end. Dashed lines indicate indirect effects.
doi:10.1371/journal.pone.0079145.g004

also harbor the potential for the erroneous exclusion of true positives.

The fact that all four candidate variants were also found in unaffected family members, per se does not contradict potential causality as it is known from other autosomal dominant forms of PD that even among members of a single family, penetrance of known PD mutations can vary widely. Of individuals who harbor the *LRRK2* p.Gly2019Ser mutation, for example, only 28% will develop PD by the age of 59 [24]. Thus, predicted penetrance of the variants identified in our family are in line with what is reported in the literature for other forms of autosomal dominant PD.

Plexin A4, *PLXNA4*, which functions as a receptor for class 3 semaphorins, holds a firmly established role in axon guidance in the development of the central and peripheral nervous systems. For example, PlxnA4 has been shown to restrict inappropriate spreading of mossy fibers within the CA3 region of the murine hippocampus [25], to direct basal dendritic arborization in layer V cortical neurons [26] and sympathetic axons [15,27] as well as lamination and synapse formation in the outer retina [28] in the mouse.

PLXNA4 has also been implicated in neurodegenerative conditions. In the discovery stage of a large family-based GWAS assessing low-frequency (MAF≤5%) variants in late-onset Alzheimer's disease an intronic SNP in *PLXNA4* (rs277484, MAF = 2.0% in 1000genomes) yielded the most significant association signal ($p = 9.0 \times 10^{-10}$). Replication, however, is still ongoing [17]. Similarly, preliminary results have suggested decreased *PLXNA4* expression in the motor cortex of individuals with amyotrophic lateral sclerosis when compared to controls, although the sample size of the study was very limited (n = 5) [29].

PLXNA4 itself has not previously been implicated in PD. Yet, a number of studies have suggested an involvement of axonal guidance pathways in PD. An early GWAS identified a SNP in semaphorin 5A (*SEMA5A)* as the best association signal [16] and systems biology-based follow-up studies reported an overrepresentation of axonal guidance factors in subthreshold association signals [30] which were shown to predict susceptibility to PD [31]. However, both the association signal and the pathway analysis proved difficult to replicate in other cohorts [32–34] which may be due to the fact that as one of the very first GWAS it was not conducted to the current quality standards. Expression studies of different brain regions, on the other hand, have repeatedly found an overrepresentation of differentially expressed axonal guidance pathways in individuals with PD when compared to controls [30,35–37]. Axonal guidance pathways have also been implicated in the proper targeting of dopaminergic neurons from the murine mesencephalon to the ipsilateral striatum [38].

At the moment, both functional and genetic data addressing a role of *PLXNA4* as a PD gene are inconclusive. The identification of additional larger families with PD in which *PLXNA4* p.Ser657-Asn or p.Arg302His segregate with the phenotype or the replication of the finding of an excess of very rare variants (MAF≤0.02%) in an independent case/control sample would lend further support to a possible role of modifying or causal variants in *PLXNA4* in PD and to the interesting hypothesis of axonal guidance dysfunction in neurodegenerative conditions.

Supporting Information

Figure S1 Filtering scheme for variants identified by exome sequencing in the two affected family members examined.
(TIF)

File S1 Supporting Methods and Tables. Table S1 in File S1, Clinical Phenotype of Affected Individuals in PARK_0005. **Table S2 in File S1,** Non-Synonymous and Indel Variants Identified in Variant Screening of *PLXNA4*.
(DOC)

Acknowledgments

We are gratefully indebted to Katja Junghans, Susanne Lindhof, Jelena Golic, Sybille Frischholz und Regina Feldmann at the Insitut für Humangenetik, Helmholtz Zentrum München, Munich, Germany, for their expert technical assistance. We would also like to thank Lucas Schirmer, Neurologische Klinik und Poliklinik, Technische Universität München, Munich, Germany, for expert help in performing the FACS analyses.

Author Contributions

Conceived and designed the experiments: ECS DC DCE BMM CT JW. Performed the experiments: ECS IS DC DCE EG. Analyzed the data: ECS IS DC DCE SE PL BMM. Contributed reagents/materials/analysis tools: BM AZ DH WP TB BB MJM AP CG CT. Wrote the paper: ECS IS JW.

References

1. Bonifati V, Fabrizio E, Vanacore N, De Mari M, Meco G (1995) Familial Parkinson's disease: A clinical genetic analysis. Can J Neurol Sci 22: 272–279.
2. Payami H, Larsen K, Bernard S, Nutt J (1994) Increased risk of Parkinson's disease in parents and siblings of patients. Ann Neurol 36: 659–661.
3. Bonifati V, Rizzu P, van Baren MJ, Schaap O, Breedveld GJ, et al. (2003) Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science 299: 256–259.
4. Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, et al. (1998) Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 392: 605–608.
5. Paisan-Ruiz C, Jain S, Evans EW, Gilks WP, Simon J, et al. (2004) Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron 44: 595–600.
6. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, et al. (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 276: 2045–2047.
7. Valente EM, Abou-Sleiman PM, Caputo V, Muqit MM, Harvey K, et al. (2004) Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science 304: 1158–1160.
8. Zimprich A, Biskup S, Leitner P, Lichtner P, Farrer M, et al. (2004) Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 44: 601–607.
9. Vilariño-Güell C, Wider C, Ross OA, Dachsel JC, Kachergus JM, et al. (2011) VPS35 Mutations in Parkinson Disease. Am J Hum Genet 89: 162–167.
10. Zimprich A, Benet-Pagès A, Struhal W, Graf E, Eck SH, et al. (2011) A Mutation in VPS35, Encoding a Subunit of the Retromer Complex, Causes Late-Onset Parkinson Disease. Am J Hum Genet 89: 168–175.
11. Schulte EC, Mollenhauer B, Zimprich A, Bereznai B, Lichtner P, et al. (2012) Variants in eukaryotic translation initiation factor 4G1 in sporadic Parkinson's disease. Neurogenetics 13: 281–285.
12. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 7: 575–576.
13. Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. Nucleic Acid Res 30: 3894–3900.
14. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. (2010) A method and server for predicting damaging missense mutations. Nat Methods 7: 248–249.
15. Suto F, Ito K, Uemura M, Shimizu M, Shinkawa Y, et al. (2005) Plexin-A4 Mediates Axon-Repulsive Activities of Both Secreted and Transmembrane Semaphorins and Plays Roles in Nerve Fiber Guidance. J Neurosci 25: 3628–3637.
16. Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, et al. (2005) High-Resolution Whole-Genome Association Study of Parkinson Disease. Am J Hum Genet 77: 685–693.
17. Jung G, Buros J, Lunetta KL, Foroud TM, Mayeux R, et al. (2011) Identification of Rare Variants for Late-Onset Alzheimer Disease in a Family-Based Genome-Wide Association Study. Abstract No. 221. Presented at the

Annual Meeting of the American Society of Human Genetics 2011, Montreal, Canada.
18. Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, et al. (2012) An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science 337: 100–104.
19. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, et al. (2012) Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. Science 337: 64–69.
20. Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, et al. (2012) A mutation in *APP* protects against Alzheimer's disease and age-related cognitive decline. Nature. 488: 96–99.
21. Liu DJ, Leal SM (2012) Estimating Genetic Effects and Quantifying Missing Heritability Explained by Identified Rare-Variant Associations. Am J Hum Genet 91: 585–596.
22. Johnson JO, Mandrioli J, Benatar M, Abramzon Y, Van Deerlin VM, et al. (2010) Exome sequencing reveals VCP mutations as a cause of familial ALS. Neuron 68: 857–864.
23. Sobreira NL, Cirulli ET, Avramopoulos D, Wohler E, Oswald GL, et al. (2010) Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. PLoS Genet 6: e1000991.
24. Healy DG, Falchi M, O'Sullivan SS, Bonifati V, Durr A, et al. (2008) Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. Lancet Neurol 7: 583–590.
25. Suto F, Tsuboi M, Kamiya H, Mizuno H, Kiyama Y, et al. (2007) Interactions between Plexin-A2, Plexin-A4, and Semaphorin 6A Control Lamina-Restricted Projection of Hippocampal Mossy Fibers. Neuron 53: 535–547.
26. Tran TS, Rubio ME, Clem RL, Johnson D, Case L, et al. (2009) Secreted semaphorins control spine distribution and morphogenesis in the postnatal CNS. Nature 462: 1065–1069.
27. Waimey KE, Huang PH, Chen M, Cheng HJ (2008) Plexin-A3 and Plexin-A4 Restrict the Migration of Sympathetic Neurons but not Their Neural Crest Precursors. Dev Biol 315: 448–458.
28. Matsuoka RL, Jiang Z, Samuels IS, Nguyen-Ba-Charvet KT, Sun LO, et al. (2012) Guidance-cue control of horizontal cell morphology, lamination, and synapse formation in the mammalian outer retina. J Neurosci 32: 6859–6868.
29. Körner S, Wilzek N, Thau S, Knippenberg S, Dengler R, et al. (2012) mRNA Expression von Axon-Leitproteinen (Axon guidance proteins) in humanem post-mortem Motorkortex bei Amyothropher Lateralsklerose. Abstract V24. Presented at the Annual Meeting of the German Society for Neurology 2012, Hamburg, Germany.
30. Edwards YJ, Beecham GW, Scott WK, Khuri S, Bademci G, et al. (2011) Identifying Consensus Disease Pathways in Parkinson's Disease Using an Integrative Systems Biology Approach. PLoS one 6: e16917.
31. Lesnick TG, Papapetropoulos S, Mash DC, Ffrench-Mullen J, Shehadeh L, et al. (2007) A Genomic Pathway Approach to a Complex Disease: Axon Guidance and Parkinson Disease. PLoS Genet 3: e98.
32. Bialecka M, Kurzawski M, Klodowska-Duda G, Opala G, Tan EK, et al. (2006) Polymorphism in semaphoring 5A (Sema5A) gene is not a marker of Parkinson's disease risk. Neurosci Lett 399: 121–123.
33. Clarimon J, Scholz S, Fung HC, Hardy J, Eerola J et al. (2006) Conflicting results regarding the semaphorin gene (SEMA5A) and the risk for Parkinson disease. Am J Hum Genet 78: 1082–1084.
34. Li Y, Rowland C, Xiromerisiou G, Lagier RJ, Schrodi SJ, et al. (2008) Neither replication nor simulation supports a role for the axon guidance pathway in the genetics of Parkinson's disease. PLoS One 3: e2707.
35. Bossers K, Heerhoff G, Balesar R, van Dongen JW, Kruse CG, et al. (2009) Analysis of gene expression in Parkinson's disease: Possible involvement of neurotrophic support and axon guidance in dopaminergic cell death. Brain Pathol 19: 91–107.
36. Sutherland GT, Matigian NA, Chalk AM, Anderson MJ, Silburn PA, et al. (2009) A cross-study transcriptional analysis of Parkinson's disease. PLoS one 4: e4955.
37. Wang K, Li M, Bucan M (2007) Pathway-based approaches for analysis of genomewide association studies. Am J Hum Genet 81: 1278–1283.
38. Kawano H, Horie M, Honma S, Kawamura K, Takeuchi K, et al. (2003) Aberrant trajectory of ascending dopaminergic pathway in mice lacking Nkx2.1. Exp Neurol 182: 103–112.
39. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA. Available: http://evs.gs.washington.edu/EVS/. Accessed October 23, 2012.

BRIEF REPORT

Mitochondrial Membrane Protein Associated Neurodegenration: A Novel Variant of Neurodegeneration With Brain Iron Accumulation

Eva C. Schulte, MD,[1,2,3] Malte C. Claussen, MD,[1] Angela Jochim, MD,[1] Tobias Haack, MD,[2,3] Monika Hartig, MD,[2,3] Maja Hempel, MD,[2,3] Holger Prokisch, PhD,[2,3] Ursula Haun-Jünger, MD,[4] Juliane Winkelmann, MD,[1,2,3] Bernhard Hemmer, MD,[1] Annette Förschler, MD,[5] and Rüdiger Ilg, MD[1]*

[1]*Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany;* [2]*Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany;* [3]*Institut für Humangenetik, Klinikum recht der Isar, Technische Universität München, Munich, Germany;* [4]*Sozialpädiatrisches Zentrum, Klinikum Bremen-Mitte, Bremen, Germany;* [5]*Abteilung für Neuroradiologie, Klinikum rechts der Isar, Technische Universität München, Munich, Germany*

ABSTRACT

Background: Recently, mutations in an open-reading frame on chromosome 19 (*C19orf12*) were identified as a novel genetic factor in neurodegeneration with brain iron accumulation (NBIA). Because of the mitochondrial localization of the derived protein, this variant is referred to as mitochondrial membrane protein-associated neurodegeneration with brain iron accumulation (MPAN).

Methods/Results: We describe the clinical phenotype and MRI of 3 newly identified individuals with MPAN due to either previously reported or novel homozygous or compound heterozygous genetic alterations in *C19orf12*.

Conclusions: MPAN is characterized by a juvenile-onset, slowly progressive phenotype with predominant lower limb spasticity, generalized dystonia, and cognitive impairment. Typical additional features include axonal motor neuropathy and atrophy of the optic nerve. MRI showed iron deposition in the globus pallidus and substantia nigra without the eye-of-the-tiger sign, which is typical for PKAN, the most frequent form of NBIA. © 2012 *Movement* Disorder Society

Key Words: neurodegeneration with brain iron accumulation (NBIA); genetics; MRI; mitochondrial membrane protein-associated neurodegeneration (MPAN); *C19orf12*

We describe 3 cases of mitochondrial membrane protein-associated neurodegeneration (MPAN), a novel form of neurodegeneration with brain iron accumulation (NBIA, formerly known as Hallervorden–Spatz syndrome). NBIA subsumes a group of hereditary neurodegenerative disorders characterized by iron accumulation primarily in the basal ganglia, presenting with a variety of neurologic changes with prominent progressive motor and behavioral symptoms.[1,2] Although very rare (1–3 cases/1 million population), significant progress has been made in stratifying this heterogeneous group into genetic and phenotypic entities. Accounting for up to 50% of cases, pantothenate kinase–associated neurodegeneration (PKAN), attributed to autosomal recessive mutations in the *PANK2* gene, is the most common form.[3,4] Other NBIA variants include infantile and noninfantile neuraxonal dystrophy (collectively also termed *PLA2G6*-associated neurodegeneration [PLAN]),[5,6] static encephalopathy of childhood with neurodegeneration in adulthood, fatty acid hydroxylase-associated neurodegeneration, Kufor–Rakeb syndrome, and Woodhouse–Sakati syndrome.[2,5,6] Recently, an open-reading frame on chromosome 19, *C19orf1*, was identified as a novel genetic factor in NBIA.[7] In a cohort of 52 NBIA patients from Poland, deletions and missense mutations were found in 19, rendering *C19orf12* the second most common genetic locus affected in NBIA.[7] Both genetically and clinically, NBIA ascribed to

---

Additional Supporting Information may be found in the online version of this article.

Eva C. Schulte and Malte C. Claussen contributed equally to the article.

***Correspondence to:** Rüdiger Ilg, Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Ismaningerstr. 22, 81675 Munich, Germany; ilg@lrz.tum.de

Funding agencies: This study was financed exclusively by in-house funding from the Neurologische Klinik & Poliklinik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany.

Relevant conflicts of interest/financial disclosures: Holger Prokisch receives research support from the Impulse and Networking Fund of the Helmholtz Association within the framework of the Helmholtz Alliance for Mental Health in an Ageing Society (HA-215) and the German Federal Ministry of Education and Research (BMBF)–funded German Center for Diabetes Research (DZD e.V.), Systems Biology of Metabotypes grant (SysMBo 0315494A), and the German Network for Mitochondrial Disorders (mitoNET 01GM0867). Rüdiger Ilg is supported by the medical faculty of the Technische Universität München (KKF).

Full financial disclosures and author roles may be found in the online version of this article.

Received: 25 January 2012; **Revised:** 10 September 2012; **Accepted:** 25 September 2012

Published online in Wiley Online Library (wileyonlinelibrary.com).

DOI: 10.1002/mds.25256

Table 1. Summary of clinical, genetic, and MRI features observed in 3 individuals suffering from MPAN

	MPAN_1	MPAN_2	MPAN_3
Sex	F	M	M
Age of onset	7	3	6
Age at presentation	34	19	18
C19orf12 mutation	c.32 C>T (p.Thr11Met) c.204_214del (p.Gly69ArgfsX10)	c.53 A>G (p.Asp18Gly) c.395 T>A (p.Leu132Gln)	c.204_214del (p.Gly69ArgfsX10) c.204_214del (p.Gly69ArgfsX10)
Pyramidal signs	+	++	++
Spasticity	+	++	+
Oromandibular dystonia	−	+	−
Generalized dystonia	+	+	+
Parkinsonism	−	−	−
Dysarthria	No spontaneous speech	+	+
Optic atrophy	−	++	(+)
Eye movements	Unremarkable	Not assessable	Saccadic pursuit
Motor axonal neuropathy	++	++	−
Psychiatric signs	+	++	−
Cognitive impairment	++	++	+
Wheelchair bound (age)		13	
Additional clinical features	Low-frequency hand tremor	None	Strabism
Serum creatine kinase (U/L)	115 (normal < 174)	589 (normal < 174)	143 (normal < 174)
MRI findings	Bilateral T2 hypointensity in GP and SN No eye-of-the-tiger sign T1 hyperintensity in CN and Put	Bilateral T2 hypointensity in GP and SN No eye-of-the-tiger sign T1 hyperintensity in CN and Put	Bilateral T2 hypointensity in GP & SN, No eye-of-the-tiger sign T1 hyperintensity in CN and Put
Family history	1 Unaffected sister No neurologic/psychiatric diseases	No siblings No neurologic/psychiatric diseases	2 Unaffected brothers No neurologic/psychiatric diseases
Ethnicity	German from Russia	Turkish	Polish

GP, globus pallidus; SN, substantia nigra; CN, caudate nucleus; Put, putamen.

variants in *C19orf12* represents a distinct group that has been termed MPAN.[7] Based on this discovery, we identified 3 previously unreported individuals with MPAN who had so far been classified as idiopathic forms of NBIA. We describe the phenotype and MRI that, according to the current state of knowledge, is believed to be characteristic of this subgroup of NBIA.

MPAN

All 3 individuals visited our department for routine treatment reevaluation of a slowly progressive psychomotor disorder. As required by clinical regulations, all provided written informed consent for genetic testing. The 3 individuals presented with progressive psychomotor deterioration (gait impairment, clumsiness, dysarthria, and progressive cognitive deficits). Initial symptoms consisted of general clumsiness along with retarded motor and scholastic development. During the course of the disease, gait disturbance from severe spasticity of the lower limbs became most prominent. Pyramidal signs (ie, upgoing plantars) were found in all 3 individuals. In contrast to PKAN, generalized dystonia was less pronounced and parkinsonism, which can occur in PKAN, was not present. Disease progression in 2 individuals was comparatively slow, with preserved ambulation at 10 and 25 years after disease onset, whereas the third individual was wheelchair bound by age 13. Two individuals showed impulsive behavior. Retinitis pigmentosa, frequently observed in PKAN, was not found in any of our individuals. Instead, bilateral atrophy of the optic nerve was present in 2 of the 3 cases (Table 1). Similar to a significant part of the Polish MPAN cohort,[7] in 2 of the cases, nerve conduction studies revealed prominent generalized axonal motor neuropathy, a feature not typically seen in PKAN. Additional clinical features in single individuals included low-frequency bilateral action tremor of the hands and congenital strabism. Increased serum creatine kinase (CK) levels have been reported in individuals with MPAN.[8] In our cases, only MPAN_2 had an elevated serum CK (584 U/L; normal < 174 U/L). We also noted slightly increased serum lactate dehydrogenase levels in all 3 individuals (MPAN_1, 369 U/L; MPAN_2, 270 U/L; and MPAN_3, 289 U/L; normal < 244 U/L), whereas all other routine blood tests were unremarkable. Electroencephalography was normal in all cases. MRI showed prominent T2 hypointensity of the globus pallidus highly suggestive of iron deposition[6] in all 3 MPAN subjects. In contrast to the PKAN subtype, where the so-called eye-of-the-tiger sign, a central T2

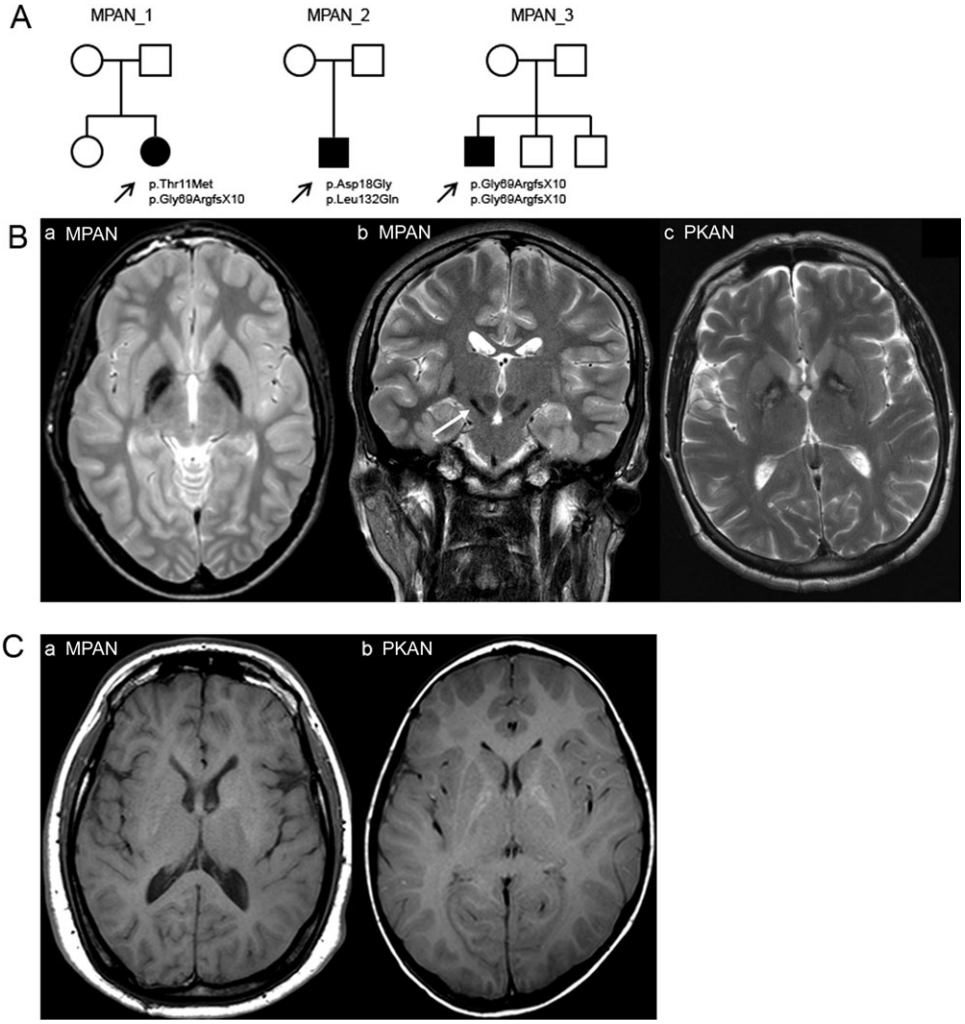


FIG. 1. Three cases of MPAN, a novel subtype of NBIA. **A**: Pedigrees of families of MPAN_1 to MPAN_3 (open symbols, unaffected family members; closed symbols, affected individuals; squares, males; circles, females; arrow, proband, the only individual for whom genotypic information is available). **B**: Axial 3-D T2- (a) and coronal T2-weighted (b) MRI (3 Tesla) of individual MPAN_3 and T2-weighted image of an individual with PKAN in axial view (c). Signal loss indicates iron deposition in the globus pallidus (a) and substantia nigra (white arrow in b), but no eye-of-the-tiger sign. Within the globus pallidus the small layer of white matter that separates the globus pallidus internus and externus becomes visible. In contrast, the image of a PKAN patient (c) displays the typical central hyperintensity within the hypointense globus pallidus, which constitutes the pupil of the eye of the tiger. **C**: Axial T1-weighted images of MPAN (a) and PKAN (b) at the level of the basal ganglia. In MPAN, increased T1 signal of the caudate nucleus and the putamen leads to harmonization of the basal ganglia with the surrounding white matter. In PKAN, the caudate nucleus and the putamen display normal gray matter signal. Here, only the globus pallidus is hyperintense, probably as a result of iron deposition.

hyperintense area within the T2 hypointense (iron-rich) globus pallidus, is nearly pathognomonic, the eye-of-the-tiger sign was not seen in any of our individuals. In contrast to other subtypes of NBIA, all individuals showed T2 hypointensity of the substantia nigra (Fig. 1) and additional T1 hyperintensity of the caudate nucleus and putamen. In the most severely affected individual, additional frontotemporal atrophy was present.

Genetic evaluation of the coding regions of *C19orf12* in the affected individuals only was performed by Sanger sequencing. One individual (MPAN_3) was homozygous for an 11-bp deletion (c.204_214del [p.Gly69ArgfsX10]), which is predicted to result in a truncated protein and which is the most frequent variant in *C19orf12* observed in MPAN to date.[7] MPAN_1 also carried this deletion on 1 allele and a missense variant (c.32 C>T [p.Thr11Met]) on the other. The p.Thr11Met missense variant has previously been described in 1 homozygous and 1 compound heterozygous individual with NBIA.[7] MPAN_2 was compound heterozygous for 2 novel missense variants (c.53 A>G [p.Asp18Gly] and c.395 T>A [p.Leu132Gln]; Table 1). No additional variants were found in *PANK2* (screened in all) or *PLA2G6* (in MPAN_1 and MPAN_2 only). Family history for neurologic or psychiatric diseases was negative in all individuals. Parental consanguinity was not reported in any of the 3 cases. Siblings and parents, although likely heterozygous carriers, were phenotypically unaffected, supporting an autosomal recessive mode of inheritance (Fig. 1).

Discussion

Description of *C19orf12* as a novel NBIA locus[7] has led to the identification of a distinct form of NBIA termed MPAN. Mitochondrial localization and analysis of genes coregulated with *C19orf12* have indicated mitochondrial dysfunction and altered lipid metabolism in NBIA pathogenesis.[7] Findings in our 3 individuals support recent data from a Polish cohort[7] suggesting that MPAN could account for a significant portion of NBIA formerly classified as idiopathic.

Although a number of clinical features are shared between PKAN, MPAN, and other variants of NBIA, there are several important differences. Compared with PKAN, onset is later and psychomotor regression more gradual. A slowly progressive gait disorder from generalized dystonia and spastic paraparesis and cognitive impairment constitute the main features of MPAN. The majority of individuals reported so far show additional optic nerve atrophy, and about half also have motor axonal neuropathy.[7,8] The eye-of-the-tiger sign, pathognomonic for PKAN, was not found in our MPAN individuals and was only described in 1 individual from the recently published Polish cohort.[7] Instead, all 3 individuals showed additional prominent nigral T2 hypointensity as a sign of iron deposition also in the substantia nigra (Fig. 1). This pattern is similar to subjects with PLAN,[6] who also show optic atrophy in some cases.[9] In contrast to PLAN, our subjects showed no cerebellar atrophy but additional T1 hyperintensity of the caudate nucleus and putamen. Depending on the tissue concentration, iron can either be hyper- or isointense in T1.[6] The combination of T1 hyperintensity and normal gray matter signal in T2 suggests either a discrepancy between paramagnetic effects on T1 and T2 as reported for manganese or could be a result of a low iron concentration, which has not yet led to signal extinction in the T2 sequence.[10]

Further instrumental diagnostic and genetic examination of the parents would be interesting to evaluate whether a heterozygous carrier status coincides with mild disease features such as asymptomatic polyneuropathy, signs of subclinical brain iron deposition, or decreased visual acuity. From a genetics perspective, it is noteworthy that MPAN_3, who is of Polish origin, was homozygous for *C19orf12* p.Gly69ArgfsX10, the original and most common genetic variant identified in the previously reported Polish NBIA cohort. It may be possible that this variant arose in Eastern Europe, whereas in other regions such as, for example, Turkey, from where both parents of MPAN_2 originate, other genetic variants of *C19orf12* may have arisen independently. This conception is further supported by a recent report of 2 Turkish brothers with an NBIA phenotype similar to the 1 of our 3 individuals (spasticity, dystonia, optic atrophy, and peripheral neuropathy) who harbor a novel homozygous missense mutation (c.362 T>A, p.Leu121Gln) in *C19orf12*.[8] Interestingly, the genetic variant found in MPAN_2 (p.Leu132Gln) is close to the p.Leu121Gln mutation and a previously identified missense mutation (p.Lys142Gln), but not in the predicted transmembrane domain where previously reported MPAN mutations are found. Clinically, MPAN_2 was most severely affected with the earliest disease onset and the most rapid progression. One possible explanation is that the variants identified in MPAN_2 harbor the most deleterious effect on MPAN-relevant *C19orf12* function or the presence of additional modifying genetic factors.

So far, the clinical management of NBIA has been limited to symptomatic treatment of motor symptoms including bilateral pallidal deep brain stimulation in cases with severe generalized dystonia.[11] In PKAN, the CoA precursors calcium panthothenate and panthetine are discussed as potential therapeutic agents,[12] and results of a small phase II pilot study indicate possible reduction of pallidal iron deposition by the iron-chelating agent deferiprone. The clinical benefit here is unclear, however.[13] Moreover, it remains to be elucidated whether PKAN and MPAN share a common disease mechanism or merely have enhanced iron deposition in the brain as a unifying feature.

At present, it is important to realize that iron deposition in the basal ganglia and the substantia nigra without the characteristic eye-of-the-tiger sign alongside relatively late psychomotor regression with prominent spasticity of the lower limbs, atrophy of the optic nerve, and sometimes severe motor axonal neuropathy should raise suspicion for MPAN, and *C19orf12* should be assessed for genetic alterations.

Video. The MPAN phenotype. Videos of MPAN_2 showing generalized action-induced dystonia predominantly of the upper extremities, together with retro- and torticollis and oromandibular dystonia, spasticity of the lower limbs with upgoing plantar reflexes (Babinski sign)—note that at the time of videotaping, spasticity of the lower extremities was relatively mild under therapy with baclofen 62.5 mg/day)—and interdigital muscular atrophy caused by axonal motor polyneuropathy.

Acknowledgment: We are greatly indebted to the 3 individuals who participated in this study and their families. We thank Angelika Klucken and her colleagues at Hoffnungsbaum e.V. for the good collaboration and their great commitment in supporting individuals with NBIA.

References

1. Hayflick SJ, Westaway SK, Levinson B, et al. Genetic, clinical, and radiographic delineation of Hallervorden-Spatz syndrome. N Engl J Med 2003;348:33–40.
2. Gregory A, Polster BJ, Hayflick SJ. Clinical and genetic delineation of neurodegeneration with brain iron accumulation. J Med Genet 2009;46:73–80.

3.1.2 Rare variants and common variants in the same place

Schulte EC, Knauf F, Kemlink D, Schormair B, Gieger C, Lichtner P, Meitinger T, Winkelmann J. Variant Screening of the Coding Regions of MEIS1 in Patients with Restless Legs Syndrome. Neurology 2011,76:1106-1108. PMID: 21422461. **(IF 2011: 8.3)**

Schulte EC*, Kousi M*, Tan P, Schormair B, Knauf F, Lichtner P, Trenkwalder C, Högl B, Frauscher B, Berger K, Fietze I, Gross N, Stiasny-Kolster K, Oertel W, Bachmann C.G, Paulus W, Zimprich A, Peters A, Gieger C, Meitinger T, Müller-Myshok B, Katsanis N, Winkelmann J. Targeted Resequencing and Systematic In Vivo Functional Testing Identifies Rare Variants in MEIS1 as Significant Contributors to Restless Legs Syndrome. Am J Hum Genet 2014, 95:85-95. PMID: 24995868 **(IF 2014: 10.9)**

In line with the "common disease, common variant" paradigm, GWAS have been extremely successful in identifying common genetic risk variants for a wide range of common, genetically complex traits and disorders. However, to date, they have—for nearly all traits examined—not been able to uncover the entire underlying heritability. The portion of heritability of common, complex traits that has not been revealed by GWAS has frequently been termed the "missing heritability"(Maher, 2008). Under the "common disease, rare variant" hypothesis, it has been proposed that at least some of this missing heritability could lie in a collection of rare variants that also contribute to common disorders(Cirulli and Goldstein, 2010; Gibson, 2012).

Two projects have tested the "common disease, rare variant" hypothesis for RLS. One possibility is that such rare variants in common, complex neuropsychiatric conditions could—as has been illustrated for non-neuropsychiatric conditions—lie in regions of the genome that also contain common genetic variants associated with the same disorder (e.g.(Bonnefond et al., 2012; Rivas et al., 2011)). Prior to the initiation of the projects depicted here, GWAS had described a total of six genomic loci harboring seven genes with common variants that increase an individual's risk for RLS(Schormair et al., 2008; Winkelmann et al., 2007). Accordingly, this was the starting point of the investigations highlighted here. The most significantly associated SNP for RLS was found in an intron of the homeobox transcription factor *MEIS1*. In a first step, the coding regions plus the exon/intron boundaries (±10 bp) of *MEIS1* were screened for rare variants using Idaho® Light Scanner high-resolution melting curve analysis followed by Sanger sequencing in a set of 188 RLS cases and 188 population controls belonging to the KORA cohort. The identified rare genetic variants were further

evaluated using Sequenom® MALDI-TOF mass spectrometry in a larger set of 1014 patients and 735 controls. In doing so, we were able to replicate a rare, non-synonymous coding variant in exon 8 of *MEIS1* (p.Arg272His)(Schulte et al., 2011), which had previously been described in a Canadian RLS family(Vilarino-Guell et al., 2009), with reduced penetrance. Due to the relatively small sample size, the other detected rare variants could not be linked conclusively to the RLS phenotype.

To more definitively answer the question in how far rare genetic variants in and around coding regions of genes within genomic loci known to harbor common risk variants identified by GWAS contribute to the genetic architecture of RLS, the project depicted above was starkly increased in size. For one, all six additional genes within the loci known at the time—namely, *PTPRD*, *TOX3*, *BTBD9*, *MAP2K5*, *SKOR1*, and the non-coding RNA *BC034767*—were screened for rare variants as depicted above and subsequently genotyped in 3262 case individuals and 2944 control individuals. In the screening analysis, 49 variants with MAF < 5% were identified across all cases and controls. Combined across all seven genes, rare coding variants were significantly more common in individuals with RLS than in controls (77 case subjects vs. 46 control subjects; p=0.023, logistic regression meta-analysis; OR=1.51), independent of the common risk genotypes. When tested in the larger case-control sample, the same picture emerged. There was a distinct excess of very rare variants with a KORA-derived MAF < 0.1% in individuals with RLS (total: 57 case versus 16 control subjects; $p=4.99 \times 10^{-4}$; OR=2.50; non-synonymous variants only: 23 case vs. 5 control subjects, p=0.0019; OR=3.92; logistic regression). Secondly, as the signal detected above appeared to be most pronounced for *MEIS1*, we expanded the high-resolution melting curve discovery sample from 188 cases and 188 controls to 3,760 cases and 3,542 controls. Stratification of variants with MAF < 5% across different regions of the gene revealed an enrichment in the 5'UTR (16 case subjects vs. 2 control subjects; p=0.001; logistic regression; OR=7.56) as well as among non-synonymous variants in isoform 2 (34 case subjects vs. 15 control subjects; p=0.007; logistic regression; OR=2.31).Comparable results were obtained using other burden statistics like the sequence kernel association test (SKAT) or SKAT with Madsen-Browning weights(Madsen and Browning, 2009; Wu et al., 2011) which attempt to account for differing allele frequencies and differing directions of effect among the identified rare variants.

In order to more fully understand the functional implications of rare non-synonymous genetic variants in *MEIS1* on the RLS phenotype, we subsequently assessed all 17

identified non-synonymous variants by *in vivo* complementation in zebrafish embryos. Two distinct neurodevelopmental phenotypes (size of the optic tectum and rhombomere patterning), which had previously been linked to a loss of *meis1* function in zebrafish(Erickson et al., 2010; Waskiewicz et al., 2001), were generated using two different splice-blocking morpholinos (i.e. anti-sense RNAs) and rescued using human *MEIS1* mRNA harboring the different rare non-synonymous variants. mRNAs containing the different variants were either able to completely (benign allele) or incompletely (hypomorphic allele) rescue the phenotype or had no rescuing effect at all (functional null allele). Layering these data over the incidence and distribution of these variants in our case/control dataset revealed an excess of null alleles in individuals with RLS (14 in case subjects versus 2 in control subjects; p=0.0012; OR=7.48). Importantly, these null alleles compromised the function of the canonical *MEIS1* isoform but not that of a second isoform known to contain alternative starting and 3' sequences. Benign and hypomorphic alleles were distributed evenly among cases and controls.

The investigations depicted above established a significant overall enrichment of rare variants in genes within the GWAS loci in individuals with RLS that was most pronounced for very rare variants and for variants in *MEIS1*. Further, these investigations highlighted the potential of combined sequencing plus systematic functional annotation of rare variation at GWAS loci to detect the true risk burden by increasing both effect size and significance levels observed in burden testing. They also highlight the concept of allelic series—a series in which different variants within a single gene have different effects on a given phenotype—in the context of common, complex neuropsychiatric disorders. The isoform specificity of the RLS-linked variants is intriguing because it holds great promise to better understanding the pathophysiology underlying the disorder by uncovering the spatial and temporal expression pattern of this isoform.

Clinical/Scientific Notes

E.C. Schulte, MD
F. Knauf, MSc
D. Kemlink, MD
B. Schormair, PhD
P. Lichtner, PhD
C. Gieger, PhD
T. Meitinger, MD
J. Winkelmann, MD

VARIANT SCREENING OF THE CODING REGIONS OF *MEIS1* IN PATIENTS WITH RESTLESS LEGS SYNDROME

Restless legs syndrome (RLS) is a common and genetically complex neurologic disease presenting with an urge to move the legs and dysesthesias in the evening and at times of rest. Genome-wide association studies have linked single nucleotide polymorphisms in *MEIS1* and 3 other loci to an increased susceptibility to RLS.[1-3] However, to date, only one potentially causal variant has been reported.[4] Therefore, we screened the coding regions and exon-intron boundaries of *MEIS1* for variants, which by exerting a strong phenotypic effect could provide a basis for assessing the function of the gene in RLS.

Methods. Using Idaho LightScanner high-resolution melting curve analysis, we screened DNA of 188 patients with RLS of a first discovery sample (72.8% female, mean age 60.0 ± 11.2 years), all harboring RLS risk alleles of *MEIS1* (G/T or G/G for rs2300478), for aberrant melting patterns (e-Methods on the *Neurology*® Web site at www.neurology.org). Exons showing changes suggestive of variants were sequenced on an ABI Prism 3730 sequencer, and variants identified were subsequently genotyped in an independent German sample (henceforth termed "second sample") consisting of 735 patients with RLS (70.8% female, mean age 61.5 ± 14.2 years) and 735 unrelated control subjects (74.5% female, mean age 59.8 ± 11.3 years) by matrix-assisted laser desorption ionization/time-of-flight mass spectrometry. Disease segregation was evaluated in one family with the p.R272H mutation of exon 8 of *MEIS1*, previously related to RLS.[4] RLS in all patients was diagnosed in accordance with standard diagnostic criteria[5] (e-Methods).

Standard protocol approvals, registrations, and patient consents. Ethics review board approval and participants' written informed consent were obtained.

Results. In the discovery sample, we identified 3 novel nonsynonymous nucleotide substitutions in exons 3 (p.H81Q) and 6 (p.S204T) and in one transcript containing exon 13 (p.M453T) of *MEIS1* in one patient each. In addition, 2 patients showed the p.R272H variant.

Genotyping in the second sample revealed p.H81Q in 2 control subjects and p.M453T in 2 case patients and 3 control subjects, whereas p.S204T was observed in one case patient. p.R272H was not found in any additional case patients or control subjects (figure, A).

Segregation analysis performed on 7 family members revealed 4 affected individuals who presented the p.R272H variant (figure, B). One family member (III-1) and one married-in individual (III-3) were affected but did not show the p.R272H variant. p.R272H was not found in one unaffected individual (II-1).

Because all p.R272H patients were of Czech heritage, we further evaluated the presence of a putative p.R272H founder mutation in 279 Czech patients with RLS (63.1% female, mean age 55.8 ± 14.9 years), but did not find any additional carriers.

Discussion. Screening of the coding regions of *MEIS1* in patients with RLS revealed 3 novel variants. In all cases, patients reported one or more relatives with at least a suspicion of RLS; however, families were small and individuals in many instances not available for further study so that pathogenicity of the variants was not evaluated. We also confirmed the p.R272H variant of *MEIS1*[4] in 2 patients with RLS, one belonging to a family in which the RLS trait seems to be inherited in an autosomal dominant fashion. As observed previously, the p.R272H variant is located within the highly conserved TALE homeobox domain, which is essential for dimerization and transcription activation and disruption of which is known to be detrimental.[4] Therefore, this variant is considered the most likely candidate for an RLS-linked pathogenic mutation to date. In our sample, it was only present in RLS-affected individuals. However, segregation analysis also revealed affected individuals without the variant, suggesting that because RLS is a common disease, these cases could represent phenocopies, that is, similar phenotypes due to different genetic alterations. As opposed to the phenotype described in the North American p.R272H family, disease representation was rather homogeneous in our family with severe and early onset of symptoms. The second patient with p.R272H reported only a daughter

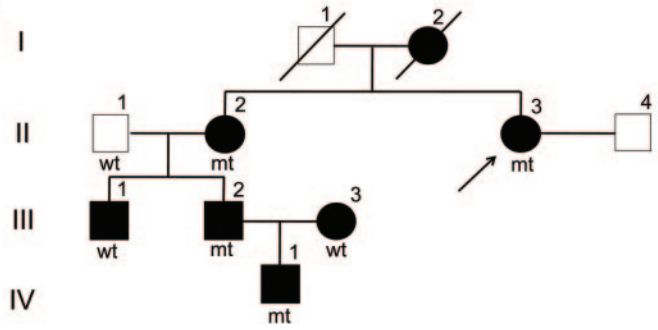
Supplemental data at www.neurology.org

Figure Variants of the coding regions of *MEIS1*

A

Region		Genomic position (chr. 2; NCBI 36/hg18)	Nucleotide/amino acid substitution		First sample	Second sample		
						Cases	Controls	MAF
MEIS1 exon 3	novel	66,520,482	CA[C/G] - His>Gln	H81Q	1/186	0/0/722	0/2/717	0.001
MEIS1 exon 6	novel	66,523,664	[T/A]CA - Ser>Thr	S204T	1/188	0/1/723	0/0/723	<0.001
MEIS1 exon 8	Ref. (4)	66,592,857	C[G/A]T - Arg>His	R272H	2/365	0/1/723	0/0/725	<0.001
MEIS1 exon 13	novel	66,652,131	A[T/C]G - Met>Thr	M453T	1/178	0/2/716	0/3/722	0.002

B



(A) Table illustrating the nucleotide and amino acid substitutions of novel coding variants and p.R272H of *MEIS1* as well as their frequency in the first (mutation carriers/total number of patients with restless legs syndrome [RLS] tested) and the second sample (homozygotes/heterozygotes/homozygotes in case patients and control subjects). (B) Segregation analysis of the p.R272H variant of *MEIS1* in an RLS family. Men are represented by squares and women by circles; a diagonal line indicates a deceased individual. For individuals sequenced, mt indicates those carrying the p.R272H variant and wt indicates noncarriers. The arrow denotes the index patient. MAF = minor allele frequency.

with possible RLS during pregnancy, which could be suggestive of variable expressivity of the p.R272H variant as well as the presence of additional modifying factors. Overall, segregation analyses remain inconclusive because of the small size of pedigrees, yet support the notion that p.R272H *MEIS1* could be causally related to RLS.

All additional variants affect amino acids highly conserved in vertebrates. Although not located within a known functional protein domain, bioinformatics algorithms[6] predict p.S204T and p.H81Q to be disease-causing, whereas p.M453T is likely to be functionally neutral. However, the fact that, as opposed to p.H81Q and p.M453T, p.S204T was only found in RLS-affected individuals renders this variant a second potential candidate for a disease-causing genetic alteration.

However, one limitation of this study is the fact that the control subjects used are general population control subjects and thus we cannot exclude the possibility that variants also found in control subjects (p.H81Q and p.M453T) are not related to RLS. Accordingly, these warrant further replication in an independent dataset.

Our results show that exonic variants in *MEIS1* are not common in RLS. However, it is still possible that rare exonic variants of strong effect could play a causative role in RLS in rare cases, as is known for other complex diseases,[7] and their study is important because they could provide significant clues toward understanding of the disease mechanism.

From the Neurologische Klinik und Poliklinik (E.C.S., J.W.) and Institut für Humangenetik (T.M., J.W.), Klinikum rechts der Isar, Technische Universität München, Munich; Institut für Humangenetik (F.K., B.S., P.L., T.M., J.W.) and Institut für Epidemiologie (C.G.), Helmholtz Zentrum München, Munich, Germany; and Department of Neurology (D.K.), Charles University, 1st Faculty of Medicine and General Teaching Hospital, Prague, Czech Republic.

Study funding: Supported by in-house institutional funding from the Helmholtz Zentrum München, Munich, Germany.

Disclosure: Dr. Schulte receives a postdoctoral fellowship from the Technische Universität München, Munich, Germany. F. Knauf reports no disclosures. Dr. Kemlink receives research support from the Czech Ministry of Education. Dr. Schormair, Dr. Lichtner, and Dr. Gieger report no disclosures. Dr. Meitinger serves on scientific advisory boards for Estonian Biobank and the European Genetics Foundation; serves on the editorial board of the European Journal of Human Genetics*; receives license fee payments for a patent from KIRIN breweries; and has served as a legal consultant. Dr. Winkelmann serves on a scientific advisory board for UCB; has received speaker honoraria from UCB and Boehringer Ingelheim; has filed a patent re: Winkelmann et al. Nat Genet 2007; and receives research support from the German RLS Foundation and the Fritz Thyssen Foundation.*

Received July 8, 2010. Accepted in final form October 10, 2010.

Address correspondence and reprint requests to Dr. Juliane Winkelmann, Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Ismaningerstr. 22, 81675 Munich, Germany; winkelmann@lrz.tum.de

ACKNOWLEDGMENT

The authors thank all individuals who participated in this study and Jelena Golic for technical assistance in performing Sequenom genotyping.

1. Winkelmann J, Schormair B, Lichtner P, et al. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. Nat Genet 2007;39:1000–1006.
2. Stefansson H, Rye DB, Hicks A, et al. A genetic risk factor for periodic limb movements in sleep. N Engl J Med 2007;357:639–647.
3. Schormair B, Kemlink D, Roeske D, et al. PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. Nat Genet 2008;40:946–948.
4. Vilariño-Güell C, Chai H, Keeling BH, et al. MEIS1 p.R272H in familial restless legs syndrome. Neurology 2009;73:243–245.
5. Allen RP, Picchietti D, Hening WA, et al. Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology: a report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. Sleep Med 2003;4:101–119.
6. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 2010;7:575–576.
7. McClellan J, King MC. Genetic heterogeneity in human disease. Cell 2010;141:210–216.

Andrew McKeon, MD
Sean J. Pittock, MD
Vanda A. Lennon, MD, PhD

CSF COMPLEMENTS SERUM FOR EVALUATING PARANEOPLASTIC ANTIBODIES AND NMO-IgG

The detection of neural-reactive immunoglobulin G (IgG) autoantibodies aids the diagnosis of organ-specific autoimmune neurologic disorders. Many paraneoplastic autoantibodies reliably predict a particular cancer type and are accompanied by varied neurologic presentations of subacute onset.[1] The detection of neuromyelitis optica (NMO)–IgG predicts a relapsing inflammatory demyelinating disorder predominated by optic neuritis and transverse myelitis.[2] When an autoimmune neurologic disorder is suspected, serologic testing of serum is frequently undertaken before more invasive CSF evaluation. However, CSF evaluation can complement testing of serum when suspicion for an autoimmune etiology persists despite a negative serum result. Here we report, for a 25-year period of testing by standardized indirect immunofluorescence protocols, the frequency of neural autoantibody detection in serum and CSF.

Methods. The immunofluorescence protocols we used were validated in this laboratory for detection of paraneoplastic antibodies (anti-neuronal nuclear antibody [ANNA]-1; ANNA-2; ANNA-3; Purkinje cell cytoplasmic antibody [PCA]-1; PCA-2; PCA-Tr; collapsin response-mediator protein [CRMP]-5-IgG; amphiphysin antibody; antiglial/neuronal nuclear antibody [AGNA]-1; NMDA receptor antibody) and NMO-IgG. We searched the Mayo Clinic Neuroimmunology Laboratory database (January 1986 to March 2010) for all patient samples submitted for service evaluation of paraneoplastic or NMO-IgG. We included both Mayo Clinic and non-Mayo patients for whom both serum and CSF were submitted, and reviewed available oncologic data for patients with antibodies identified by CSF testing.

Results. Testing was performed on a clinical service basis for a median of 12 years (range 2–25 years). The antibody detection rate in all specimens ranged from 0.01% for PCA-Tr to 7% for NMO-IgG (table).

In patients for whom paired serum and CSF samples were tested, the antibody detection rate ranged from 0.08% (PCA-Tr) to 9% (NMDA receptor antibody). One or more neural autoantibodies were detected in 462 patients (497 antibodies detected). In 405 of those 462 patients, both serum and CSF yielded a positive result (88%). In 57 patients, serum or CSF alone was positive (12%). Among those patients, serum alone yielded a positive result in 31 (54%) and CSF alone in 26 (46%). For classic paraneoplastic antibodies, CSF alone yielded a positive result in 20 patients, twice as commonly as serum alone (10 patients). For NMO-IgG, serum alone yielded a positive result in 21 patients 3.5 times more commonly than CSF alone (6 patients).

Discussion. From our review of a 25-year experience with immunofluorescence testing on a service basis in the Mayo Clinic Neuroimmunology Laboratory, we found that the rate of clinically pertinent autoantibody detection was highest when both serum and CSF were tested. It is plausible that this finding may reflect a greater likelihood of physicians deciding to test both serum and CSF in patients with the highest index of clinical suspicion.

When both serum and CSF were tested, CSF was more commonly informative than serum for paraneoplastic antibody detection. This raises concern that clinically important neural antibodies may be missed when only serum is tested. This finding was most prominent for NMDA receptor-specific IgG. Consistent with this finding, Kumar et al.[3] recently reported 3 patients, each of whom had NMDA receptor IgG detected in CSF but not in serum. Where there is a high suspicion for

Targeted Resequencing and Systematic In Vivo Functional Testing Identifies Rare Variants in *MEIS1* as Significant Contributors to Restless Legs Syndrome

Eva C. Schulte,[1,2,22] Maria Kousi,[3,22] Perciliz L. Tan,[3] Erik Tilch,[2,4] Franziska Knauf,[2] Peter Lichtner,[2,4] Claudia Trenkwalder,[5,6] Birgit Högl,[7] Birgit Frauscher,[7] Klaus Berger,[8] Ingo Fietze,[9] Magdolna Hornyak,[1,10,11] Wolfgang H. Oertel,[12] Cornelius G. Bachmann,[13,14] Alexander Zimprich,[15] Annette Peters,[16] Christian Gieger,[17] Thomas Meitinger,[2,4,18] Bertram Müller-Myhsok,[18,19,20] Nicholas Katsanis,[3,23] and Juliane Winkelmann[1,2,4,18,21,23,*]

Restless legs syndrome (RLS) is a common neurologic condition characterized by nocturnal dysesthesias and an urge to move, affecting the legs. RLS is a complex trait, for which genome-wide association studies (GWASs) have identified common susceptibility alleles of modest (OR 1.2–1.7) risk at six genomic loci. Among these, variants in *MEIS1* have emerged as the largest risk factors for RLS, suggesting that perturbations in this transcription factor might be causally related to RLS susceptibility. To establish this causality, direction of effect, and total genetic burden of *MEIS1*, we interrogated 188 case subjects and 182 control subjects for rare alleles not captured by previous GWASs, followed by genotyping of ~3,000 case subjects and 3,000 control subjects, and concluded with systematic functionalization of all discovered variants using a previously established in vivo model of neurogenesis. We observed a significant excess of rare *MEIS1* variants in individuals with RLS. Subsequent assessment of all nonsynonymous variants by in vivo complementation revealed an excess of loss-of-function alleles in individuals with RLS. Strikingly, these alleles compromised the function of the canonical *MEIS1* splice isoform but were irrelevant to an isoform known to utilize an alternative 3′ sequence. Our data link *MEIS1* loss of function to the etiopathology of RLS, highlight how combined sequencing and systematic functional annotation of rare variation at GWAS loci can detect risk burden, and offer a plausible explanation for the specificity of phenotypic expressivity of loss-of-function alleles at a locus broadly necessary for neurogenesis and neurodevelopment.

Introduction

Restless legs syndrome (RLS [MIM 102300]) is a common neurologic condition with an age-dependent prevalence of up to 10% in Europe and North America.[1] It is characterized by an irresistible urge to move the legs accompanied by disagreeable, often painful, sensations in the lower limbs at night. Moving the affected legs or walking leads to prompt but only temporary relief.[1] As a consequence, individuals suffer from persistent insomnia, leading to an impairment of quality of life and mental health. RLS is a highly familial trait but genetically complex, with estimates of narrow-sense heredity between 54% and 69% as derived from twin studies.[2,3]

Genome-wide association studies (GWASs) in large RLS case/control samples have identified common susceptibility alleles at six loci that together explain about 7% of the heritability.[4] Among many models that can explain some of the "missing heritability,"[5] we considered the possibility that a collection of rare variants of strong effect, which cannot be identified by means of GWASs,[6,7] might be a contributory factor. Although a single potentially causal rare variant has been described in *MEIS1* (MIM 601739),[8,9] to date no variants of strong effect have been established. Nonetheless, for some other complex genetic diseases, such as diabetes or chronic inflammatory bowel disease, rare variants have recently been identified within known GWAS loci,[10,11] supporting the concept of allelic series in complex genetic disorders. We therefore sought to assess the potential contribution of rare variants to disease burden both by using standard statistical methods and by assessing the incidence and contribution of

[1]Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, 81675 Munich, Germany; [2]Institut für Humangenetik, Helmholtz Zentrum München, 85764 Munich, Germany; [3]Center for Human Disease Modeling, Department of Cell Biology, Duke University, Durham, NC 27710, USA; [4]Institut für Humangenetik, Technische Universität München, 81675 Munich, Germany; [5]Paracelsus Elena Klinik, 34128 Kassel, Germany; [6]Klinik für Neurochirurgie, Georg August Universität, 37075 Göttingen, Germany; [7]Department of Neurology, Medizinische Universität Innsbruck, 6020 Innsbruck, Austria; [8]Institut für Epidemiologie und Sozialmedizin, Westfälische Wilhelms Universität Münster, 48149 Münster, Germany; [9]Zentrum für Schlafmedizin, Charité Universitätsmedizin, 10117 Berlin, Germany; [10]Interdisziplinäres Schmerzzentrum, Albert-Ludwigs Universität Freiburg, 79106 Freiburg, Germany; [11]Diakoniewerk München-Maxvorstadt, 80799 Munich, Germany; [12]Klinik für Neurologie, Philipps Universität Marburg, 35039 Marburg, Germany; [13]Abteilung für Neurologie, Paracelsus Klinikum Osnabrück, 49076 Osnabrück, Germany; [14]Klinische Neurophysiologie, Georg August Universität, 37075 Göttingen, Germany; [15]Department of Neurology, Medizinische Universität Wien, 1090 Vienna, Austria; [16]Institute of Epidemiology II, Helmholtz Zentrum München, 85764 Munich, Germany; [17]Institute of Genetic Epidemiology, Helmholtz Zentrum München, 85764 Munich, Germany; [18]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; [19]Max-Planck Institut für Psychiatrie München, 80804 Munich, Germany; [20]Institute of Translational Medicine, University of Liverpool, Liverpool L69 3BX, UK; [21]Department of Neurology and Neurosciences, Center for Sleep Sciences and Medicine, Stanford University, Palo Alto, CA 94304, USA
[22]These authors contributed equally to this work
[23]These authors contributed equally to this work
*Correspondence: winkelmann@stanford.edu
http://dx.doi.org/10.1016/j.ajhg.2014.06.005.

functionally annotated variants relevant to MEIS1 biology (Figure S1 available online). For this purpose, we exploited two major resources: a well-phenotyped, ethnically homogeneous RLS cohort and an experimentally tractable method to assay MEIS1 functionality grounded on previously defined in vivo observations on the roles of this protein in neurogenesis, wherein suppression of *meis1* in zebrafish embryos led to a quantitative reduction of the optic tectum, a major site of neurogenesis in the developing brain, and malformation of rhombomeres 3 and 5, which represent early hindbrain structures shown previously to be defective in the absence of *meis1*.[12,13] Our data showed a significant enrichment of rare variants across both *MEIS1* and also all seven RLS-associated genes. Based upon population statistics alone, only one single low-frequency variant in the 3′ UTR (rs11693221) showed significant association with RLS. However, the combinatorial usage of functional annotation and statistical analyses highlighted a major contribution of loss-of-function variants in *MEIS1* and suggested that rare alleles in this locus pose significant RLS risk to individuals.

Materials and Methods

Participants

Both case and control populations were entirely of German and Austrian descent. In the case subjects, diagnosis was based on the diagnostic criteria of the International RLS Study Group[1] as assessed in a personal interview conducted by an RLS expert. In keeping with the previous GWASs,[4,14] we excluded individuals with secondary RLS due to uremia, dialysis, or anemia resulting from iron deficiency. The presence of secondary RLS was determined by clinical interview, physical and neurological examination, blood chemistry, and nerve conduction studies whenever deemed clinically necessary. Participants' written informed consent was obtained prior to the initiation of the study. The institutional review boards of the contributing authors approved the study. The primary review board was located in Munich at the Bayerische Ärztekammer and Technische Universität München.

Genotyping by High-Resolution Melting Curve Analysis

In a first step, we used Idaho LightScanner high-resolution melting curve analysis (Biofire) to screen the coding regions and exon/intron boundaries of *PTPRD* (MIM 601598), *BC034767*, *TOX3* (MIM 611416), *BTBD9* (MIM 611237), *MEIS1*, and *MAP2K5* (MIM 602520) for variants. Due to the high GC content, the coding regions ±10 bp of *SKOR1* (MIM 611273) could not be subjected to LightScanner analysis and were Sanger sequenced instead. Included in the screening were 188 German RLS-affected case subjects and 182 general population control subjects belonging to the KORA cohort[13] based in Southern Germany. Where possible, the 188 case subjects used were half homozygous and half heterozygous for the published risk alleles.[4,14–16] The same set of control subjects was used for all screening experiments. *MEIS1*, *TOX3*, and *BC034767* variants identified in the 188 case subjects have already been published.[4,8] In the case of an altered melting pattern suggestive of variants, Sanger sequencing ensued to identify the underlying variant. The same method was used to screen the coding regions of *MEIS1* isoforms 1 and 2 ±10 bp as well as the 5′ and 3′ UTRs in 3,760 RLS case subjects of German and Austrian descent (62.2 ± 12.8 years; 30.8% male) and 3,542 general-population control subjects (55.1 ± 13.8 years; 40.1% male) belonging to the S4 and F4 surveys of KORA.[17] For the discovery sample, group comparisons between case and control subjects were performed in R[18] for each gene and each type of variant separately, with and without the common risk allele genotype as covariate, using logistic regression (logreg) of the phenotype on aggregate minor allele counts of variants of each type.[19] To account for a possible bias introduced by the comparison of different sets of risk-genotype-selected cases to a constant set of unselected controls, we performed a meta-analysis using rmeta to evaluate the contribution of rare coding variants across all seven genes.

Empirical p values were calculated with 1,000 permutations of the phenotype and assessing the ratio of p values equal to or smaller than the p value belonging to the original phenotype. For the large-scale screening of *MEIS1*, both logreg of the phenotype on number of variants as well as sequence kernel association tests (SKAT)[20] with and without Madsen-Browning weights[21] were performed. Empirical p values are based on 10,000 permutations of the phenotype, calculating the ratio of test statistics equal to or larger than the test statistic of the original phenotype.

Genotyping by Mass Spectrometry

Genotyping was carried out on the MassARRAY system by MALDI-TOF mass spectrometry with iPLEX Gold chemistry (Sequenom). Primers were designed with AssayDesign v.3.1.2.2 with iPLEX Gold default parameters. No assay could be designed for seven variants, largely those located in the extremely GC-rich gene *SKOR1*. Further, three assays failed two or more times and were, therefore, not pursued further. Automated genotype calling was carried out with SpectroTYPER v.3.4. Genotype clustering was visually checked by an experienced evaluator. SNPs with a call rate <90% were excluded. The genotyping sample consisted of 3,262 case subjects of German and Austrian descent (65.3 ± 11.3 years; 29.3% male) and 2,944 general population control subjects (56.1 ± 13.3 years; 48.7% male) from the KORA F4 survey.[17] For the most part, case and control subjects used in both genotyping approaches were drawn from the same samples. Both logreg of the phenotype on number of variants as well as SKATs[20] with and without Madsen-Browning weights[21] were performed. Empirical p values are based on 200 permutations of the phenotype, calculating the ratio of test statistics larger than the test statistic of the original phenotype.

In Vivo Complementation in Zebrafish Embryos, In Situ Hybridization, and Whole-Mount Immunostaining

Splice-blocking morpholinos (MOs) against *meis1* and *map2k5* were designed and obtained from Gene Tools. We injected 1 nl of diluted MO (4 ng for *meis1*_MO1, 3 ng for *meis1*_MO2, and 5 ng for *map2k5*) and/or RNA (75 pg for *meis1*, 100 pg for *map2k5*) into wild-type zebrafish embryos at the 1- to 2-cell stage. To evaluate hindbrain organization, injected embryos were raised until the 10- to 13-somite stage, corresponding to 14 to 16 hr post-fertilization (hpf); they were then dechorionated and fixed in 4% paraformaldehyde (PFA) overnight. Fixed embryos were transferred to 100% methanol at −20°C for at least 2 hr and were then processed after standard protocols[22] using a digoxygenin-labeled antisense riboprobe against *krox20*. For analysis of the

optic tectum, injected embryos were fixed overnight at 72 hpf in 4% PFA and stored in 100% methanol at −20°C. For acetylated tubulin staining, embryos were fixed in Dent's fixative (80% methanol, 20% DMSO) overnight at 4°C. The embryos were permeabilized with proteinase K followed by postfixation with 4% PFA. PFA-fixed embryos were washed first in PBS and subsequently in IF buffer (0.1% Tween-20, 1% BSA in PBS) for 10 min at room temperature. The embryos were incubated in blocking buffer (10% FBS, 1% BSA in PBS) for 1 hr at room temperature. After two washes in IF buffer for 10 min each, embryos were incubated in the primary antibody (anti-acetylated tubulin [T7451, mouse, Sigma-Aldrich], 1:1,000) in blocking solution, overnight at 4°C. After two additional washes in IF buffer for 10 min each, embryos were incubated in the secondary antibody solution (Alexa Fluor goat anti-mouse IgG [A21207, Invitrogen], 1:1,000) in blocking solution, for 1 hr at room temperature.

For RNA rescue and overexpression experiments, the human wild-type mRNAs of isoforms 1 (RefSeq accession number NM_002398.2/ENST00000272369) and 2 (no RefSeq ID/ENST00000381518) of *MEIS1* as well as the canonical isoform of *MAP2K5* (RefSeq NM_145160.2/ENST00000178640) were cloned into the pCS2+ vector and transcribed in vitro using the SP6 Message Machine kit (Ambion). All variants identified in isoform 1 as well as all additional variants only coding in isoform 2 plus two functional null variants from isoform 1 (p.Ser204Thr [c.610T>A] and p.Arg272His [c.815G>A]; both RefSeq NM_002398.2/ENST00000272369) were introduced with Phusion high-fidelity DNA polymerase (New England Biolabs) and custom-designed primers. Additionally, a non-naturally occurring homeobox domain-dead variant of *MEIS1* (p.Arg276Ala + p.Asn325Ala [c.826_827delinsGC + c.973_974delinsGC]; RefSeq NM_002398.2/ENST00000272369) was created. All the experiments were repeated in triplicate and significance of the morphant phenotype was judged with Student's t test.

Results

Variant Screening of Seven RLS GWAS Candidate Genes

To assess low-frequency (1% < minor allele frequency [MAF] < 5%) coding variation at the known RLS susceptibility loci encompassing *PTPRD*, *TOX3*, *BTBD9*, *MEIS1*, *MAP2K5*, *SKOR1*, and the noncoding RNA *BC034767*,[4,14–16] we screened the coding regions and exon-intron boundaries (±10 bp) in 188 German individuals with RLS and 182 control subjects belonging to the KORA population cohort[13] based in Southern Germany. A total of 49 variants with MAF < 5% were identified in case and control subjects together (Table S1). When collapsed across all seven genes, rare and low-frequency nonsynonymous variants showed a trend toward being more frequent in case than in control subjects (39 in case subjects versus 24 in control subjects; $p = 0.103$; logreg meta-analysis; odds ratio [OR] = 1.40). Nonsynonymous and synonymous variants combined, however, showed a stronger enrichment in individuals with RLS (77 case subjects versus 46 control subjects; $p = 0.023$, logreg meta-analysis; OR = 1.51). Addition of the common risk allele genotype[4,14–16] into the analysis as a covariate decreased the permuted p value (nonsynonymous only: $p = 0.079$; all variants: $p = 0.008$; n = 1,000 permutations), suggesting that the enrichment of rare variants across all loci is independent of the common risk allele genotype. However, the degree of interdependence between common and rare variants differed between genes (Table S2).

Within *MEIS1*, synonymous or nonsynonymous variants with a MAF below 5% were present in nine case subjects but only one control subject ($p = 0.021$) although no marked difference in the amount of variation was obvious in any of the other genes (Table S1). Here, the p value increased modestly after addition of the common risk variant (rs2300478) as covariate ($p = 0.080$; 1,000 permutations), indicating some interdependence of rare and common alleles at the *MEIS1* locus. Variants did not seem to cluster within specific regions of the examined genes (Figure S2).

Genotyping of Identified Low-Frequency and Rare Variants

To assess a possible association with the RLS phenotype, we next genotyped 39 of the 49 identified variants in 3,262 German and Austrian RLS case subjects (65.3 ± 11.3 years; 29.3% male) and 2,944 KORA control subjects (56.1 ± 13.3 years; 48.7% male). Of the 49 variants, 10 could not be included for technical reasons (see Material and Methods section above). Although variants (either altogether or only nonsynonymous ones) with MAF < 5% and MAF < 1% were not significantly enriched in RLS, we observed a distinct excess of very rare variants with a KORA-derived MAF < 0.1% (total: 57 case versus 16 control subjects, $p = 4.99 \times 10^{-4}$, OR = 2.50; nonsynonymous only: 23 case versus 5 control subjects, $p = 0.0019$; OR = 3.92; logreg; Figure 1).

We then went on to assess whether more sophisticated tests used to analyze rare variant associations of bidirectional effects (SKAT)[20] and differing allele frequency (SKAT with Madsen-Browning weights)[21] would change the association signal. However, although frequency-weighted results were still superior to unweighted results in most cases, overall SKAT analysis led to an increase in association p values both for the joined analysis across all loci as well as for *MEIS1* alone (Table S2).

We next asked whether this signal is distributed across all seven tested genes, or whether a specific subset of tested loci cause the apparent enrichment of rare nonsynonymous variation in individuals with RLS. A significant association with RLS was observed for *MEIS1*; the aggregate risk conferred by the variants showed a large increase as we transitioned from rare (MAF < 1%) to very rare (MAF < 0.1%) (Figure 1). More specifically, rare variants with MAF < 1% (total: 116 case subjects versus 67 control subjects, $p = 0.0064$, OR = 1.51; nonsynonymous only: 39 versus 14, $p = 0.0024$, OR = 2.46) and MAF < 0.1% (both total and nonsynonymous only: 9 versus 1: $p = 0.014$; OR = 8.14; Figure 1) were seen more frequently in cases than in controls. SKAT (all: $p = 0.049$;

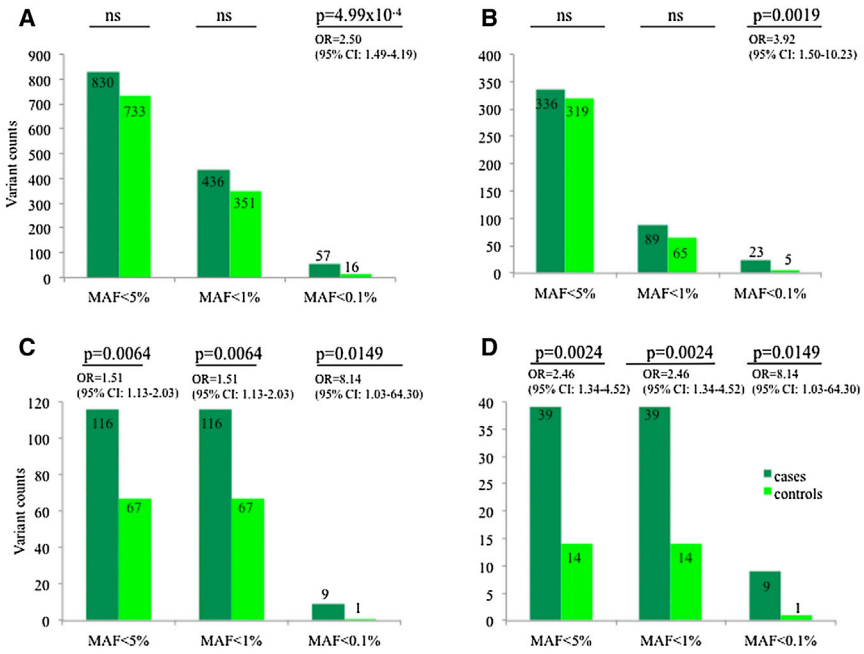


Figure 1. Excess of Rare Coding Variants at RLS-Associated GWAS Loci

Frequency assessment of 39 low-frequency and rare variants identified in the coding sequences of seven genes associated with RLS[4,14–16] in 3,262 case subject and 2,944 control subjects revealed an excess of both overall (A) and nonsynonymous (B) variants with MAF < 0.1% across all examined loci. The same held true for the overall (C) and nonsynonymous (D) variants at the *MEIS1* locus.

nonsynonymous only: p = 0.009) and Madsen-Browning-weighted SKAT (all: p = 0.019; nonsynonymous only: p = 0.029; 10,000 permutations each) substantiated this finding (Table S2). No low-frequency coding variants with MAF between 1% and 5% were found within the coding regions of *MEIS1*.

After exclusion of *MEIS1*, logreg (all variants: p = 0.009; nonsynoymous only: p = 0.024; 2,000 permutations) for variants with MAF < 0.1% and SKAT using Madsen-Browning weights (all variants: p = 0.019; nonsynonymous only: p = 0.019; 100 permutations) showed a nominally significant enrichment across all other six genes. In *PTPRD* alone, rare variants of all classes were also encountered more frequently in individuals with RLS (all variants with MAF < 0.1%: $p = 9.99 \times 10^{-4}$, logreg; all variants: p = 0.029, SKAT with MB) whereas this enrichment was not observed for nonsynonymous variants only. Of note, several individual rare variants in *MEIS1* (p.Arg272His [c.815G>A] and p.Met453Thr [c.1359T>C] [ENST00000381518]), *TOX3* (p.Ala233Ala [c.699T>C]; RefSeq NM_001080430.2/ENST00000219746), and *PTPRD* (c.551−4C>G and p.Pro278Pro [c.834T>G]; both RefSeq NM_002839.3/ENST00000381196) were associated nominally with RLS in the large case/control sample; however, associations did not withstand correction for multiple testing (Table S1).

Assessment of Rare Variation in *MEIS1*

The excess of low-frequency and rare variants at RLS-associated GWAS loci was most pronounced for *MEIS1*. Therefore, we sought to expand our analysis to a more comprehensive investigation of genetic variation with regard to frequency and location within *MEIS1* by screening the coding regions ±10 bp as well as the 5′ and 3′ UTRs for variants with MAF < 5% in 3,760 German and Austrian RLS case subjects (62.2 ± 12.8 years; 30.8% male) and 3,542 KORA control subjects (55.1 ± 13.8 years; 40.1% male). We identified a total of 75 such variants (Tables S3 and S4); 28 of these lie in either the canonical isoform 1 (ENST00000272369) or a longer isoform 2 (ENST00000398506) of *MEIS1* that encodes an alternate start site and an alternate C terminus. All synonymous and nonsynonymous variants identified in the initial screening of 188 case subjects and 182 control subjects were observed again.

Overall, we observed an excess of variants with MAF < 5% across all examined regions of *MEIS1* (1,383 variant counts in case subjects versus 606 in control subjects, $p = 1.04 \times 10^{-61}$; not permuted), which was driven primarily by a low-frequency variant, rs11693221, in the 3′ UTR ($MAF_{cases} = 13.55\%$, $MAF_{controls} = 3.58\%$; $p = 1.27 \times 10^{-89}$; OR = 4.42 [95% CI: 3.83–5.11]; not permuted; Figure 2, Tables S3 and S4). After exclusion of this variant, the remainder of individuals with low-frequency and rare variants across all regions of *MEIS1* was similar in case and control subjects (432 versus 396; p = 0.68). However, stratification of variants according to their localization showed an excess of rare variants with MAF < 5% in the 5′ UTR (16 case subjects versus 2 control subjects; logreg: p = 0.001, OR = 7.56; SKAT: p = 0.01; SKAT with MB: p = 0.006) and among nonsynonymous variants in isoform 2 (34 case subjects versus 15 control subjects; logreg: p = 0.007, OR = 2.31; SKAT: p = 0.004; SKAT with MB: p = 0.0005) (Figure 2, Table S3).

Functional Analysis of Rare Nonsynonymous Variants in *MEIS1* by In Vivo Complementation in Zebrafish Embryos

Resequencing of any locus is certain to reveal rare variants in both case and control subjects which, bereft of a means of preselecting for variants relevant to protein function, can dampen or extinguish bona fide association signals. We, therefore, considered a paradigm grounded on prior knowledge of *MEIS1* to test whether each of the rare discovered variants in our study have an effect on function

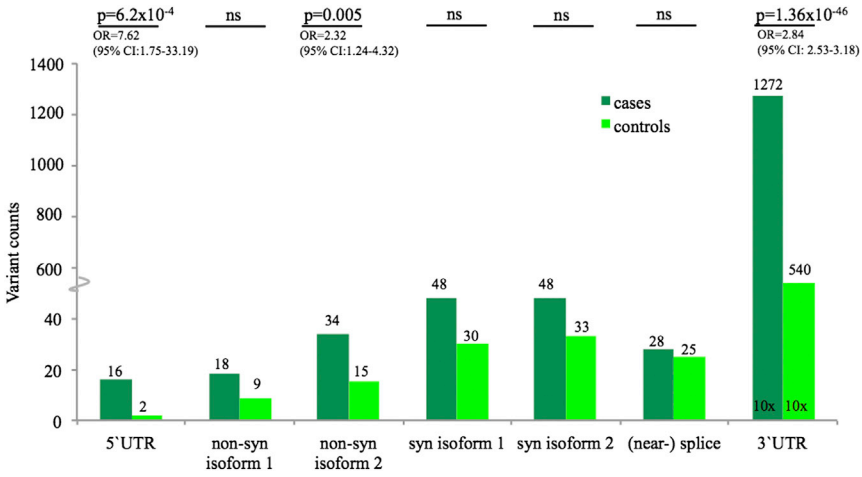


Figure 2. Variant Screening of the Coding Regions and UTRs of *MEIS1* in 3,760 Individuals with RLS and 3,542 KORA Control Subjects

Stratification according to variant localization shows an excess of rare variants in both the 5′ UTR and among nonsynonymous coding variants. Low-frequency and rare variants in the 3′ UTR were also more frequent in case subject than in control subjects. No difference was observed in the number of individuals carrying synonymous coding or (near-) splice variants.

and to then use this information to assess the burden of deleterious genetic lesions in our case/control RLS study. During zebrafish development, suppression of *meis1* has been shown to impact neurogenesis, a phenotype captured prominently by the quantitative reduction of the size of the optic tectum 72 hpf[12] as well as the disruption of hindbrain patterning at 14 hpf.[13] We first tested the ability of human *MEIS1* mRNA to rescue the optic tectum size phenotype and to thus establish a baseline assay for the evaluation of the identified nonsynonymous variants. To this end, we designed two independent MOs that target different exon-intron splice junctions of the endogenous zebrafish *meis1*. Both MOs gave rise to the same phenotype, bolstering our confidence in the specificity of the assay: we observed a reduction of the size of the optic tectum by ~30% when injecting 4 ng of *meis1*_MO1 targeting exon-intron boundary 2, and by ~20% when injecting 3 ng of a previously characterized MO (*meis1*_MO2) against the acceptor site of exon 2[23] (Figures S3 and S4). In both cases, the phenotype was rescued significantly and reproducibly ($p < 0.0001$; performed in triplicate, scored blind to injection cocktail) by coinjection with 75 pg of human capped *MEIS1* mRNA (Figures S3 and S4). The optic tectum phenotype could be rescued by either the canonical *MEIS1* isoform or by an isoform utilizing an alternative 3′ terminus. By contrast, injection of the "domain-dead" human mRNA (MEIS1_DD) bearing two variants engineered to ablate DNA binding ability (p.Arg276Ala+Asn325Ala) was indistinguishable from MO alone ($p < 0.68$). As a test of the relevance of the phenotype to RLS, MO suppression of *map2k5,* another GWAS-associated RLS gene, yielded a similar phenotype with regard to the size of the optic tectum, which was also specific as it could be rescued by coinjection with human *MAP2K5* mRNA (Figure S5).

Given these data, we proceeded to perform in vivo complementation assays on all 13 nonsynonymous coding variants identified in isoform 1, as well as in the 4 nonsynonymous variants that lie in the unique sequence of isoform 2 (Figure 3, Table S4), wherein human mRNA bearing one test variant was coinjected with MO and compared to the rescue ability of WT mRNA (n = at least 51 embryos tested per injection, Figure 3 and Tables S5 and S6). We classified variants as benign (rescue indistinguishable from WT), hypomorphs (mutant rescue significantly worse than WT but better than MO alone), or functionally null (indistinguishable from MO alone). Among the 13 variants of isoform 1, we identified three benign, four hypomorphic, and six null variants (Figures 3 and 4). Overexpression of *MEIS1* WT mRNA or mRNAs harboring each of the 13 variants had no effect on optic tectum size (Figure S6).

Layering these data over the incidence and distribution of these variants in our case/control data set, we found a significant excess of functionally null variants in individuals with RLS compared to control subjects (14 in case subjects versus 2 in control subjects; $p = 0.0012$; OR = 7.48 [95% CI: 1.68–33.40]) (Figures 3 and 5) whereas hypomorphic (2 in case subjects versus 4 in control subjects; $p > 0.05$) and benign (2 in cases subject versus 2 in control subjects; $p > 0.05$) variants showed similar distributions in case and control subjects.

To corroborate these data, we designed a second assay at an earlier developmental time point, grounded on the known requirement of *meis1* for the organization of the hindbrain; in triplicate experiments, ~30% of the embryos suppressed for *meis1* developed hindbrain defects that reproduced previously reported *meis1* phenotypes.[13] These consisted of significant widening of rhombomere 3 and/or 5 (r3 and r5); shortening of the distance between r3 and r5; thinning of either one of those structures; or absence of r3 and/or r5 altogether (Figure 4). This phenotype was rescued by coinjection with 75 pg of WT human *MEIS1* mRNA ($p = 0.045$). Assessment of six variants from our series (three variants scored as null and three variants scored as benign in the optic tectum assay) and blind triplicate scoring confirmed this result: p.Ser204Thr, p.His239Tyr (c.715C>T), and p.Arg272His were verified as null variants and p.Ala122Val (c.365C>T), p.Ser175Asn (c.524G>A), and p.Met366Leu (c.1096A>T) were validated as benign (all RefSeq NM_002398.2/ENST00000272369).

By contrast, all four nonsynonymous variants exclusive to isoform 2 scored benign in our assay. Strikingly, two

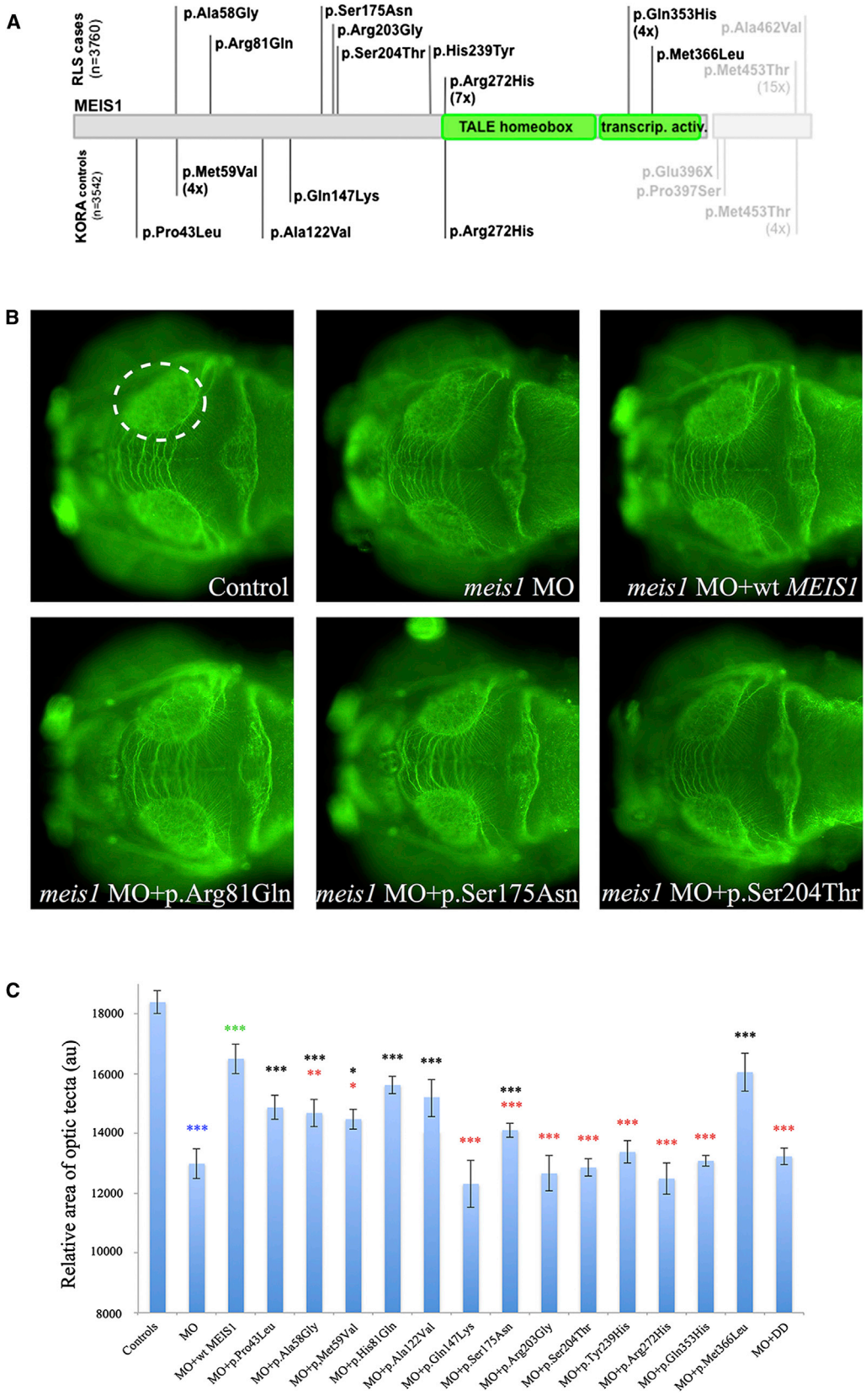


Figure 3. Functional Assessment of Rare Nonsynonymous Variants in *MEIS1* by In Vivo Complementation in Zebrafish Embryos
(A) Location and frequency of nonsynonymous *MEIS1* variants examined in zebrafish. Variants found in case subjects are given above the gene, those found in control subjects below. The short, canonical isoform 1 of *MEIS1* is given in dark gray (ENST00000272369); the additional amino acids in the longer isoform 2 in light gray (ENST00000398506).

(legend continued on next page)

variants found to be functional nulls in the isoform 1 complementation assay (p.Ser204Thr and p.Arg272His) were able to fully rescue the tectum size phenotype in isoform 2 (Figures 5 and S6), suggesting that the contribution of rare variants to RLS is mediated specifically by reduced activity of MEIS1 isoform 1-encoded protein.

Discussion

Previous GWASs have established the genomic locus encompassing *MEIS1* as the most significant susceptibility region for RLS.[4,14] The most likely candidate gene in this region is *MEIS1*, a TALE homeobox transcription factor known to be involved in specifying spinal motor neuron pool identity and connectivity[24] as well as proximo-distal limb patterning[25] and expressed in forebrain neurons and astrocytes[26] during embryonic development. A common RLS-linked intronic variant in *MEIS1* was also shown to induce differential forebrain enhancer activity during development.[27] Additional studies in the context of RLS have suggested a link between *MEIS1* and iron metabolism in the central nervous system.[28,29]

The excess of rare alleles of functional effect in RLS case subjects compared to control subjects shown here substantiates *MEIS1* as the causal genetic factor underlying the observed associations. Moreover, it implicates loss of function as the underlying mechanism, at least with regard to rare variants. We also observed a new association with a low-frequency variant (rs11693221) located in the 3′ UTR of the ENSEMBL-derived canonical isoform 1 of *MEIS1* (ENST00000272369) that represents the largest single-allele genetic risk factor for RLS identified to date. In the 1000 Genomes data,[30] the linkage disequilibrium (LD, r^2) between rs11693221 and the most significantly associated common variant within *MEIS1* (rs12469063)[4,14] is low ($r^2 = 0.080$). In the same data set,[30] rs11693221 is in high LD ($r^2 > 0.8$) solely with a single low-frequency variant (rs113851554, $r^2 = 0.83$). rs113851554 is located within a highly conserved noncoding region in intron 8 and has previously also been shown to be associated with RLS in a Canadian case/control sample.[29] Follow-up analyses are needed to fully dissect all functional effects underlying the *MEIS1* association signal.

Nonetheless, the finding of an associated low-frequency variant (rs11693221) in the 3′ UTR of isoform 1 in conjunction with the observed excess of rare variants in the 5′ UTR of *MEIS1* in individuals with RLS implicates the UTRs in disease pathogenesis, potentially through regulating expression and/or mRNA stability. The excess of rare noncoding variants in the 100 bp surrounding the exons of nine genes associated with asthma[31] and the fact that fine-mapping studies located about 22% of 36 GWAS association signals for celiac disease to either the 5′ or the 3′ noncoding regions (UTRs and several kilobases up- or downstream)[32] could indicate that these regions are indeed important in the context of complex genetic diseases and might be overlooked by the current surge of whole-exome sequencing studies.

The functional experiments conducted in zebrafish allowed us to differentiate between potentially benign and pathological sequence variation and thereby increased both effect size and significance levels observed in burden testing. Moreover, the identified rare nonsynonymous *MEIS1* variants, which showed an effect on optic tectum size in zebrafish embryos, were restricted to isoform 1 of *MEIS1* (ENST00000272369). Previous studies have demonstrated that rare variants can exhibit isoform-specific effects on a given phenotype, such as in the case of *DNAJB6* in limb-girdle muscular dystrophy.[33] Because the pathophysiology of RLS is just beginning to be elucidated, it will be of importance to see when and where this isoform of *MEIS1* is expressed as the temporal and spatial expression patterns of the different *MEIS1* transcripts are currently unknown. Given our observations, we speculate that understanding the differential biological roles of isoform 1 will help dissect the subset of MEIS1 functions relevant to the etiopathology of RLS.

Recent studies have implicated allelic series of variants of different frequency and effect sizes at loci identified in the context of GWASs in complex genetic diseases. In single cases, individual rare variants were shown to be associated with the phenotype[11,34,35] whereas in other cases it was the collective of rare variation either within a single gene[10] or across a number of GWAS-identified loci.[36,37] Our data substantiate this role of the whole of genetic variation, from common to low-frequency to rare variation, at a GWAS-identified locus in the genetic architecture of a complex genetic disease. Interestingly, the addition of synonymous variants to the burden analyses yielded a more significant enrichment of rare variants in many situations. Whether this is due to increased power, fine-scale population substructure, artificial signal amplification driven by high LD between the synonymous and the nonsynonymous or causal variants, or a true causal contribution of rare synonymous

(B) At 72 hpf, zebrafish larvae were stained as whole mounts using an antibody against acetylated tubulin and the size of the optic tecta was measured for phenotypic read out. Control, morpholino injection, and rescue by human WT mRNA are shown in the upper panels. The lower panels illustrate the effects of different alleles tested.
(C) Quantification of optic tectum area in zebrafish larvae at 72 hpf (n = at least 51 per genotype). Benign alleles show a significant difference with regard to the MO injection, hypomorphic alleles a significant difference with regard to both the MO injection and the rescue (MO plus WT) injection, and null alleles are significantly different from the rescue only. Asterisks denote significance levels as determined by Student's t test. Color of asterisks as follows: blue, MO versus control; green, rescue versus MO; black, allele versus MO; red, allele versus WT rescue. Abbreviations are as follows: MO, morpholino; WT, wild-type. Error bars represent standard deviations across all examined embryos.

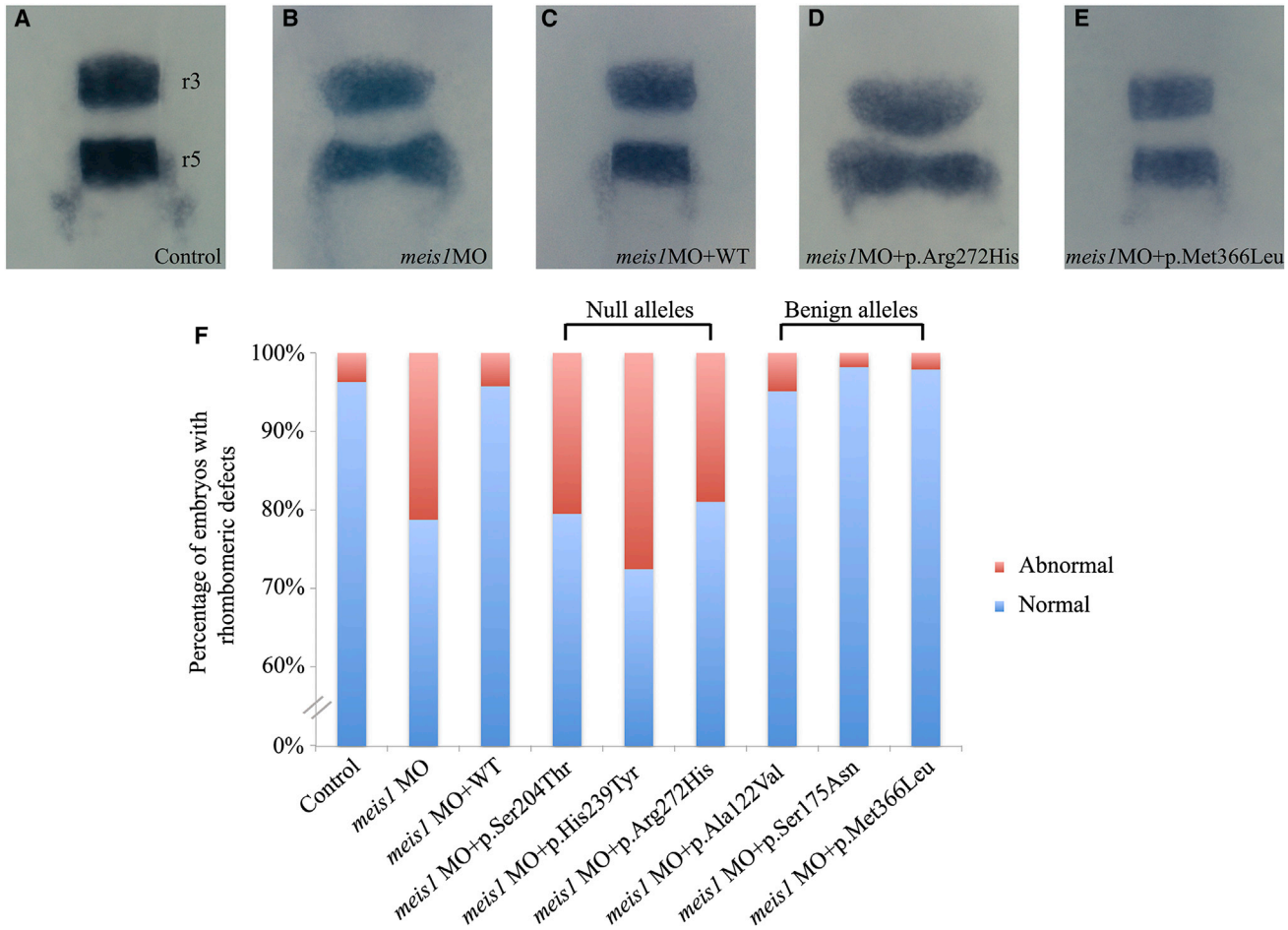


Figure 4. Functional Assessment of Null and Benign *MEIS1* Variants by In Vivo Complementation in Zebrafish Embryos and Evaluation of Hindbrain Patterning

(A–E) At 14–16 hpf, developing zebrafish embryos were evaluated for the integrity of rhombomeres 3 and 5 (r3 and r5) by in situ hybridization with a riboprobe against *krox20*. Upon disruption of *meis1*, we observed rhombomeric defects that involved widening of the evaluated structures (B and D) or shortening of the distance between r3 and r5 (D), as well as thinning or absence of the evaluated structures.

(F) Quantification showing that the aberrant phenotypes were especially pronounced in the morphant embryos and embryos coinjected with MO+null mRNA ($n \geq 26$ embryos per genotype). Abbreviations are as follows: MO, morpholino; WT, wild-type.

variation as was recently reported for sporadic Alzheimer disease[38] cannot be ascertained within the bounds of this study. We also note that we were unable to establish a single rare variant of large effect involved in RLS at the examined loci, possibly due to a lack of power in sight of the large amount of background rare genetic variation and the excess of singletons known to exist in the human genome.[39–41] However, systematic functional annotation of such singletons has improved our interpretative ability and has suggested that, in addition to the risk conferred by common and low-frequency alleles, rare variants contribute significantly to the genetic burden in RLS. Our data are consistent with previous observations wherein the rarer a genetic variant, the more likely it is to harbor a functional effect[42] and extends these to a disease context. It is particularly notable that, subsequent to functional tagging of alleles, we observed association with RLS only for functionally null variants, but not for hypomorphs, potentially intimating a threshold effect on total MEIS1 function necessary to drive pathology. Nonetheless, although our positive and negative controls for the in vivo complementation assay support previously reported high specificity and sensitivity for the approach,[43] and despite the fact that we achieved full concordance of allele effect tagging by two independent in vivo complementation assays, it will be important to validate our observations further in an independent model system; the evaluation of the two functionally null variants *MEIS1* p.Arg272His and p.Gln353His (c.1059G>C) (both RefSeq NM_002398.2/ENST00000272369) found primarily or exclusively in case subjects, located in the homeobox and transcription activation domains, respectively, in animal models could prove worthwhile to establish their relevance to the RLS phenotype and to further explore the pathophysiology of the disorder.

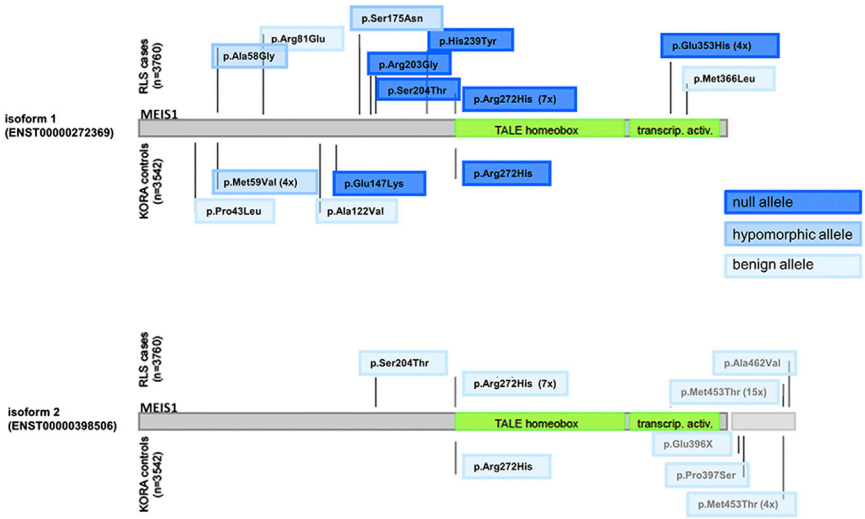


Figure 5. Functional Annotation of Rare, Nonsynonymous Variants in Isoforms 1 and 2 of *MEIS1* According to the Effect on Optic Tectum Size in Zebrafish Embryos

When tested in *MEIS1* isoform 1, an excess of rare null alleles was present among individuals with RLS. When tested in isoform 2, none of the variants show a functional effect, suggesting an isoform specificity with regard to a potential involvement in RLS. Numbers in parentheses behind variants indicate the total variant count in either case subjects or control subjects. If no number is given, the variant is a singleton.

Supplemental Data

Supplemental Data include six figures and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2014.06.005.

Acknowledgments

We are gratefully indebted to Katja Junghans, Susanne Lindhof, Jelena Golic, Sybille Frischholz, and Regina Feldmann at the Institut für Humangenetik, Helmholtz Zentrum München (Munich, Germany) for their expert technical assistance in performing Sequenom genotyping and Light Scanner analyses. Moreover, we thank Rene Rezsohazy at the Université de Louvain (Louvain, Belgium) for his advice in the construction of the domain-dead construct of *MEIS1*. This study was funded by in-house institutional funding from Technische Universität München and Helmholtz Zentrum München, a grant entitled "Functional Analysis of Rare Variants in Restless Legs Syndrome" from the Else Kröner-Fresenius-Stiftung (2013_A124), by seed funding from the Center for Human Disease Modeling, Duke University, and by P50 MH094268 to N.K. Recruitment of case and control cohorts was supported by institutional (Helmholtz Zentrum München) and government funding from the German Bundesministerium für Bildung und Forschung (03.2007-02.2011 FKZ 01ET0713). W.H.O. is a Senior Research Professor of the Charitable Hertie Foundation, Frankfurt/Main, Germany. N.K. is a Distinguished Brumley Professor.

Received: December 9, 2013
Accepted: June 10, 2014
Published: July 3, 2014

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://browser.1000genomes.org
CRAN – Package rmeta, http://cran.r-project.org/web/packages/rmeta/index.html
dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq

Accession Numbers

The dbSNP accession numbers of the variants reported in this paper are available in Table S7.

References

1. Allen, R.P., Picchietti, D., Hening, W.A., Trenkwalder, C., Walters, A.S., and Montplaisi, J.; Restless Legs Syndrome Diagnosis and Epidemiology workshop at the National Institutes of Health; International Restless Legs Syndrome Study Group (2003). Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. Sleep Med. *4*, 101–119.
2. Desai, A.V., Cherkas, L.F., Spector, T.D., and Williams, A.J. (2004). Genetic influences in self-reported symptoms of obstructive sleep apnoea and restless legs: a twin study. Twin Res. *7*, 589–595.
3. Xiong, L., Jang, K., Montplaisir, J., Levchenko, A., Thibodeau, P., Gaspar, C., Turecki, G., and Rouleau, G.A. (2007). Canadian restless legs syndrome twin study. Neurology *68*, 1631–1633.
4. Winkelmann, J., Czamara, D., Schormair, B., Knauf, F., Schulte, E.C., Trenkwalder, C., Dauvilliers, Y., Polo, O., Högl, B., Berger, K., et al. (2011). Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. PLoS Genet. *7*, e1002171.
5. Maher, B. (2008). Personal genomes: The case of the missing heritability. Nature *456*, 18–21.
6. Gibson, G. (2011). Rare and common variants: twenty arguments. Nat. Rev. Genet. *13*, 135–145.
7. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753.
8. Schulte, E.C., Knauf, F., Kemlink, D., Schormair, B., Lichtner, P., Gieger, C., Meitinger, T., and Winkelmann, J. (2011).

Variant screening of the coding regions of MEIS1 in patients with restless legs syndrome. Neurology *76*, 1106–1108.
9. Vilariño-Güell, C., Chai, H., Keeling, B.H., Young, J.E., Rajput, A., Lynch, T., Aasly, J.O., Uitti, R.J., Wszolek, Z.K., Farrer, M.J., and Lin, S.C. (2009). MEIS1 p.R272H in familial restless legs syndrome. Neurology *73*, 243–245.
10. Bonnefond, A., Clément, N., Fawcett, K., Yengo, L., Vaillant, E., Guillaume, J.L., Dechaume, A., Payne, F., Roussel, R., Czernichow, S., et al.; Meta-Analysis of Glucose and Insulin-Related Traits Consortium (MAGIC) (2012). Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat. Genet. *44*, 297–301.
11. Rivas, M.A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C.K., Boucher, G., Ripke, S., Ellinghaus, D., Burtt, N., et al.; National Institute of Diabetes and Digestive Kidney Diseases Inflammatory Bowel Disease Genetics Consortium (NIDDK IBDGC); United Kingdom Inflammatory Bowel Disease Genetics Consortium; International Inflammatory Bowel Disease Genetics Consortium (2011). Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat. Genet. *43*, 1066–1073.
12. Erickson, T., French, C.R., and Waskiewicz, A.J. (2010). Meis1 specifies positional information in the retina and tectum to organize the zebrafish visual system. Neural Dev. *5*, 22.
13. Waskiewicz, A.J., Rikhof, H.A., Hernandez, R.E., and Moens, C.B. (2001). Zebrafish Meis functions to stabilize Pbx proteins and regulate hindbrain patterning. Development *128*, 4139–4151.
14. Winkelmann, J., Schormair, B., Lichtner, P., Ripke, S., Xiong, L., Jalilzadeh, S., Fulda, S., Pütz, B., Eckstein, G., Hauk, S., et al. (2007). Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. Nat. Genet. *39*, 1000–1006.
15. Schormair, B., Kemlink, D., Roeske, D., Eckstein, G., Xiong, L., Lichtner, P., Ripke, S., Trenkwalder, C., Zimprich, A., Stiasny-Kolster, K., et al. (2008). PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. Nat. Genet. *40*, 946–948.
16. Stefansson, H., Rye, D.B., Hicks, A., Petursson, H., Ingason, A., Thorgeirsson, T.E., Palsson, S., Sigmundsson, T., Sigurdsson, A.P., Eiriksdottir, I., et al. (2007). A genetic risk factor for periodic limb movements in sleep. N. Engl. J. Med. *357*, 639–647.
17. Wichmann, H.E., Gieger, C., and Illig, T.; MONICA/KORA Study Group (2005). KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen *67* (*Suppl 1*), S26–S30.
18. Team, R.C. (2014). R: A language and environment for statistical computing (Vienna: R Foundation for Statistical Computing).
19. Morris, A.P., and Zeggini, E. (2010). An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genet. Epidemiol. *34*, 188–193.
20. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. *89*, 82–93.
21. Madsen, B.E., and Browning, S.R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet. *5*, e1000384.
22. Thisse, C., and Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat. Protoc. *3*, 59–69.
23. Cvejic, A., Serbanovic-Canic, J., Stemple, D.L., and Ouwehand, W.H. (2011). The role of meis1 in primitive and definitive hematopoiesis during zebrafish development. Haematologica *96*, 190–198.
24. Dasen, J.S., Tice, B.C., Brenner-Morton, S., and Jessell, T.M. (2005). A Hox regulatory network establishes motor neuron pool identity and target-muscle connectivity. Cell *123*, 477–491.
25. Mercader, N., Leonardo, E., Azpiazu, N., Serrano, A., Morata, G., Martínez, C., and Torres, M. (1999). Conserved regulation of proximodistal limb axis development by Meis1/Hth. Nature *402*, 425–429.
26. Barber, B.A., Liyanage, V.R., Zachariah, R.M., Olson, C.O., Bailey, M.A., and Rastegar, M. (2013). Dynamic expression of MEIS1 homeoprotein in E14.5 forebrain and differentiated forebrain-derived neural stem cells. Ann. Anat. *195*, 431–440.
27. Spieler, D., Kaffe, M., Knauf, F., Bessa, J., Tena, J.J., Giesert, F., Schormair, B., Tilch, E., Lee, H., Horsch, M., et al. (2014). Restless legs syndrome-associated intronic common variant in Meis1 alters enhancer function in the developing telencephalon. Genome Res. *24*, 592–603.
28. Catoire, H., Dion, P.A., Xiong, L., Amari, M., Gaudet, R., Girard, S.L., Noreau, A., Gaspar, C., Turecki, G., Montplaisir, J.Y., et al. (2011). Restless legs syndrome-associated MEIS1 risk variant influences iron homeostasis. Ann. Neurol. *70*, 170–175.
29. Xiong, L., Catoire, H., Dion, P., Gaspar, C., Lafrenière, R.G., Girard, S.L., Levchenko, A., Rivière, J.B., Fiori, L., St-Onge, J., et al. (2009). MEIS1 intronic risk haplotype associated with restless legs syndrome affects its mRNA and protein expression levels. Hum. Mol. Genet. *18*, 1065–1074.
30. Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J., and de Bakker, P.I. (2008). SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics *24*, 2938–2939.
31. Torgerson, D.G., Capurso, D., Mathias, R.A., Graves, P.E., Hernandez, R.D., Beaty, T.H., Bleecker, E.R., Raby, B.A., Meyers, D.A., Barnes, K.C., et al. (2012). Resequencing candidate genes implicates rare variants in asthma susceptibility. Am. J. Hum. Genet. *90*, 273–281.
32. Trynka, G., Hunt, K.A., Bockett, N.A., Romanos, J., Mistry, V., Szperl, A., Bakker, S.F., Bardella, M.T., Bhaw-Rosun, L., Castillejo, G., et al.; Spanish Consortium on the Genetics of Coeliac Disease (CEGEC); PreventCD Study Group; Wellcome Trust Case Control Consortium (WTCCC) (2011). Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat. Genet. *43*, 1193–1201.
33. Sarparanta, J., Jonson, P.H., Golzio, C., Sandell, S., Luque, H., Screen, M., McDonald, K., Stajich, J.M., Mahjneh, I., Vihola, A., et al. (2012). Mutations affecting the cytoplasmic functions of the co-chaperone DNAJB6 cause limb-girdle muscular dystrophy. Nat. Genet. *44*, 450–455, S1–S2.
34. Holm, H., Gudbjartsson, D.F., Sulem, P., Masson, G., Helgadottir, H.T., Zanon, C., Magnusson, O.T., Helgason, A., Saemundsdottir, J., Gylfason, A., et al. (2011). A rare variant in MYH6 is associated with high risk of sick sinus syndrome. Nat. Genet. *43*, 316–320.
35. Raychaudhuri, S., Iartchouk, O., Chin, K., Tan, P.L., Tai, A.K., Ripke, S., Gowrisankar, S., Vemuri, S., Montgomery, K., Yu, Y., et al. (2011). A rare penetrant mutation in CFH confers

high risk of age-related macular degeneration. Nat. Genet. *43*, 1232–1236.

36. Diogo, D., Kurreeman, F., Stahl, E.A., Liao, K.P., Gupta, N., Greenberg, J.D., Rivas, M.A., Hickey, B., Flannick, J., Thomson, B., et al.; Consortium of Rheumatology Researchers of North America; Rheumatoid Arthritis Consortium International (2013). Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis. Am. J. Hum. Genet. *92*, 15–27.
37. Johansen, C.T., Wang, J., Lanktree, M.B., Cao, H., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., et al. (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat. Genet. *42*, 684–687.
38. Cruchaga, C., Karch, C.M., Jin, S.C., Benitez, B.A., Cai, Y., Guerreiro, R., Harari, O., Norton, J., Budde, J., Bertelsen, S., et al.; UK Brain Expression Consortium; Alzheimer's Research UK Consortium (2014). Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature *505*, 550–554.
39. Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., et al.; NHLBI Exome Sequencing Project (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature *493*, 216–220.
40. Kiezun, A., Garimella, K., Do, R., Stitziel, N.O., Neale, B.M., McLaren, P.J., Gupta, N., Sklar, P., Sullivan, P.F., Moran, J.L., et al. (2012). Exome sequencing and the genetic basis of complex traits. Nat. Genet. *44*, 623–630.
41. Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., et al.; Broad GO; Seattle GO; NHLBI Exome Sequencing Project (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science *337*, 64–69.
42. Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., St Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S.A., Fraser, D., et al. (2012). An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science *337*, 100–104.
43. Zaghloul, N.A., Liu, Y., Gerdes, J.M., Gascue, C., Oh, E.C., Leitch, C.C., Bromberg, Y., Binkley, J., Leibel, R.L., Sidow, A., et al. (2010). Functional analyses of variants reveal a significant role for dominant negative and common alleles in oligogenic Bardet-Biedl syndrome. Proc. Natl. Acad. Sci. USA *107*, 10602–10607.

3.1.3 Rare variants shaping the clinical phenotype

Schulte EC, Kurz A, Alexopoulos P, Hampel H, Peters A, Gieger C, Rujescu D, Diehl-Schmid J, Winkelmann J. Excess of rare coding variants in PLD3 in late but not early-onset Alzheimer's disease. Hum Genome Var 2015, 2:14028. PMID: 27081517. **(IF 2019: 1.3)**

Schulte EC, Mollenhauer B, Zimprich A, Bereznai B, Lichtner P, Haubenberger D, Pirker W, Brücke T, Molnar MJ, Peters A, Gieger C, Trenkwalder C, Winkelmann J. Variants in eurkaryotic translation initiation factor 4G1 in sporadic Parkinson's disease. Neurogenetics 2012, 13: 281-285. PMID: 22707335 **(IF 2012: 3.4)**

Even if not causal, rare variants shape the clinical phenotype of neuropsychiatric disorders. This can mean that specific genes or variants only play a role in disease modification in specific sets of individuals. The two studies described here examined the contribution of rare genetic variation in two specific genes—*PLD3*(Cruchaga et al., 2014) and *EIF4G1*(Chartier-Harlin et al., 2011)—that had been implicated very recently prior to the depicted projects, in AD and PD, respectively. The two projects sought to both independently replicate and refine the role of rare genetic variants in *PLD3* and *EIF4G1* in disease pathogenesis.

PLD3 had been identified as a late-onset familial AD gene in family-based WES(Cruchaga et al., 2014). To understand the extent to which rare variants in *PLD3* also contribute to early- and late-onset AD in Central Europe, we screened the coding regions plus intron/exon boundaries of *PLD3* for rare variants (MAF < 5%) in 1,089 individuals with AD, 182 individuals with frontotemporal lobar degeneration (FTLD) and 1,456 population controls. 32 such variants were identified. Burden testing revealed an excess of rare variants in individuals with late- but not early-onset AD (p=0.034; $\chi 2$ test; OR=1.46)(Schulte et al., 2015b), arguing that rare variants in *PLD3* may contribute uniquely to late- but not early-onset AD in families as well as in sporadic cases and that this contribution is not population-specific.

The second study centered around rare variants in *EIF4G1* in individuals with PD(Schulte et al., 2012). Several such variants had been implicated in multi-incident families with autosomal-dominant PD by means of linkage analyses. The frequency of four of five previously described *EIF4G1* variants plus seven rare non-synonymous variants identified in targeted resequencing of all 33 exons of *EIF4G1* in 376 individuals with PD were genotyped in a total of 975 individuals with PD and 1014 KORA-AGE

controls. Genotyping was carried out by MALDI-TOF mass spectrometry on the Sequenom® platform. None of the rare variants were statistically significantly associated with PD. The previously reported variants were either not found in our sample or they were present at equal frequencies in case and control individuals (c.2056G > T (p.Gly686Cys)) or only in controls (c.3614G > A (p.Arg1205His)). Especially the latter had previously only been found in individuals with PD and reported to be potentially causal for PD.

Our study was one of the first of numerous studies (e.g. (Deng et al., 2015; Huttenlocher et al., 2015; Lesage et al., 2012; Nichols et al., 2015; Siitonen et al., 2013; Zhao et al., 2013)) to question the causative role of this variant and *EIF4G1* as a whole with regard to PD. Because the *EIF4G1* locus naturally holds a lot of genetic variation, much larger case-control samples than those included in the original study, our study, or any of the other replication or follow-up studies published to date will be needed to conclusively answer the question of a potential involvement of *EIF4G1* and, along with it the mechanism of translation initiation, in PD. This situation places a spotlight on the need for high-quality replication studies in the field of genetics of neuropsychiatric disorders, especially when it comes to rare genetic variants. Rare genetic variants are more likely to be population-specific and subject to founder effects than common or intermediate frequency variants, making careful, independent replications and cautious conclusions on causality vital.

OPEN

Citation: Human Genome Variation (2015) 2, 14028; doi:10.1038/hgv.2014.28

www.nature.com/hgv

DATA REPORT

Excess of rare coding variants in *PLD3* in late- but not early-onset Alzheimer's disease

Eva C Schulte[1,2], Alexander Kurz[3], Panagiotis Alexopoulos[3], Harald Hampel[4,5], Annette Peters[6], Christian Gieger[7], Dan Rujescu[8], Janine Diehl-Schmid[3] and Juliane Winkelmann[1,2,9,10,11]

Recently, mutations in phospholipase D3 (*PLD3*) were reported in late-onset Alzheimer's disease (AD). By screening the coding regions of *PLD3* for variants in a European cohort of 1,089 AD cases, 182 individuals with frontotemporal lobar degeneration and 1,456 controls, we identified 32 variants with a minor allele frequency $< 5\%$ and observed an excess of rare variants in individuals with late- but not early-onset AD ($P = 0.034$, χ^2-test; odds ratio = 1.46).

Human Genome Variation (2015) **2,** 14028; doi:10.1038/hgv.2014.28; published online 9 January 2015

Genome-wide association studies and linkage analyses have identified at least 25 genes associated with sporadic and familial Alzheimer's disease (AD).[1] These genes include classical genetic factors contributing to familial, early-onset forms of AD, such as the β-amyloid precursor protein and the presenilins (PSEN1 and PSEN2),[2–5] as well as several more recently discovered genes that harbor common variants associated with increased risk of late-onset AD (LOAD).[6–9] Together, these genes explain ~61% of the population attributable risk of AD,[8,10] and novel genetic factors continue to be revealed. Most recently, rare variants in phospholipase D3 (*PLD3*) were implicated in late-onset familial and sporadic AD both by family-based whole-exome sequencing and by genotyping and gene-based resequencing.[11] Here we assessed the role of *PLD3* variants in central European AD and frontotemporal lobar degeneration (FTLD) patients, particularly investigating the role of *PLD3* variants in early-onset AD (EOAD).

Using Idaho LightScanner high-resolution melting curve analysis (Biofire Diagnostics, Inc., Salt Lake City, UT, USA), we screened the 13 coding exons and exon–intron boundaries (±10 bp) of *PLD3* in 1089 German AD case subjects (75.6 ± 18.6 years, 59.3% female, including 139 cases with an age of onset younger than 65 (61.5 ± 5.5 years, 53.2% female)), 138 German FTLD cases (63.7 ± 8.1 years, 42.0% female) and 1,456 general population controls belonging to the KORA general population cohort[12] (58.3 ± 12.0 years, 48.2% female) based in southern Germany. When altered melting patterns suggested variants, Sanger sequencing ensued to identify the underlying genetic alteration. Gene-based burden tests (cohort allelic sum test) and single-variant association tests were performed using χ^2 analysis.

All subjects were diagnosed according to the NINCDS-ADRDA criteria or the revised Neary *et al.*[13] criteria as appropriate by a senior psychiatrist specializing in dementias. Ethics review board approval and participants' written informed consent were obtained before the initiation of the study.

We observed a total of 32 variants with minor allele frequency (MAF) $< 5\%$ within the coding regions ± 10 bp, including: 3 near-splice variants, 1 deletion, 9 synonymous and 16 non-synonymous variants and 3 newly introduced stop codons (Figure 1, Table 1). Eight of the coding variants (27.5%) had been observed previously.[11] Notably, we found the variant most significantly associated with the AD phenotype in the previous study,[11] *PLD3* p.Val232Met (rs145999145), more frequently in controls ($n = 6$; MAF = 0.20%; 55.6 ± 12.6 years; 33% older than 65 years) than in AD patients ($n = 1$; MAF = 0.05%; 66 years). The single AD individual harboring the p.Val232Met variant presented with AD at age 64 and reported that her mother had also suffered from dementia. Moreover, the synonymous variant (*PLD3* p.Ala442Ala; rs4819) that had been significantly associated with LOAD in the original publication[11] did not show association with the AD phenotype in our sample in either LOAD or EOAD cases (Table 1). However, we did identify significant associations between two different variants in *PLD3* and AD in our sample: *PLD3* p.Ile364Ile (rs51787324; $P_{nominal} = 3.0 \times 10^{-8}$; $P_{corrected} = 9.6 \times 10^{-7}$; χ^2 test; odds ratio (OR) = 63.50 (95% confidence interval (CI): 3.85–1,046.15)) in both the LOAD-only and combined AD sample and *PLD3* p.Asn284Ser (rs200274020; $P_{nominal} = 5.0 \times 10^{-6}$; $P_{corrected} = 1.6 \times 10^{-5}$; χ^2 test; OR = 52.66 (95% CI: 2.52–1,099.83)). To date, these variants have only been found in individuals with either LOAD[11] or EOAD (this study).

A priori power calculation based on published variant frequencies and effect sizes[11] using the Purcell power calculator[14] suggested that a sample size of 1,089 AD cases and an equal number of general population controls would be sufficient to reach 100% power to detect an excess of rare variants with a MAF $< 5\%$ under an autosomal dominant model and 36% power for recessive effects. Accordingly, we also performed gene-based burden tests for variants with a KORA-derived MAF $< 5\%$. Interestingly, although we observed an excess (Figure 1,

[1]Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany; [2]Institut für Humangenetik, Helmholtz Zentrum München, Neuherberg, Germany; [3]Klinik und Poliklinik für Psychiatrie und Psychotherapie, Klinikum rechts der Isar, Technische Universität München, Munich, Germany; [4]AXA Research Fund; [5]Sorbonne Universités, Université Pierre et Marie Curie, Paris 06, Institut de la Mémoire et de la Maladie d'Alzheimer (IM2A) & Institut du Cerveau et de la Moelle épinière (ICM), Département de Neurologie, Hôpital de la Pitié-Salpétrière, Paris, France; Université Pierre et Marie Curie-Paris 6, AP-HP, Hôpital de la Salpêtrière, Paris, France; [6]Institute of Epidemiology II, Helmholtz Zentrum München, Neuherberg, Germany; [7]Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany; [8]Klinik für Psychiatrie, Universität Halle, Halle, Germany; [9]Institut für Humangenetik, Technische Universität München, Munich, Germany; [10]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany and [11]Department of Neurology and Neurosciences, Stanford University, Palo Alto, CA, USA.
Correspondence: J Winkelmann (winkelmann@lrz.tum.de or winkelmann@stanford.edu)
Received 5 September 2014; revised 28 September 2014; accepted 28 October 2014

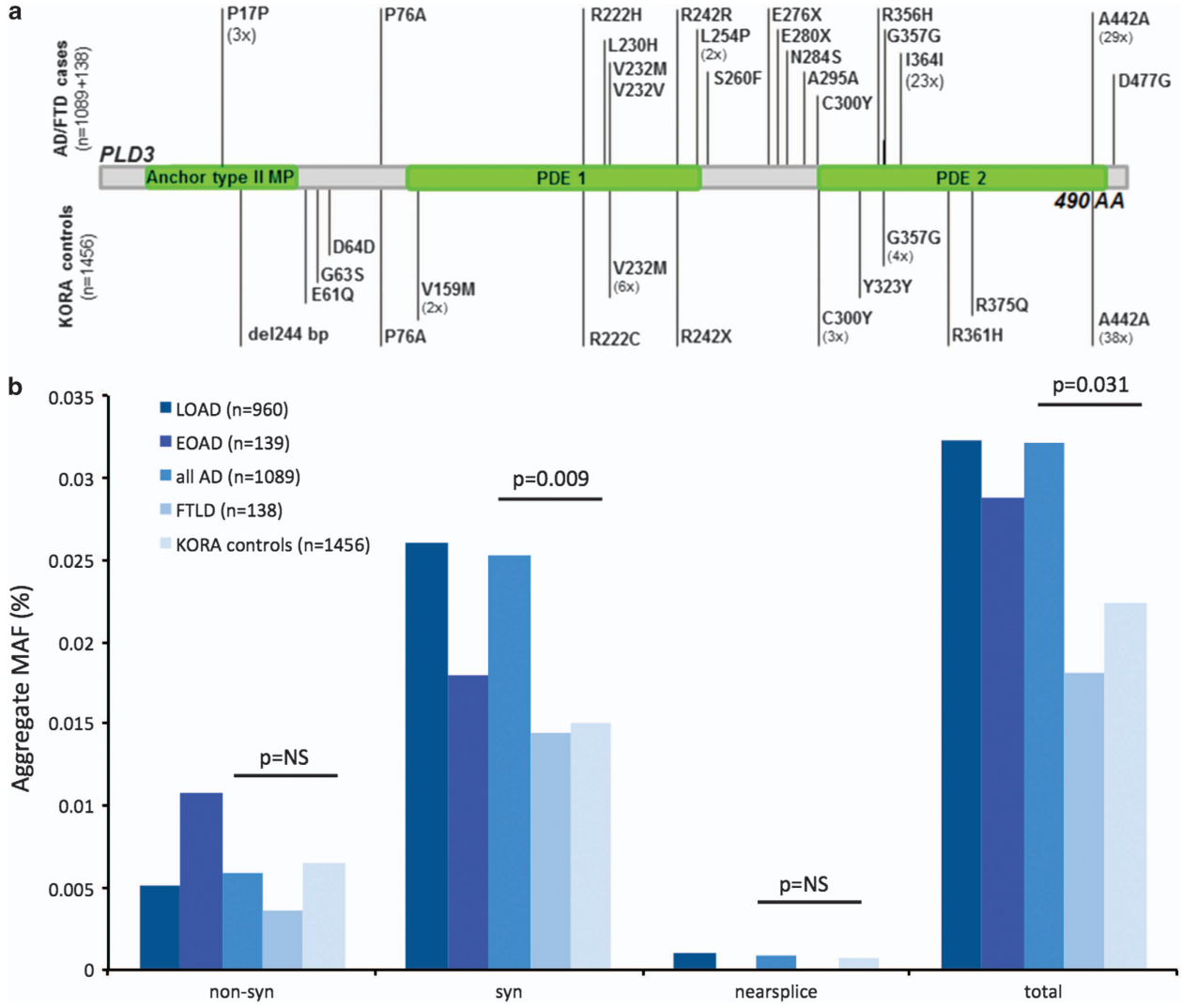


Figure 1. Rare coding variants in *PLD3* in AD, FTLD and control subjects. (**a**) Schematic representation of rare coding variants identified in *PLD3* in AD and FTLD case subjects (above the gene) and KORA general population controls (below the gene). Numbers in parentheses indicate variant counts. If no number is given, variants were identified only once. (**b**) Aggregate minor allele frequencies (MAFs) for variants of different classes in the different subsamples. *P*-values represent gene-based burden tests for rare variants in *PLD3* (MAF < 5%) based on variant counts (Supplementary Table S1) and calculated using χ^2 test statistics. AA, amino acids; AD, Alzheimer's disease; EOAD, EOAD, early-onset AD; FTLD, frontotemporal lobar degeneration; NS, not significant.

Supplementary Tables S1–S4) of rare variants of all classes in LOAD ($P = 0.034$, χ^2 test; OR = 1.46 (95% CI: 1.02–2.07)) and LOAD+EOAD combined ($P = 0.031$, χ^2 test; OR = 1.45 (95% CI: 1.03–2.05)), the same was not observed for EOAD alone ($P = 0.54$, χ^2 test; OR = 1.28 (95% CI: 0.60–2.73)). This excess of rare variants included mainly synonymous variants (LOAD: $P = 0.007$, χ^2 test, OR = 1.74 (95% CI: 1.15–2.62); LOAD+EOAD combined: $P = 0.009$, χ^2 test, OR = 1.68 (95% CI: 1.13–2.52)). In the EOAD cases, however, non-synonymous variants were encountered at a MAF = 1.1%, which was almost twice as frequent as in either controls (MAF = 0.7%) or LOAD (MAF = 0.5%; Figure 1, Supplementary Table S2). This result fell short of statistical significance, possibly because of the small number of EOAD cases ($n = 139$) in our sample.

Because an overlap in the genetic architecture of different dementia syndromes as well as neurodegenerative conditions[15–17] has been described, we also assessed the contributions of rare variants in *PLD3* to the genetic framework of our FTLD samples. Rare variants in *PLD3* were found at an equal or lower frequency in FTLD case subjects relative to controls (Figure 1, Supplementary Tables S1–S3), making a large-scale contribution of rare genetic variants in *PLD3* to the genetics of FTLD unlikely.

Both our data as well as previously published data[11] point to a significant contribution of a number of different rare synonymous variants in *PLD3* to the LOAD phenotype. This is especially interesting in light of the fact that although over 50 human diseases associated with synonymous mutations have been reported to date, few examples exist with regard to neuropsychiatric disorders.[18] Functional assays have demonstrated that *PLD3* can directly alter β-amyloid precursor protein processing and β-amyloid formation,[11] and the apparent lack of contribution of rare *PLD3* variants in another neurodegenerative phenotype (that is, FTLD) indirectly supports this notion. It is known that *PLD3* p.Ala442Ala is associated with lower expression of total *PLD3* mRNA as well as lower levels of exon 11-containing transcripts. Whether a similar mechanism could be implicated in the association between *PLD3* p.Ile364Ile and, to a much lesser extent, p.Pro17Pro remains to be elucidated. Human Splicing Finder[19] predicts that p.Ile364Ile ablates an enhancer, whereas p.Pro17Pro generates a

Table 1. Rare variants (MAF < 5%) identified by high-resolution melting curve analysis is 1,089 AD case subjects, 138 FTD case subjects and 1,456 KORA general population controls

Gene	*Exon*	*Genome position*	*Variant*	*AA subst.*	*LOAD (n = 960)*	*EOAD (n = 139)*	*FTD (n = 138)*	*Controls (n = 1456) (dbSNP139)*	*Reference cases/controls*	*MAF AD*	$P_{nominal}$	*Class*	*PolyPhen2 (MAF)*	*1000G (MAF)*	*ESP (cases/ controls)*	*Cruchaga et al.*[11]
PLD3	Int 4	chr19:40872671	c.103-9T>A					1	Novel	0.00%/0.03%	NS	Nearsplice	N/A	N/F	N/A	N/A
PLD3	Int 5	chr19:40873599	c.246-4 C>G					1	Novel	0.00%/0.03%	NS	Nearsplice	N/A	N/F NA	N/A	N/A
PLD3	Int 9	chr19:40877785	c.879+5G>A		1 (homo)				Novel	0.05%/0.00%	NS	Nearsplice	N/A	N/F	N/A	N/A
PLD3	4	chr19:40872540	c.51 C>T	P17P	3				rs200094590	0.13%/0.00%	0.03	Syn	N/A	0.04%	0.02%	N/F
PLD3	4/5	chr19:40872502_745	del 244 bp	del exons 4 & 5				1	novel	0.00%/0.03%	NS	Deletion	N/a	N/F	N/F	N/F
PLD3	5	chr19:40872758	c.181G>C	E61Q				1	Novel	0.00%/0.03%	NS	Non-syn	Benign	N/F	N/F	N/F
PLD3	5	chr19:40872764	c.187G>A	G63S				1	rs142070038	0.00%/0.03%	NS	Non-syn	Benign	0.09%	0.16%	0.08%/0.12%
PLD3	5	chr19:40872769	c.192 C>T	D64D				1	Novel	0.00%/0.03%	NS	Syn	N/A	N/F	N/F	N/F
PLD3	5	chr19:40872803	c.226 C>G	P76A	1			1	rs138674695	0.05%/0.03%	NS	Non-syn	Benign	0.03%	0.03%	0.08%/0.00%
PLD3	7	chr19:40875860	c.475G>A	V159M				2	rs374184677	0.00%/0.06%	NS	Non-syn	Deleterious	N/F	0.02%	N/F
PLD3	8	chr19:40876130	c.664 C>T	R222C				1	rs200077325	0.00%/0.03%	NS	Non-syn	Deleterious	0.70%	N/F	N/F
PLD3	8	chr19:40876131	c.665G>A	R222H			1		Novel	0.00%/0.00%	0.001	Non-syn	Deleterious	N/F	N/F	N/F
PLD3	9	chr19:40877590	c.689T>A	L230H	1				Novel	0.05%/0.00%	NS	Non-syn	Deleterious	N/F	N/F	N/F
PLD3	9	chr19:40877595	c.694T>A	V232M		1		6	rs145999145	0.05%/0.20%	NS	Non-syn	Deleterious	0.30%	0.48%	1.36%/0.79%
PLD3	9	chr19:40877597	c.696G>T	V232V	1				Novel	0.05%/0.00%	NS	Syn	N/A	N/F	0.02%	N/F
PLD3	9	chr19:40877625	c.724 C>T	R242X				1	Novel	0.00%/0.03%	NS	Stop	N/A	N/F	N/F	N/F
PLD3	9	chr19:40877627	c.726 AA>GG	R242R	1				Novel	0.05%/0.00%	NS	Syn	N/A	N/F	N/F	N/F
PLD3	9	chr19:40877662	c.761TT>CC	L254P	1 (homo)				Novel	0.09%/0.00%	NS	Non-syn	Deleterious	N/F	N/F	N/F
PLD3	9	chr19:40877680	c.779 C>T	S260F	1				Novel	0.05%/0.00%	NS	Non-syn	Deleterious	N/F	N/F	N/F N/F
PLD3	9	chr19:40877727	c.826G>T	E276X	1				Novel	0.05%/0.00%	NS	Stop	N/A	N/F	N/F	N/F
PLD3	9	chr19:40877739	c.838G>T	E280X	1				Novel	0.05%/0.00%	NS	Stop	N/A	N/F	N/F	N/F
PLD3	9	chr19:40877752	c.851A>G	N284S		2			rs200274020	0.09%/0.00%	5.0×10^{-6}	non-syn	Deleterious	0.06%	N/F	0.02%/0.00%
PLD3	10	chr19:40880393	c.885G>A	A295A	1				Novel	0.05%/0.00%	NS	syn	N/A	N/F	N/F	N/F
PLD3	10	chr19:40880475	c.899G>A	C300Y	1			3	rs146083475	0.05%/0.1%	NS	Non-syn	Deleterious	N/F	0.09%	0.10%/0.04%
PLD3	10	chr19:40880477	c.969 C>T	Y323Y				1	rs369989744	0.00%/0.03%	NS	Syn	N/A	N/F	0.02%	N/F
PLD3	11	chr19:40882563	c.1067G>A	R356H	1				Novel	0.05%/0.00%	NS	Non-syn	Deleterious	N/F	N/f	N/F
PLD3	11	chr19:40882536	c.1071 C>T	G357G	1			4	rs147721393	0.05%/0.14%	NS	Syn	N/A	N/F	0.10%	N/F
PLD3	11	chr19:40882578	c.1082G>A	R361H				1	Novel	0.00%/0.03%	NS	Non-syn	Deleterious	N/F	N/F	N/F
PLD3	11	chr19:40882588	c.1092 C>T	I364I	15 (het)/4 (homo)				rs57187324	1.05%/0.00%	3.0×10^{-8}	Syn	N/A	N/F	0.96%	4.62%/2.33%[a]
PLD3	11	chr19:40882620	c.1124G>A	R375Q				1	rs374352480	0.00%/0.03%	NS	Non-syn	Deleterious	N/F	0.00%	N/F
PLD3	13	chr19:40883933	c.1326G>A	A442A	20	5	4	39	rs4819	1.45%/1.30%	NS	Syn	N/A	0.95%	1.59%	2.09%/0.90%
PLD3	13	chr19:40884037	c.1430G>A	D477G	1				rs141721330	0.05%/0.00%	NS	Non-syn	Benign	0.04%	0.02%	0.02%/0.02%

Abbreviations: AA, amino acid; AD, Alzheimer's disease; EOAD, early-onset AD; ESP, Exome Sequencing Project; FTD, frontotemporal dementia; subst, substitution; Int, intron; LOAD, late-onset AD; MAF, minor allele frequency; N/A, not available; N/F, not found; Non-syn, non-synonymous; NS, nonsignificant; PLD3, phospholipase D3; syn, synonymous. [a]MAFs derived from the African-American sample because the variant was not identified in individuals of European ancestry by Cruchaga *et al.*[11]

novel enhancer site. However, experimental evidence is lacking to date. In this context, although we did not observe a similar role of rare synonymous variants in EOAD cases, it seems noteworthy that we identified several non-synonymous *PLD3* variants in our small EOAD sample. One could hypothesize that non-synonymous variants with possibly larger effects might also contribute to this comparatively more severe phenotype. However, given the dearth of statistical significance to support this assumption, it currently remains a hypothesis at best.

When considering the individual variants identified in our screening and their contribution to AD genetics, an additional caveat would have to be that association *P*-values and ORs appear inflated possibly due to the small total number of variants, the small sample sizes (for EOAD) or the possible existence of unaccounted population substructure. Conversely, because we used general population controls, we are unable to exclude the possibility that some of the controls have or will develop AD, thus underestimating the calculated effect sizes. ORs between 50 and 60 typically suggest (near) monogenic disease. However, OR estimates for the total rare genetic variation in *PLD3* and its contribution to the AD phenotype (that is, ORs of ~ 1.5–2.0) seem more realistic because it is likely that these variants contribute to AD risk but are not causal by themselves.

In summary, our data corroborate the role of rare variants in *PLD3* and further highlight the significant contribution of rare synonymous variants in this gene to the genetic architecture of LOAD. Interestingly, the association between *PLD3* and LOAD was largely driven by variants not significantly associated with the phenotype in the original study,[11] whereas the individual variants showing significant associations in the original study could be replicated directly. While rare variants overall or synonymous variants alone do not seem to play a large role in bringing about EOAD, the role for non-synonymous *PLD3* variants in EOAD remains open to debate.

HGV DATABASE

The relevant data from this Data Report are hosted at the Human Genome Variation Database at http://dx.doi.org/10.6084/m9.figshare.hgv.526, http://dx.doi.org/10.6084/m9.figshare.hgv.530, http://dx.doi.org/10.6084/m9.figshare.hgv.532, http://dx.doi.org/10.6084/m9.figshare.hgv.534, http://dx.doi.org/10.6084/m9.figshare.hgv.536, http://dx.doi.org/10.6084/m9.figshare.hgv.538, http://dx.doi.org/10.6084/m9.figshare.hgv.540, http://dx.doi.org/10.6084/m9.figshare.hgv.542, http://dx.doi.org/10.6084/m9.figshare.hgv.544, http://dx.doi.org/10.6084/m9.figshare.hgv.546, http://dx.doi.org/10.6084/m9.figshare.hgv.548, http://dx.doi.org/10.6084/m9.figshare.hgv.550, http://dx.doi.org/10.6084/m9.figshare.hgv.552, http://dx.doi.org/10.6084/m9.figshare.hgv.554, http://dx.doi.org/10.6084/m9.figshare.hgv.556, http://dx.doi.org/10.6084/m9.figshare.hgv.558, http://dx.doi.org/10.6084/m9.figshare.hgv.560, http://dx.doi.org/10.6084/m9.figshare.hgv.562, http://dx.doi.org/10.6084/m9.figshare.hgv.564, http://dx.doi.org/10.6084/m9.figshare.hgv.566.

COMPETING INTERESTS

The authors declare no conflict of interest.

REFERENCES

1 Hardy J, Bogdanovic N, Winblad B, Portelius E, Andreasen N, Cedazo-Minguez A *et al.* Pathways to Alzheimer's disease. *J Internal Med* 2014; **275**: 296–303.
2 Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M *et al.* Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* 1995; **375**: 754–760.
3 Chartier-Harlin MC, Crawford F, Houlden H, Warren A, Hughes D, Fidani L *et al.* Early-onset Alzheimer's disease caused by mutations at codon 717 of the beta-amyloid precursor protein gene. *Nature* 1991; **353**: 844–846.
4 Rogaev EI, Sherrington R, Rogaeva EA, Levesque G, Ikeda M, Liang Y *et al.* Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature* 1995; **376**: 775–778.
5 Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* 1991; **349**: 704–706.
6 Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML *et al.* Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature Genet* 2009; **41**: 1088–1093.
7 Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM *et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nature Genet* 2011; **43**: 429–435.
8 Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nature Genet* 2011; **43**: 436–441.
9 Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genet* 2013; **45**: 1452–1458.
10 Medway C, Morgan K. Review: the genetics of Alzheimer's disease; putting flesh on the bones. *Neuropathol Appl Neurobiol* 2014; **40**: 97–105.
11 Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, Guerreiro R *et al.* Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* 2014; **505**: 550–554.
12 Wichmann HE, Gieger C, Illig TGroup MKS. KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* 2005; **67**(Suppl 1): S26–S30.
13 Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S *et al.* Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* 1998; **51**: 1546–1554.
14 Purcell S, Cherny SS, Sham PC.. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003; **19**: 149–150.
15 Harms M, Benitez BA, Cairns N, Cooper B, Cooper P, Mayo K *et al.* C9orf72 hexanucleotide repeat expansions in clinical Alzheimer disease. *JAMA Neurol* 2013; **70**: 736–741.
16 Hutton M, Lendon CL, Rizzu P, Baker M, Froelich S, Houlden H *et al.* Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* 1998; **393**: 702–705.
17 Pastor P, Ezquerra M, Munoz E, Martí MJ, Blesa R, Tolosa E *et al.* Significant association between the tau gene A0/A0 genotype and Parkinson's disease. *Ann Neurol* 2000; **47**: 242–245.
18 Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C. Exposing synonymous mutations. *Trends Genet* 2014; **30**: 308–321.
19 Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; **37**: e67.

Supplemental Information for this article can be found on the *Human Genome Variation* website (http://www.nature.com/hgv)

Neurogenetics (2012) 13:281–285
DOI 10.1007/s10048-012-0334-9

SHORT COMMUNICATION

Variants in *eukaryotic translation initiation factor 4G1* in sporadic Parkinson's disease

Eva C. Schulte · Brit Mollenhauer · Alexander Zimprich · Benjamin Bereznai · Peter Lichtner · Dietrich Haubenberger · Walter Pirker · Thomas Brücke · Maria J. Molnar · Annette Peters · Christian Gieger · Claudia Trenkwalder · Juliane Winkelmann

Received: 5 March 2012 / Accepted: 30 May 2012 / Published online: 16 June 2012

Abstract Recently, mutations in *eukaryotic translation initiation factor 4G1* (*EIF4G1*) were reported as a rare cause of familial Parkinson's disease (PD). We screened the 33 exons of *EIF4G1* by high-resolution melting curve analysis for variants in our Central European cohort of 376 PD cases. Variant frequency was assessed in a total of 975 PD cases and 1,014 general population controls. Eight novel nonsynonymous and four synonymous variants were identified. In our cohort, novel and previously identified nonsynonymous variants were very rare. Although it is possible that our general population controls also comprise individuals who have or could develop PD in the future, the presence of the original mutation (*EIF4G1* p.Arg1205 His) in three controls only, raises questions about the causality of this variant with regard to PD.

Keywords Genetics · Rare variants · Parkinson's disease · *EIF4G1*

Introduction

Genome-wide association studies and linkage analyses have identified at least 19 genes associated with idiopathic Parkinson`s disease (PD). Most recently, variants in *eukaryotic translation initiation factor 4G1* (*EIF4G1*) were implicated in familial PD, linking dysfunctional mRNA translation initiation to PD pathogenesis. [1] Here, we assess the role of *EIF4G1* variants in our Central European PD cohort.

E. C. Schulte · J. Winkelmann (✉)
Neurologische Klinik und Poliklinik, Klinikum rechts der Isar,
Technische Universität München,
Ismaningerstr. 22,
81675 Munich, Germany
e-mail: winkelmann@lrz.tum.de

E. C. Schulte · P. Lichtner · J. Winkelmann
Institut für Humangenetik, Helmholtz Zentrum München,
Munich, Germany

B. Mollenhauer · C. Trenkwalder
Paracelsus Elena Klinik,
Kassel, Germany

B. Mollenhauer · C. Trenkwalder
Neurologische Klinik, Georg August Universität,
Göttingen, Germany

A. Zimprich · D. Haubenberger · W. Pirker
Department of Neurology, Medical University of Vienna,
Vienna, Austria

B. Bereznai · M. J. Molnar
Center for Molecular Neurology, Department of Neurology,
Semmelweis University,
Budapest, Hungary

P. Lichtner · J. Winkelmann
Institut für Humangenetik, Technische Universität München,
Munich, Germany

T. Brücke
Department of Neurology, Wilhelminenspital,
Vienna, Austria

A. Peters
Institute of Epidemiology II, Helmholtz Zentrum München,
Munich, Germany

C. Gieger
Institute of Genetic Epidemiology, Helmholtz Zentrum München,
Munich, Germany

Methods

Using Idaho® melting curve analysis, we screened the 33 exons and exon–intron boundaries of *EIF4G1* in a discovery sample of 376 German PD patients (71.1±9.4 years, 31.6 % female). When altered melting patterns suggested variants, Sanger sequencing ensued. To assess variant frequency, we genotyped the novel as well as four of the five variants previously described in PD (*EIF4G1* c.1505C > T (p.Ala502Val), c.2056G > T (p.Gly686Cys), c.3490A > C (p.Ser1164Arg), and c.3614G > A (p.Arg1205His)) [1] in 975 familial and sporadic PD cases from Austria (n=486, 58.7±11.3 years, 35.4 % female, family history known in n=413, 33.4 % thereof positive for PD in a first or second degree relative), Germany (n=450, 376 of which comprised the discovery sample, 70.2±9.7 years, 32.2 % female, family history known in n=105, 24.7 % thereof positive for PD in a first or second degree relative), and Hungary (n=39, 50.4±10.8 years, 53.9 % female, family history known in n=39, 28.2 % thereof positive for PD in a first- or second-degree relative) and 1,014 general population controls belonging to the KORA-AGE cohort (76.0±6.6 years, 50.1 % female) [2] by matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry on the Sequenom platform. The KORA-AGE cohort is a follow-up study of the initial surveys, enriched for older individuals. Individuals known to take dopaminergic medication where excluded from the control sample. All individuals included in this study were Caucasian. German and Austrian PD samples and KORA controls originate from the same geographic region. The small number of Hungarian patients either have an early age of onset or are index patients of larger PD families and were, therefore, genotyped as well. For technical reasons, four novel variants could not be included in the genotyping assay. Haplotype analysis in carriers of the original c. 3614G > A (p.Arg1205His) variant was performed using haplotype-tagging SNPs rs4912537, rs2178403, rs2293605, rs1879244, and rs2230571 and polymorphic markers D3S3609, D3S3578, and D3S3583 by Sanger sequencing. All subjects were diagnosed according to the UK Brain Bank criteria by a senior neurologist specializing in movement disorders. Ethics review board approval and participants' written informed consent were obtained.

Results

In addition to several common and rare synonymous variants, we identified seven nonsynonymous variants, not previously reported in PD, in six individuals. These include c.47C > T (p.Pro16Leu), c.211C > T (p.Pro71Ser, rs113810947), c.953C > T (p.Thr318Ile),c.1622T > G (p.Val541Gly), c.1648G > C (p.Ala550Pro, rs111924994), c.2093G > C (p.Gly698Ala), and c.2149G > C (p.Ala717Pro, rs11396765) as well as c.1456C > T (p.Pro486Ser, rs112545306) previously reported in two individuals suffering from PD [3] (Fig. 1). Similar to the phenotype described [1], all individuals presented with classic PD with an age of onset at 64.5±5.5 years and positive response to dopaminergic therapy. Where available, family history was negative (Table 1).

Overall, the identified variants were very rare in our population. Four—c.47C > T (p.Pro16Leu), c.953C > T (p.Thr318Ile), c.1622T > G (p.Val541Gly), and c.2093G > C (p.Gly698Ala)—were validated in the PD individual in whom they were first identified but were not found in any additional PD subjects. Of these, c.953C > T (p.Thr318Ile), c.1622T > G (p.Val541Gly), and c.2093G > C (p.Gly698Ala) were not present in controls, while c.47C > T (p.Pro16Leu) was identified in three controls. Of the previously reported [1] variants, c.1505C > T (p.Ala502Val) and c.3490A > C (p.Ser1164Arg) were not seen in the 1989 individuals assessed. Five cases and three controls, on the other hand, were heterozygous for c.2056G > T (p.Gly686Cys). Surprisingly, the original mutation, c.3614G > A (p.Arg1205His), which had, so far, only been identified in PD cases [1], was only present in three controls. In the original publication, all eight PD probands heterozygous for c. 3614G > A (p.Arg1205His; out of 4,708 cases and 4,576 controls) shared the same minimal haplotype [1]. Genotyping of five haplotype-tagging SNPs and three

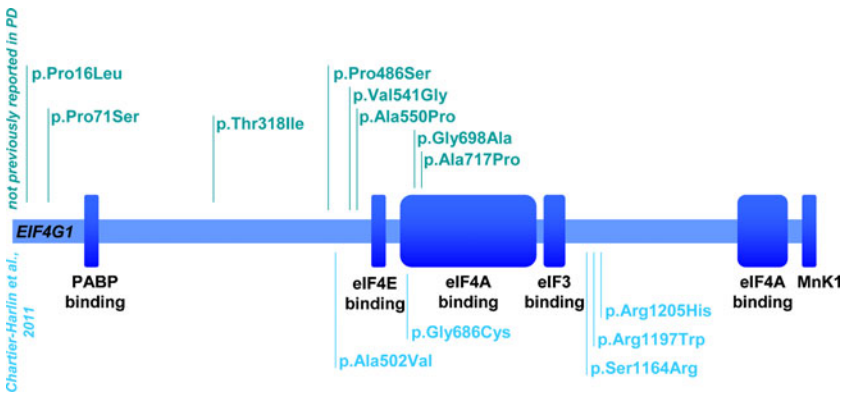


Fig. 1 *EIF4G1* sheme depicting novel and previously described [1] missense variants in individuals with PD and their relative location in relation to known and predicted functional domains. *PABP* polyadenylate binding protein, *eIF* eukaryotic translation initiation factor

Table 1 Rare *EIF4G1* variants identified in individuals with PD, clinical phenotype, and variant frequencies

EIF4G1 variant	Genomic position (hg19)	Exon	Family history	AoO	DD	IS	B	R	RT	PI	L-Dopa/DA	Frequency		Frequency NHLBI-ESP (EA)	Mutation Taster	PolyPhen2
												Cases	Controls			
p.Pro16Leu (c.47C>T)	chr3:184,033,631	1	n/a	n/a	n/a	n/a	+	+	−	++	+	1/975	3/1014	Not found	dc	n/s
p.Pro71Leu (c.211C>T)	chr3:184,035,172	4	n/a	59	4	RT	+	+	+	+	+	n/a	n/a	0.000285	Poly	prob.dam.
p.Ala239Ala (c.717A>G)	chr3:184,039,089	9										1/975	0/1014	0.000142		
p.Thr318lle (c.953C>T)	chr3:184,039,325	9	neg	62	3	B	+	+	−	+	+	1/975	0/1014	Not found	poly	prob.dam.
p.Pro486Ser (c.1456C>T)	chr3:184,039,828	9	n/a	70	1	B	+	+	+	+	+	n/a	n/a	0.00057	poly	benign
p.Val541Gly (c.1622T>G)	chr3:184,040,345	11	n/a	72	3	TR	+	+	+	+	+	1/975	0/1014	Not found	poly	poss. dam.
p.Ala550Pro (c.1648G>C)	chr3:184,040,371	11	n/a	72	3	RT	+	+	+	+	+	n/a	n/a	0.001994	poly	benign
p.Gly698Ala (c.2093G>C)	chr3:184,041,200	14	neg	65	4	B	+	+	−	+	+	1/975	0/1014	Not found	dc	prob. dam.
p.Ala717Pro (c.2149G>C)	chr3:184,041,256	14	neg	59	9	B	+	+	−	+	+	n/a	n/a	Not found	poly	benign
p.Pro992Pro (c.2971A>G)	chr3:184,043,282	20										14/975	10/1014	0.00616		
p.Val141Val (c.4251C>T)	chr3:184,049,143	29										n/a	n/a	0.0245		
p.Ala1517Ala (c.4551C>T)	chr3:184,049,807	32										n/a	n/a	0.00818		
Chartier_A502V		10										0/975	0/1014	0.000285		
Chartier_G686C		14										5/975	3/1014	Not found		
Chartier_S1164R		24										0/975	0/1014	No found		
Chartier_R1197W		24										n/a	n/a	Not found		
Chartier_R1205H		24										0/975	3/1014	Not found		

Four nonsynonymous and two synonymous variants present in our sample in addition to the *EIF4G1* variants previously identified in familial PD [1] were genotyped in 975 cases and 1,014 controls. Additional clinical information and in silico predictions of the damaging potential of the amino acid exchange as assessed by MutationTaster [5] and PolyPhen2 [6] are presented for the newly identified missense variants. Additionally, variant frequencies as found in the approximately 3,500 European American exomes found in the NHLBI exome sequencing project (NHLBI-ESP) are noted for all newly identified and previously reported [1, 3, 4]

n/a Not available, *neg* negative, *AoO* age of onset, *DD* disease duration, *IS* initial symptom, *B* bradykinesia, *R* rigor, *RT* resting tremor, *PI* postural instability, *D* dementia, *DA* dopamine agonist, *dc* disease causing, *poly* polymorphism, *n/s* not scored, *prob. dam.* probably damaging, *poss. dam.* possibly damaging

microsatellite markers indicated that two of our three c.3614G > A (p.Arg1205His) controls could share this minimal haplotype (Table 2). Seven out of the 12 variants identified in our PD cohort were also found in the approximately 3,500 European American exomes pertaining to the NHLBI exome sequencing project [4] (Table 1).

Discussion

Of the newly identified variants, c.47C > T (p.Pro16Leu), c.211C > T (p.Pro71Ser), c.953C > T (p.Thr318Ile), c.1622T > G (p.Val541Gly), and c.2093G > C (p.Gly698Ala) are predicted to damage protein structure, while c.1456C > T (p.Pro486Ser) and c.2149G > C (p.Ala717Pro) are likely functionally neutral [5, 6] (Table 1). Of these, c.2093G > C (p.Gly698Ala) emerges as the best potentially pathogenic candidate. Contrary to most other amino acids affected, the glycine in position 698 is conserved in all vertebrates. The variant, moreover, was ranked most likely to be damaging by two prediction algorithms [5, 6] and is located in the eIF3/eIF4A binding domain necessary for formation of the translation initiation complex (Fig. 1). However, caution is mandated as a nearby variant (c.2056G > T (p.Gly686Cys)), previously only found in two individuals with PD [1], was present in five cases and three controls in our much smaller sample, suggesting that population-specific effects can misconstrue frequency assessment especially with regard to rare genetic variation. Consequently, further assessment of the role of *EIF4G1* variants in PD is warranted.

Haplotype analysis in the three control subjects harboring c. 3614G > A (p.Arg1205His) supports the idea of an ancestral founder mutation. Linkage analysis and segregation in the original family [1] back pathogenicity of this variant and this is not necessarily disparaged by the presence of the variant in our controls. First, we used general population controls and it is not unlikely that some of the controls may have or may develop PD. Second, it is possible that this mutation shows incomplete penetrance or that other protective factors exist. However, the presence of c. 3614G > A (p.Arg1205His) in our control cohort could also indicate that its role in PD pathogenesis is questionable as has just now also been suggested for the *EIF4G1* p.Ala502Val variant initially also reported by Chartier-Harlin et al. [1, 3]. Overall, the *EIF4G1* locus naturally holds a lot of genetic variance [1, 3]. Accordingly, much larger case–control samples than those used in either the original [1], a follow-up [3], or our study will be necessary to answer this question.

Although not common, it still cannot be excluded that rare exonic *EIF4G1* variants of strong effect could play a causative role in PD in rare cases. And their study is important as they can provide significant clues in understanding disease mechanism. This idea is supported by the fact that *LRRK2*, which harbors both rare and common genetic variation contributing to PD development [7, 8], has recently also been implicated in dysfunctional mRNA translation initiation [9].

Table 2 Haplotype of *EIF4G1* p.Arg1205His carriers

Marker ID	KORA_315	KORA_330	KORA_944
D3S3609	163/179	163/167	163/165
rs4912537	***T***	***T****/C*	***T****/C*
rs2178403	***G***	***G****/A*	***G***
rs2293605	***T****/C*	***T****/C*	***T****/C*
p.Arg1205His	***A****/G*	***A****/G*	***A****/G*
rs1879244	***T/T***	***T****/C*	***T***
rs2230571	***C***	***C***	***C****/T*
D3S3578	*240/240*	***230****/240*	***230****/240*
D3S3583	262/**272**	268/270	268/270

Since phase is unknown for all three individuals, where necessary, both alleles are given (with the one pertaining to the described haplotype in bold). Variants comprising the reported minimal haplotype [1] are in italics

Accession numbers

NCBI accessions NM_198241.2 and NP_937884.1 were used to number all variants within the *EIF4G1* gene and eIF4G1 protein. Functional domains were assessed using UniProtKB/Swiss-Prot Q04637 (accessed January 24, 2012).

Acknowledgments We are very grateful to Jelena Golic, Susanne Lindhof, Katja Junghans, Regina Feldmann, and Sybille Frischholz for expert technical assistance.

Competing interests The authors declare that they have no conflict of interest with regard to the above study. Full financial disclosures are listed below.

Dr. Schulte received a postdoctoral fellowship from Technische Universität München, Munich, Germany. Dr. Mollenhauer has received speaker honoraria from Orion Corporation and GlaxoSmithKline; serves as an Associate Editor for the *Journal of Alzheimer Disease*; holds or has pending patents re: Method of differentially diagnosing dementias; Novel ELISA-based quantification of alpha-synuclein proteins in cerebrospinal fluid and peripheral blood products using 384-well plates; and MicroRNA expression profiling of cerebrospinal fluid; serves as a consultant for Bayer Schering Pharma AG; and receives research support from Teva Pharmaceutical Industries Ltd., Desitin Pharmaceuticals, GmbH, Boehringer Ingelheim, GE Healthcare, the Michael J. Fox Foundation for Parkinson's Research, the American Parkinson's Disease Association, and the Stifterverband für die Deutsche Wissenschaft (Dr. Werner Jackstädt-Stipend). Dr. Zimprich reports no disclosures. Dr. Bereznai receives research support from the Hungarian National Innovation Office (TÁMOP-4-2-1/B-03/1/KMR-2010-001). Dr. Lichtner reports no disclosures. Dr. Haubenberger received an NINDS Intramural Competitive Fellowship and research report from the Austrian Science Fund (Erwin Schroedinger Fellowship, project# J2783-B09) and the NINDS Intramural Research Program. Dr. Pirker has received speaker honoraria and travel compensation from Boehringer Ingelheim, Novartis, Abbott Pharmaceuticals, Medtronic, and UCB. Dr.

Brücke has received honoraria for lecturing and travel compensation from CSC, USB, Boehringer Ingelheim, Novartis, Aventis, GE Healthcare, Lundbeck, Merz, GlaxoSmithKline, and Pfizer. Dr. Molnar serves/has served on scientific advisory boards for Genzyme Europe B.V., received speaker honoria from Roche, serves as the Editor-in-Chief of the Hungarian edition of *Neurology*, and receives research support from the Hungarian National Innovation Office (TÁMOP-4-2-1/B-03/1/KMR-2010-001). Dr. Peters and Dr. Gieger report no disclosures. Dr. Trenkwalder serves on scientific advisory boards for Boehringer Ingelheim, Cephalon, Inc., UCB, Novartis, Mundipharma International Limited, and Solvay Pharmaceuticals, Inc.; has received speaker honoraria from Boehringer Ingelheim, Cephalon, Inc., UCB, Novartis, Pfizer Inc, and GlaxoSmithKline; serves on the editorial boards of *Sleep Medicine* and *Movement Disorders* and as an Associate Editor for *Focus on Parkinson Disease*. Dr. Winkelmann serves on a scientific advisory board for UCB; has received speaker honoraria from UCB and Boehringer Ingelheim; has filed a patent re: Winkelmann et al. Nat Genet 2007; and receives research support from the German RLS foundation, the Deutsche Forschungsgemeinschaft (DFG) and the Fritz Thyssen Foundation.

Funding The study depicted herein was funded by institutional funding from Helmholtz Zentrum München, Munich, Germany. Recruitment of case and control cohorts was supported by institutional (Helmholtz Zentrum München, Munich, Germany) and government funding from the German Bundesministerium für Bildung und Forschung (03.2007-02.2011 FKZ 01ET0713), and the Hungarian National Innovation Office (TAMOP-4-2-1/B-03/1/KMR-2010-001).

References

1. Chartier-Harlin MC, Dachsel JC, Vilariño-Güell C et al (2011) Translation initiator EIF4G1 in familial Parkinson disease. Am J Hum Genet 89:398–406
2. Wichmann HE, Gieger C, Illig T (2005) KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 67(Suppl 1):26–30
3. Tucci A, Charlesworth G, Sheerin UM et al (2012) Study of the genetic variability in a Parkinson's disease gene: EIF4G1. Neurosci Lett (in press)
4. Exome Variant Server, NHLBI Exome Sequencing Project (ESP), Seattle, WA. http://evs.gs.washington.edu/EVS/. Accessed 14 May 2012.
5. Schwarz JM, Rödelsperger C, Schuelke M et al (2010) MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 8:575–576
6. Adzhubei IA, Schmidt S, Peshkin L et al (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248–249
7. Satake W, Nakabayashi Y, Mizuta I et al (2009) Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson`s disease. Nat Genet 41:1303–1307
8. Zimprich A, Biskup S, Leitner P et al (2004) Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 44:601–607
9. Gehrke S, Imai Y, Sokol N, Lu B (2010) Pathogenic LRRK2 negatively regulates microRNA-mediated translational repression. Nature 466:637–641

3.1.4 Rare variants crossing diagnostic boundaries

Schulte EC, Fukumori A, Mollenhauer B, Hor H, Arzberger T, Perneczky R, Kurz A, Diehl-Schmid J, Hüll M, Lichtner P, Eckstein G, Zimprich A, Haubenberger D, Pirker W, Brücke T, Bereznai B, Molnar MJ, Lorenzo-Betancor O, Pastor P, Peters A, Gieger C, Estivill X, Meitinger T, Kretzschmar HA, Trenkwalder C, Haass C, Winkelmann J. Rare variants in β-Amyloid precursor protein (APP) and Parkinson's disease. Eur J Hum Genet 2015, 23: 1328-1333. PMID: 25604855. **(IF 2015: 4.6)**

Zech M, Nübling G, Castrop F, Jochim A, **Schulte EC**, Mollenhauer B, Lichtner P, Peters A, Gieger C, Marquardt T, Vanier MT, Latour P, Klünemann HH, Trenkwalder C, Diehl-Schmid J, Perneczky R, Meitinger T, Oexle K, Haslinger B, Lorenzl S, Winkelmann J. Niemann-Pick C disease gene mutations and age-related neurodegenerative disorders. PLoS One 2013,8:e82879. PMID: 24386122 **(IF 2013: 3.5)**

While genetic variants of different frequencies in the same locus can contribute to a given phenotype as depicted above, genetic variants in the same locus may also contribute to (seemingly) distinct phenotypes in what is called pleiotropy. The projects subsumed under this section address the role of rare variants in genes of a known role in one disorder in a different disorder. A significant overlap is known to exist in the clinical presentation of many neuropsychiatric disorders. For example, rare lysosomal storage disorders like Niemann-Pick C disease (NPC) can present with Parkinson syndromes among other neuropsychiatric features. Also, many but not all individuals with PD develop cognitive dysfunction during the course of disease. These observations prompted evaluations of the role of rare genetic variants in the *NPC* genes in multiple other neurodegenerative disorders(Zech et al., 2013) as well as the role of rare genetic variants in genes with an established role in dementias like Alzheimer's disease (AD) or FTLD in PD(Schulte et al., 2015a).

Both of these projects used targeted resequencing based on Idaho® LightScanner high-resolution melting curve analyses and Sanger sequencing to assess the role of novel and known rare coding variants in genes of known relevance to a different neuropsychiatric condition. The analyzed genes had mostly been linked to specific disorders by classical family studies and included genes of a known role in NPC (*NPC1* and *NPC2*)(Zech et al., 2013) as well as in several different kinds of dementias (*APP*, *PSEN1* & *PSEN2* known to be linked to AD; *TDP-43*, *FUS*, *GRN* & *MAPT* known to be linked to FTLD)(Schulte et al., 2015a).

In 563 individuals with PD, 133 individuals with FTLD, and 94 individuals with progressive supranuclear palsy as well as 846 general population controls, the frequencies of rare variants in *NPC1* and *NPC2* were not distributed differently. Six PD individuals (1.1%) and seven control subjects (0.8%) carried known disease-associated *NPC1* or *NPC2* mutations while none such variants were found in either individuals with FTLD or with PSP. None of the subjects with rare, disease-associated *NPC1* or *NPC2* variants were homozygous or compound heterozygous for these variants. Since NPC is an autosomal-recessive disorder, this also means that in none of the individuals with PD, FTLD, or PSP, an NPC diagnosis had been missed(Zech et al., 2013).

In the second study, "dementia genes", *APP*, *PSEN1*, *PSEN2*, *TDP-43*, *FUS*, *GRN*, and *MAPT*, were screened for rare variants in a discovery sample of 188 individuals with PD and dementia and 188 individuals with PD without dementia as well as 376 population controls belonging to the KORA cohort. Using Sequenom® MALDI-TOF, 25 out of 27 identified variants with MAF < 5% were subsequently genotyped in 975 individuals with PD (without information on additional dementia symptomes), 93 individuals with Lewy body disease, 613 individuals with AD, 182 individuals with FTLD, and 1014 general population controls. PD individuals with dementia had significantly more rare variants across all seven "dementia genes" than PD individuals without dementia (10.11% of individuals vs. 4.26%; p=0.0027; $\chi 2$ test). Further, rare variants in *APP* were also more common across all PD individuals when compared to either individuals with AD or controls. When additional controls from public databases were added, one rare variant in *APP* (c.1795G>A (p.Glu599Lys)) was significantly associated with the PD phenotype (14 out of 1068 cases versus 12 out of 5310 controls; $p_{corrected}=9.5 \times 10^{-6}$, $\chi 2$ test) but was not found in either the PD cases or controls of an independent Spanish replication sample of 715 PD cases and 948 healthy controls. Rare variants in *APP* were followed-up functionally by Aβ mass spectrometry in transiently transfected HEK293 cells. One variant (c.2125G>A (p.Gly709Ser)) shifted the Aβ spectrum from Aβ40 to Aβ39 and Aβ37 which is interesting in the context of the overall mechanism of *APP* variants in neurodegeneration. However, more in-depth analyses are needed to understand a potential causal connection to (the dementia phenotype in) PD.

Taken together, the results of the two studies depicted above suggest that, in a minority of instances, rare genetic variants in genes of relevance to one neuropsychiatric disorder could—in a cross-disorder fashion—also contribute to the

overall genetic load in other neuropsychiatric conditions. However, the scale of the conducted studies is vastly too small to truly understand the extent of this contribution. Yet, the emergence of very large-scale WES studies in the field of neuropsychiatric disorders in recent years (e.g.(Fu, 2021b; Palmer, 2020; Satterstrom et al., 2019; Singh, 2020)) begins to provide the consortial groundwork that is indispensable to answer the questions above at a more appropriate scale. The coming decade will, hopefully, provide a clearer picture.

European Journal of Human Genetics (2015) 23, 1328–1333

www.nature.com/ejhg

ARTICLE

Rare variants in β-Amyloid precursor protein (APP) and Parkinson's disease

Eva C Schulte[1,2], Akio Fukumori[3,4], Brit Mollenhauer[5,6], Hyun Hor[7], Thomas Arzberger[8], Robert Perneczky[9,10], Alexander Kurz[9], Janine Diehl-Schmid[9], Michael Hüll[11], Peter Lichtner[2,12], Gertrud Eckstein[2], Alexander Zimprich[13], Dietrich Haubenberger[13], Walter Pirker[13], Thomas Brücke[14], Benjamin Bereznai[15], Maria J Molnar[15], Oswaldo Lorenzo-Betancor[16,17,18], Pau Pastor[16,17,18], Annette Peters[19], Christian Gieger[20], Xavier Estivill[7], Thomas Meitinger[2,12,21], Hans A Kretzschmar[8], Claudia Trenkwalder[5,6], Christian Haass[3,4,21] and Juliane Winkelmann*[,1,2,12,21]

Many individuals with Parkinson's disease (PD) develop cognitive deficits, and a phenotypic and molecular overlap between neurodegenerative diseases exists. We investigated the contribution of rare variants in seven genes of known relevance to dementias (β-amyloid precursor protein (*APP*), *PSEN1/2*, *MAPT* (microtubule-associated protein tau), fused in sarcoma (*FUS*), granulin (*GRN*) and TAR DNA-binding protein 43 (*TDP-43*)) to PD and PD plus dementia (PD+D) in a discovery sample of 376 individuals with PD and followed by the genotyping of 25 out of the 27 identified variants with a minor allele frequency <5% in 975 individuals with PD, 93 cases with Lewy body disease on neuropathological examination, 613 individuals with Alzheimer's disease (AD), 182 cases with frontotemporal dementia and 1014 general population controls. Variants identified in APP were functionally followed up by Aβ mass spectrometry in transiently transfected HEK293 cells. PD+D cases harbored more rare variants across all the seven genes than PD individuals without dementia, and rare variants in *APP* were more common in PD cases overall than in either the AD cases or controls. When additional controls from publically available databases were added, one rare variant in *APP* (c.1795G>A(p.(E599K))) was significantly associated with the PD phenotype but was not found in either the PD cases or controls of an independent replication sample. One of the identified rare variants (c.2125G>A (p.(G709S))) shifted the Aβ spectrum from Aβ40 to Aβ39 and Aβ37. Although the precise mechanism remains to be elucidated, our data suggest a possible role for *APP* in modifying the PD phenotype as well as a general contribution of genetic factors to the development of dementia in individuals with PD.
European Journal of Human Genetics (2015) **23,** 1328–1333; doi:10.1038/ejhg.2014.300; published online 21 January 2015

INTRODUCTION

Linkage analyses as well as genome-wide association and exome sequencing studies have uncovered at least 20 genes associated with idiopathic Parkinson's disease (PD). Still, to date, the identified genes only explain a small portion of the genetic burden in PD. It is likely that genetic factors involved in bringing about a PD phenotype comprise both genetic variants of strong effect, which alone are causative, as well as variants of weaker effect, which contribute to disease risk or phenotypic modification.

A significant overlap between different neurodegenerative diseases has been described on the neuropathologic, the genetic and the phenotypic level.[1–4] Neuropathologically, the overlap is exemplified by the coexistence of hallmark features of both Alzheimer's disease (AD) and PD in individuals with Lewy body disease.[1] On the genetic level, common genetic variants in microtubule-associated protein tau (*MAPT*) represent risk factors for PD[3,4] whereas, at the same time, rare variants of strong effect in *MAPT* have long been recognized as a cause of frontotemporal dementia (FTD).[2] Phenotypically, it is known that at least 30% of individuals with PD develop dementia[5,6] and that age has been described as a major predisposing factor for the development of cognitive impairment.[7] Accordingly, we sought to assess the contribution of genetic factors known to be involved in dementias such as AD[8–11] or FTD[2,12–14] to the PD phenotype.

[1]Klinik und Poliklinik für Neurologie, Klinikum rechts der Isar, Technische Universität München, Munich, Germany; [2]Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany; [3]Department of Biochemistry, Adolf-Butenandt-Institut, Ludwig-Maximilians Universität München, Munich, Germany; [4]German Center for Neurodegenerative Diseases (DZNE), Munich, Germany; [5]Paracelsus Elena Klinik, Kassel, Germany; [6]Neurochirurgische Klinik, Georg August Universität Göttingen, Göttingen, Germany; [7]Genomics and Disease Group, Centre for Genomic Regulation (CRG), Pompeu Fabra University (UPF) and Centro de Investigación Biomédica en Red en Epidemiología y Salud Pública, Barcelona, Spain; [8]Institut für Neuropathologie, Ludwig-Maximillians Universität München, Munich, Germany; [9]Psychiatrische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany; [10]Neuroepidemiology and Ageing Research Unit, School of Public Health, Faculty of Medicine, The Imperial College of Science, Technology and Medicine, London, UK; [11]Psychiatrische Universitätsklinik, Albert Ludwigs Universität, Freiburg, Germany; [12]Institut für Humangenetik, Technische Universität München, Munich, Germany; [13]Department of Neurology, Medical University of Vienna, Vienna, Austria; [14]Neurologische Klinik, Wilhelminenspital, Vienna, Austria; [15]Institute of Genomic Medicine and Rare Disorders, Semmelweis University, Budapest, Hungary; [16]Neurogenetics Laboratory, Division of Neurosciences, Center for Applied Medical Research, University of Navarra, Pamplona, Spain; [17]Department of Neurology, Clinica Universidad de Navarra, University of Navarra School of Medicine, Pamplona, Spain; [18]CIBERNED, Centro de Investigacion Biomedica en Red en Enfermedades Neurodegenerativas, Instituto de Salud Carlos III, Spain; [19]Institut für Epidemiologie II, Helmholtz Zentrum München, Munich, Germany; [20]Institut für Genetische Epidemiologie, Helmholtz Zentrum München, Munich, Germany; [21]Munich Cluster for Systems Neurology, SyNergy, Munich, Germany

*Correspondence: Professor Dr J Winkelmann, Klinik und Poliklinik für Neurologie, Klinikum rechts der Isar, Technische Universität München, Ismaningerstr. 22, 81675 München, Germany. Tel: +49 89 4140 4688; Fax: +49 89 4140 7681; E-mail: winkelmann@lrz.tum.de

Received 8 August 2014; revised 18 December 2014; accepted 19 December 2014; published online 21 January 2015

METHODS

Standard protocol approvals, registrations and patient consents

Ethics review board approval was obtained at all participating institutions, with the primary review board located at the Technische Universität München, Munich, Germany. All the participants provided written informed consent for participation in the study.

Participants, variant screening and genotyping

We used Idaho LightScanner (BioFire Defense, Salt Lake City, UT, USA) melting curve analysis to screen the coding regions and exon–intron boundaries of β-amyloid precursor protein (*APP*), presenilin 1 and 2 (*PSEN1* and *PSEN2*), tau (*MAPT*), TAR DNA-binding protein 43 (*TDP-43*), granulin (*GRN*) and fused in sarcoma (*FUS*) in 376 individuals with PD (188 with PD without dementia, 188 with PD plus dementia as diagnosed according to the guidelines set forth by the task force of the Movement Disorder Society[15]) and 376 KORA-AGE controls (*APP* and *MAPT* only; Supplementary Figure 1). In the case of altered melting patterns suggestive of variants, Sanger sequencing ensued.

Variants identified during the screening phase were genotyped in 975 PD cases, 93 independent neuropathologically confirmed cases of Lewy body disease, 613 AD, 182 FTD cases and 1014 controls using Sequenom MALDI-TOF mass spectrometry. For technical reasons, *MAPT* c.1637G>A (p.(R546H)) and *PSEN2* c.211C>T (p.(R71W)) were not included. Two 3 base pair (bp) deletions in *APP* were assessed by fragment analysis as described previously.[16] One variant (*APP* c.1795G>A (p.(E599K))) that showed significant association in the first sample was also assessed in a second independent sample of 715 PD cases and 948 healthy controls from Spain. Significance was judged using the χ^2-test. For the genotyping experiments, *P*-values were corrected using the Bonferroni method. *P*-values given for burden tests represent nominal *P*-values. For a detailed description, see Supplementary Figure 1.

The following transcripts and genomic sequences were used in primer design and variant annotation: *APP*—NM_000484.3, NG_007376.1; *PSEN1*—NM_000021.3, NG_007386.2; *PSEN2*—NM_000447.2, NG_007381.1; *FUS*—NM_004960.3, NG_012889.2; *GRN*—NM_002087.2, NG_007886.1; *MAPT*—NM_001123066.3, NG_007398.1; *TDP-43*—NM_007375.3, NG_008734.1. Primer sequences are available upon request.

Immunohistochemistry

Cortical and midbrain sections of the individual harboring the *APP* c.1795G>A (p.(E599K)) variant were stained for Aβ and alpha-synuclein. Staining procedure and antibodies can be found in the supplement.

Cloning, transfections and analysis of Aß-spectrum

cDNA of the pCDNA3.1+APP695sw vector containing all identified *APP* variants were transiently transfected into HEK293 cells and Aβ was analyzed by mass spectrometry as depicted in the supplement in the culture medium.

RESULTS

Variant screening of 'dementia genes' in individuals with PD

Within the coding regions and exon–intron boundaries (±10 bp) of *APP*, *PSEN1*, *PSEN2*, *MAPT*, *FUS*, *TDP-43* and *GRN*, we identified a total of 27 rare variants with minor allele frequency (MAF) <5% in 376 individuals with PD ($n=188$; 70.4±11.73 years, 28.4% female) or PD+D ($n=188$; 72.0±6.1 years, 33.0% female). Interestingly, more individuals with PD+D (10.11%) than solely PD (4.26%) harbored a rare variant with MAF <5% in any of the seven 'dementia genes' (19 PD+D individuals with a variant *vs* 8 PD individuals with a variant; $P=0.0027$, χ^2-test). Four individuals harbored the *GRN* c.1297C>T (p.(R433W)) (rs63750412) variant and one *GRN* c.103G>A (p.(G35R)). One novel variant in *PSEN1* (c.442A>G (p.(I148W))) within two amino acids of variants known to affect the function as well as three previously reported variants in *PSEN2* (c.185G>A (p.(R62H)) (rs58973334), c.211C>T (p.(R71W)) (rs140501902), c.389C>T (p.(S130L)) (rs63750197)) were found. No variants were identified in either *TDP-43* or *FUS*. Nine were also found by the NHLBI-GO exome sequencing project.[17] (Table 1) For a detailed discussion of the phenotype of variant carriers, please refer to the supplement. (Supplementary Table 1)

For *APP* and *MAPT*, the screening was performed in the above 376 PD cases and 376 KORA-AGE controls. In *APP*, 11 rare variants with MAF <5% (seven missense, two 3-bp deletions, two nearsplice variants) were seen. In total, 10 cases but only 4 general population controls carried a rare *APP* variant. None of these variants have previously been reported in individuals with a neurodegenerative condition. In *MAPT*, we identified a total of 10 rare variants (9 missense, 1 stop). Overall, seven cases and five controls harbored a rare *MAPT* variant. (Table 1,Supplementary Table 1). Analysis by common prediction algorithms yielded contradicting results for most variants (Table 1), thus warranting additional frequency assessment and functional study.

Frequency assessment in individuals with PD, AD and FTD

Frequency assessment for 25 of the 27 variants identified in the screening phase was carried out in a sample consisting of 975 PD patients (including the 376 used above), 613 AD patients, 182 FTD patients, 93 neuropathologically confirmed cases of Lewy body disease and 1014 controls (also including the 376 used above). 68.0% of the variants were very rare with MAF <0.1% in the control sample. When compared with controls, the *APP* c.1795G>A (p.(E599K)) variant was significantly more frequent in the PD phenotype than in controls ($P=0.009$, χ^2-test; Supplementary Table 2) prior to correction for multiple testing. When publically available data from the NHLBI-ESP exomes[17] (*APP* c.1795G>A (p.(E599K)) MAF=0.15% in KORA and 0.11% in NHLBI-ESP exomes *vs* 0.66% in PD cases) were added to the controls, the finding remained significant even after Bonferroni correction for multiple testing (14 out of 1068 cases *vs* 12 out of 5310 controls; $P_{\text{nominal}}=3.8\times10^{-7}$, $P_{\text{corrected}}=9.5\times10^{-6}$, χ^2-test). Exclusion of the 376 PD cases and 376 controls used in the discovery phase of the study did not alter this finding (11 out of 692 cases *vs* 11 out of 4934 controls; $P_{\text{nominal}}=4.0\times10^{-7}$, $P_{\text{corrected}}=1.0\times10^{-5}$, χ^2-test). However, when trying to replicate this finding in a Spanish PD case/control sample, we did not find any *APP* c.1795G>A (p.(E599K)) carriers in either cases or controls, possibly suggesting a population-specific effect of *APP* c.1795G>A (p.(E599K)) in Central Europeans. *APP* c.1795G>A (p.(E599K)) was the only variant identified in the 93 Lewy body disease cases. Neuropathologically, this case was indistinguishable from other LBD cases and showed no obvious special pathology. Clinically, this individual had suffered from classical, levodopa-responsive PD with an age of onset at 59 years. Her mother had also had PD. Histology revealed both Lewy bodies in the substantia nigra (SN) and some amyloid plaques in the frontal and parietal cortex and the hippocampus, in line with a diagnosis of idiopathic PD (Figure 1).

Burden tests analyzing the load of rare variants were performed for both *APP* and *MAPT*. This revealed an excess of rare variants with MAF <5% in *APP* in PD (27 individuals with a variant out of 975) when compared with either controls alone (13 out of 1014, $P=0.018$, χ^2-test), AD cases (4 out of 613, $P=0.002$, χ^2-test) or the combined sample of controls, AD and FTD cases ($P=2.22\times10^{-4}$, χ^2-test). This excess of variants with MAF <5% in *APP* in PD (17 individuals with a variant out of 599) was also seen after exclusion of the 'discovery' samples when compared with the joined sample of controls, AD and FTD cases (14 out of 1433, $P_{\text{corrected}}=0.014$, χ^2-test) and to AD cases alone (4 out of 613, $P_{\text{corrected}}=0.014$, χ^2-test). When compared with the controls only (9 out of 638), variants were nearly twice as frequent

Table 1 Rare variants in 'dementia genes' identified in variant screening

				Cases					
				PD	*PD+D*	*Controls*	*Frequency assessment*	*NHLBI-ESP*[a]	*CADD prediction*[b]
Gene	*Genomic position (hg19)*	*dbSNP138*	*Variant*	*(n = 188)*	*(n = 188)*	*(n = 376)*	*(975 PD:93LBD:613 AD:182 FTD:1014KORA)*	*(EA only)*	*C-score*
APP	chr21:27,423,376, C>T	rs149995579	p.(A201V)		1		2:0:0:0:1	A = 3/G = 8597	1.04
APP	chr21:27,394,297_27,394,299delCTT	N/A	c.722_724delAAG		1		4:0:1:0:4	not found	13.04
APP	chr21:27,354,793, A>G	N/A	c.1091-3T>C	1			1:0:0:0:0	not found	7.06
APP	chr21:27,347,438, C>T	N/A	p.(R468H)			1	0:0:0:0:1	not found	31.00
APP	chr21:27,328,030, C>T	rs201547994	p.(A500T)			1	0:0:0:0:2	not found	25.70
APP	chr21:27,284,167, C>T	rs140304729	p.(E599K)	1	2	1	13:1:3:0:3	T = 9/C = 8591	25.40
APP	chr21:27,284,163, G>A	rs200088099	p.(T600M)			1	0:0:0:0:1	not found	22.10
APP	chr21:27,269,955_27,269,957delTCC	N/A	c.1992_1994delGGA		1		4:0:0:1:0	not found	16.18
APP	chr21:27,269,961, G>A	rs200260102	p.(T663M)		1		1:0:0:0:0	not found	13.01
APP	chr21:27,264,120, C>T	rs201269325	p.(G709S)		1		1:0:0:0:0	not found	35.00
APP	chr21:27,254,092_27,254,093 delAA	rs112965435	c.2212-10_2212-11delTT		1		1:0:0:0:1	not found	1.40
GRN	chr17:42,426,635, G>A	N/A	p.(G35R)		1		1:0:0:0:0	not found	21.70
GRN	chr17:42,429,500, C>T	rs63750412	p.(R433W)	1	3		2:0:0:0:0 homo &.6:0:5:2:6 hetero	T = 24/C = 8576	16.16
MAPT	chr17:44,039,716, C>T	N/A	p.(R5C)		1		1:0:0:0:0	not found	21.00
MAPT	chr17:44,039,717, G>A	rs63750959	p.(R5H)			1	0:0:0:0:1	not found	17.89
MAPT	chr17:44,039,824, G>A	rs115239819	p.(A41T)	1			1:0:0:0:0	not found	2.41
MAPT	chr17:44,067,341, C>T	rs143956882	p.(S427F)		1	1	2:0:4:0:3	T = 18/C = 8582	20.20
MAPT	chr17:44,067,403, C>T	rs200099007	p.(R448*)	1			2:0:4:0:3	T = 2/C = 8598	41.00
MAPT	chr17:44,068,850, G>A	rs143624519	p.(A469T)	1	1	2	3:0:4:0:5	A = 23/G = 8577	15.33
MAPT	chr17:44,071,314, C>G	N/A	p.(P511R)			1	1:0:0:0:2	not found	11.87
MAPT	chr17:44,073,840, G>A	N/A	p.(R546H)			1	n/a	not found	21.80
MAPT	chr17:44,096,064, A>G	N/A	p.(I695V)	1			1:0:0:0:0	not found	9.17
MAPT	chr17:44,101,491, C>T	rs63750991	p.(T762M)	1			1:0:0:0:0	not found	24.10
PSEN1	chr14:73,640,377, A>G	N/A	p.(I148W)		1		1:0:0:0:0	not found	9.29
PSEN2	chr1:227,071,449, G>A	rs58973334	p.(R62H)		1		4:0:3:0:5	A = 22/G = 8578	15.31
PSEN2	chr1:227,071,475, C>T	rs140501902	p.(R71W)		1		n/a	T = 32/C = 8568	18.06
PSEN2	chr1:227,073,271, C>T	rs63750197	p.(S130L)		1		2:0:2:0:4	T = 9/C = 8591	28.10

Abbreviations: AD, Alzheimer's disease; APP, β-amyloid precursor protein; chr, chromosome; D, dementia; FTD, frontotemporal dementia; GRN, granulin; MAPT, microtubule-associated protein tau; N/A = not annotated, n/a = not available because this class of variants cannot be tested using the given prediction algorithm; N/S = not scored; PD, Parkinson's disease; SNP, single-nucleotide polymorphism.
Gray boxes in the controls column symbolize that the given gene was not screened in controls. The following transcripts and genomic sequences were used in variant annotation: *APP*—NM_000484.3, NG_007376.1; *PSEN1*—NM_000021.3, NG_007386.2; *PSEN2*—NM_000447.2, NG_007381.1; *FUS*—NM_004960.3, NG_012889.2; *GRN*—NM_002087.2, NG_007886.1; *MAPT*—NM_001123066.3, NG_007398.1; *TDP-43*—NM_007375.3, NG_008734.1.
[a]Exome Variant Server, NHLBI-GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/; accessed 1 December, 2012).
[b]Combined Annotation Dependent Depletion (CADD), Seattle, WA (URL: http://cadd.gs.washington.edu; accessed 10 November, 2014).

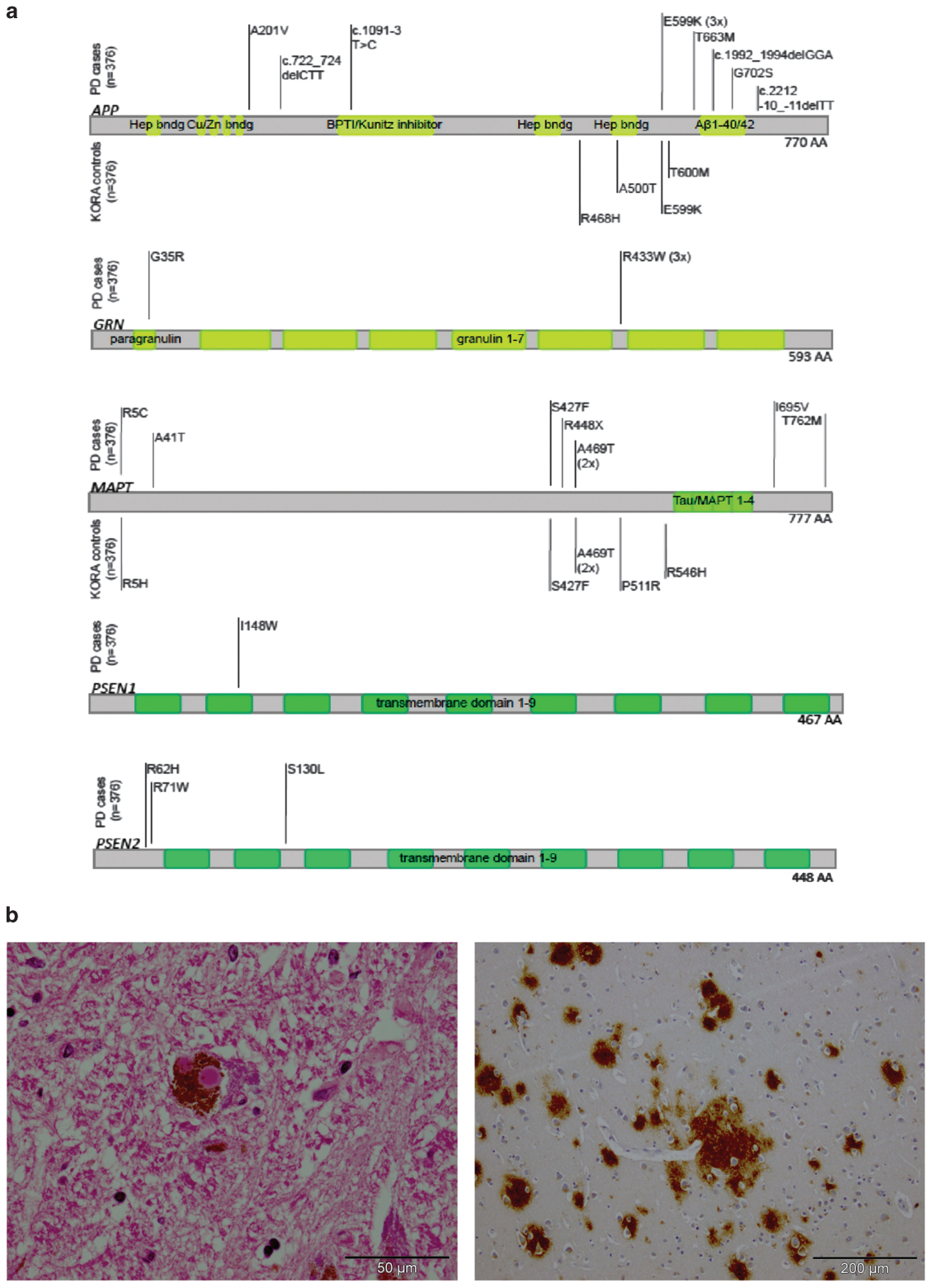


Figure 1 Location of rare variants in *APP*, *GRN*, *MAPT*, *PSEN1*, *PSEN2* and histological features of an individual harboring the c.1795G>A (p.(E599K)) variant of *APP*. (**a**) Variants with MAF <5% found in PD cases are depicted above the schematic illustration of each gene, those found in controls – if the gene was analyzed in controls – below the gene. If variants were present more than once in the discovery sample, the number of occurrences is given in parentheses. Domain annotations were taken from Uniprot (accessed 12 December, 2012). He*P*, heparin; AA, amino acids. (**b**) Depiction of a classical nigral Lewy body (left, antibody: anti-alpha-synuclein KM51, 1:1000, Novocastra/NCL-ASYN, counter stain: hematoxylin–eosine) and cortical Aβ plaques (right, antibody: 4G8, 1:2000, Signet) found in an individual with classical idiopathic PD and the *APP* c.1795G>A (p.(E599K)) variant. The neuropathology was in line with a cases of Lewy body disease (Braak stage 6) with additional Alzheimer-associated alterations (Braak and Braak Stage II), cerebral amyloid angiopathy (Thal stage 1) and beginning argyrophillic grain disease.

($MAF_{PD} = 1.41\%$ *vs* $MAF_{KORA} = 0.71\%$) but this result fell short of statistical significance ($P_{corrected} = 0.24$, χ^2-test). The frequency of rare variants in *MAPT* was similar in all the groups and remained unchanged after the omission of the initial 376 PD cases and 376 controls.

Impact of rare variants in *APP* on Aβ processing

Aβ spectral analysis was performed to further evaluate a potential functional effect of the identified coding variants in *APP*. In all but one, the Aβ spectrum reflected the wild-type situation. However, *APP* c.2125G>A (p.(G709S)), located within the Aβ domain, shifted the spectrum from Aβ40 as the main species to Aβ39 and – to a lesser extent – Aβ37 (Figure 2, Supplementary Figure 2).

DISCUSSION

Screening of seven genes known to be strong genetic factors in AD or FTD in a sample comprising both individuals with PD and PD+D revealed a number of rare variants not previously described. Interestingly, identified variants in *APP* were more common in PD with and without dementia than in either controls or AD. Next, to a mere chance occurrence, there are several possible explanations for this finding. For one, rare variants in known dementia genes could represent phenotype modifiers in PD. This is supported by the fact that in the screening sample, rare variants were more frequent in the PD+D group than in the PD group when all seven genes were analyzed together. Also, the 'dementia gene' variants could contribute to the overall 'neurodegenerative burden', which reflects an increased susceptibility for neurodegenerative conditions in general. In this scenario, an excess of genetic alterations in a specific pathway plus additional non-genetic factors could then tip the balance toward one neurodegenerative phenotype or the other or create phenotypes in which features of multiple neurodegenerative diseases and symptoms coexist. Alternatively, this could also mean that the phenotypic spectrum of AD or FTD is broader than previously recognized and could include PD-like aspects.

The boldest proposal would be that rare variants of strong effect in *APP* or *MAPT* alone could cause PD. *Mapt*$^{-/-}$ mice have recently been shown to develop not only memory deficits but also PD-specific features such as a loss of neurons in the SN and reduced locomotion.[18] Common variants in *MAPT* are an established risk factor for PD[3,4] and the relevance of allelic series – that is, both common variants of weak effect and rare variants of strong effect in one gene – to PD has already be shown.[4,19] Yet, in our sample, rare variants in *APP*, not *MAPT*, were enriched in PD. However, since a physical interaction between MAPT and APP and a role of MAPT in trafficking APP to the cell membrane has been reported,[18,20] rare variants in APP could have a similar effect with regard to PD as MAPT variants.

One of the identified APP variants (c.2125G>A (p.(G709S))) shifts the Aβ proteome spectrum from Aβ40 to Aβ39 and Aβ37 indicating that it likely interferes with γ-secretase cleavage. This could possibly be due to an alteration in the site at which APP interacts with γ-secretases, a mechanism recently postulated for increased Aβ37 production in response to an artificial APP variant (c.2095A>G (p.(K699E))) 10 amino acids N-terminal of our variant.[21] None of the other *APP* variants showed an altered Aβ spectrum. However, further studies are necessary to exclude that these variants could affect the structure and, accordingly, the aggregation potential of generated Aβ as has been demonstrated for some AD-linked variants (reviewed in Haass *et al*[22]).

Yet, from our data we cannot conclude that an Aβ-related function is truly relevant to a potential (modifying) role in PD. Next to the well-recognized role in amyloid production, recently several other functions have been identified.[23,24] APP has been described to serve as a neuronal ferroxidase, which oxidizes Fe^{2+} and loads Fe^{3+} on to the iron transport protein transferrin.[24] Moreover, iron accumulates in mice lacking *App*.[24] As iron accumulation in the SN is a known feature of PD,[25] it would be imaginable that APP dysfunction could also predispose to increased iron accumulation in the SN. *App*$^{-/-}$ mice also show increased cerebral levels of dopamine and catecholamines owing to a lack of amine catabolism via the amine oxidase function of App.[23] Increased APP expression due to *APP* variants potentially related to PD could lead to cerebral dopamine deficits and a PD phenotype. Accordingly, APP's ferroxidase[24] and amine oxidase[23] activities could even more plausibly fit a potential role in PD pathogenesis or phenotype modification and should be explored further.

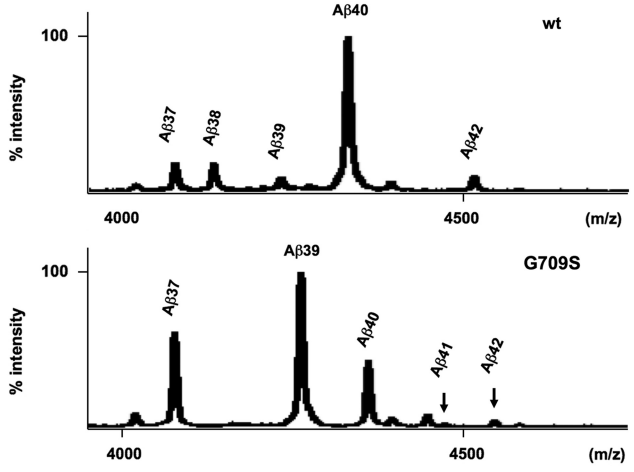


Figure 2 Functional effect of rare *APP* variants. When overexpressed in HEK293 cells, the *APP* c.2125 (p.(G709S)) variant shifts the Aβ spectrum from Aβ40 to Aβ39 and Aβ37, whereas none of the other rare APP variants examined showed alterations in the Aβ spectrum.

CONFLICT OF INTEREST

This study was funded exclusively by in-house institutional funding from the Technische Universität München and the Helmholtz Zentrum München, Munich, Germany. There was no industry sponsorship. The KORA-Age project was funded by the German Ministry for Education and Research (BMBF) FKZ 01ET0713 and 01ET1003. BM has received grants from TEVA-Pharma, Desitin, Boehringer-Ingelheim and GE Healthcare and honoraria for consultancy from Bayer Schering Pharma AG and for presentations from GlaxoSmithKline and Orion Pharma as well as travel and meeting expenses from Boehringer-Ingelheim and Novartis. DH is currently employed by the Medical University of Vienna, Austria, and received research support through the National Institute of Neurological Disorders and Stroke Intramural Research Program and the Austrian Science Fund (FWF). He serves as member of the Medical Advisory Board of the International Essential Tremor Foundation and received honoraria and conference support from Ipsen and UCB. Walter Pirker received speaker honoraria from AOP Orphan Pharma, Medtronic Inc., Novartis, Boehringer-Ingelheim, Abbott Pharm and UCB as well as travel compensation from Ipsen Pharma, Boehringer-Ingelheim and Medtronic, Inc. BB and MJM receive research support from the Hungarian National Innovation

Office (TÁMOP-4-2-1/B-03/1/KMR-2010-001). The functional analyses were supported by the European Research Council under the European Union's Seventh Framework Program (FP7/20072013)/ ERC Grant Agreement No. 321366-Amyloid (advanced grant to CH) and the general legacy of Mrs Ammer (to the Ludwig-Maximilians University/the chair of C.H.). PP is funded by a grant of the Spanish Ministry of Science and Innovation SAF-2010-22329-C02-01: 2011–2013. XE is funded by the Spanish Plan Nacional SAF-2008-00357 (NOVADIS); the Generalitat de Catalunya AGAUR 2009 SGR-1502; and the European Commission 7th Framework Program, Project No 261123 (GEUVADIS) and Project No 262055 (ESGI). JW serves on a scientific advisory board for UCB and has received speaker honoraria from UCB and Vifor Pharma. The remaining authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We are very grateful to Jelena Golic, Susanne Lindhof, Katja Junghans, Regina Feldmann and Sybille Frischholz at the Institute for Human Genetics at the Helmholtz Zentrum München for expert technical assistance.

All variants were submitted to dbSNP and received the NCBI numbers ss1399967350 to ss1399967376.

1 Deramecourt V, Bombois S, Maurage CA *et al*: Biochemical staging of synucleinopathy and amyloid deposition in dementia with Lewy bodies. *J Neuropathol Exp Neurol* 2006; **65**: 278–288.
2 Hutton M, Lendon CL, Rizzu P *et al*: Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* 1998; **393**: 702–705.
3 Pastor P, Ezquerra M, Munoz E *et al*: Significant association between the tau gene A0/A0 genotype and Parkinson's disease. *Ann Neurol* 2000; **47**: 242–245.
4 Simon-Sanchez J, Schulte C, Bras JM *et al*: Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet* 2009; **41**: 1308–1312.
5 Aarsland D, Andersen K, Larsen JP, Lolk A, Kragh-Sorensen P: Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study. *Arch Neurol* 2003; **60**: 387–392.
6 Riedel O, Klotsche J, Spottke A *et al*: Cognitive impairment in 873 patients with idiopathic Parkinson's disease. Results from the German Study on Epidemiology of Parkinson's Disease with Dementia (GEPAD). *J Neurol* 2008; **255**: 255–264.
7 Kempster PA, O'Sullivan SS, Holton JL, Revesz T, Lees AJ.: Relationships between age and late progression of Parkinson's disease: a clinico-pathological study. *Brain* 2010; **133**: 1755–1762.
8 Goate A, Chartier-Harlin MC, Mullan M *et al*: Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* 1991; **349**: 704–706.
9 Levy-Lahad E, Wasco W, Poorkaj P *et al*: Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* 1995; **269**: 973–977.
10 Rogaev EI, Sherrington R, Rogaeva EA *et al*: Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature* 1995; **376**: 775–778.
11 Sherrington R, Rogaev EI, Liang Y *et al*: Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* 1995; **375**: 754–760.
12 Cruts M, Gijselinck I, van der Zee J *et al*: Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. *Nature* 2006; **442**: 920–924.
13 Neumann M, Rademakers R, Roeber S, Baker M, Kretzschmar HA, Mackenzie IR.: A new subtype of frontotemporal lobar degeneration with FUS pathology. *Brain* 2009; **132**: 2922–2931.
14 Neumann M, Sampathu DM, Kwong LK *et al*: Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* 2006; **314**: 130–133.
15 Emre M, Aarsland D, Brown R *et al*: Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Mov Disord* 2007; **22**: 1689–1707, quiz 1837.
16 Winkelmann J, Czamara D, Schormair B *et al*: Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. *PLoS Genet* 2011; **7**: e1002171.
17 (ESP) NGESP. Exome Variant Server. Available at http://evs.gs.washington.edu/EVS/ Accessed on December 1, 2012.
18 Lei P, Ayton S, Finkelstein DI *et al*: Tau deficiency induces parkinsonism with dementia by impairing APP-mediated iron export. *Nat Med* 2012; **18**: 291–295.
19 Ross OA, Soto-Ortolaza AI, Heckman MG *et al*: Association of LRRK2 exonic variants with susceptibility to Parkinson's disease: a case-control study. *Lancet Neurol* 2011; **10**: 898–908.
20 Guo JP, Arai T, Miklossy J, McGeer PL: Abeta and tau form soluble complexes that may promote self aggregation of both into the insoluble forms observed in Alzheimer's disease. *Proc Natl Acad Sci USA* 2006;**103**:1953–1958.
21 Ousson S, Saric A, Baguet A *et al*: Substrate determinants in the C99 juxtamembrane domains differentially affect gamma-secretase cleavage specificity and modulator pharmacology. *J Neurochem* 2013; **125**: 610–619.
22 Haass C, Kaether C, Thinakaran G, Sisodia S: Trafficking and proteolytic processing of APP. *Cold Spring Harb Perspect Med* 2012; **2**: a006270.
23 Duce JA, Ayton S, Miller AA *et al*: Amine oxidase activity of beta-amyloid precursor protein modulates systemic and local catecholamine levels. *Mol Psychiatry* 2013; **18**: 245–254.
24 Duce JA, Tsatsanis A, Cater MA *et al*: Iron-export ferroxidase activity of beta-amyloid precursor protein is inhibited by zinc in Alzheimer's disease. *Cell* 2010; **142**: 857–867.
25 Dexter DT, Wells FR, Agid F *et al*: Increased nigral iron content in postmortem parkinsonian brain. *Lancet* 1987; **2**: 1219–1220.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)

OPEN ACCESS Freely available online

PLOS ONE

Niemann-Pick C Disease Gene Mutations and Age-Related Neurodegenerative Disorders

Michael Zech[1,2], Georg Nübling[6], Florian Castrop[1], Angela Jochim[1], Eva C. Schulte[1,2], Brit Mollenhauer[7,12], Peter Lichtner[2,5], Annette Peters[3], Christian Gieger[4], Thorsten Marquardt[13], Marie T. Vanier[14], Philippe Latour[14,15], Hans Klünemann[11], Claudia Trenkwalder[7,12], Janine Diehl-Schmid[8], Robert Perneczky[8,9,10], Thomas Meitinger[2,5,16], Konrad Oexle[2,5], Bernhard Haslinger[1], Stefan Lorenzl[6], Juliane Winkelmann[1,2,5,16,17]*

1 Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany, **2** Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany, **3** Institute of Epidemiology II, Helmholtz Zentrum München, Munich, Germany, **4** Institute of Genetic Epidemiology, Helmholtz Zentrum München, Munich, Germany, **5** Institut für Humangenetik, Technische Universität München, Munich, Germany, **6** Neurologische Klinik und Poliklinik, Klinikum der Universität München, Ludwig-Maximilians-Universität München, Munich, Germany, **7** Paracelsus-Elena-Klinik, Kassel, Germany, **8** Department of Psychiatry and Psychotherapy, Technische Universität München, Munich, Germany, **9** Neuroepidemiology and Ageing Research Unit, School of Public Health, Faculty of Medicine, The Imperial College of Science, Technology and Medicine, London, United Kingdom, **10** West London Cognitive Disorders Treatment and Research Unit, West London Mental Health Trust, London, United Kingdom, **11** Klinik für Psychiatrie und Psychotherapie, Universität Regensburg, Regensburg, Germany, **12** Neurochirurgische Klinik, Georg-August-Universität, Göttingen, Germany, **13** Klinik für Kinder- und Jugendmedizin, Universitätsklinikum Münster, Münster, Germany, **14** Laboratoire Gillet-Mérieux, CBPE, Hospices Civils de Lyon, Lyon, France, **15** UF de Neurogénétique Moléculaire, CBPE, Hospices Civils de Lyon, Lyon, France, **16** Munich Cluster for Systems Neurology, SyNergy, Munich, Germany, **17** Department of Neurology and Neurological Sciences and Center for Sleep Sciences and Medicine, Stanford University School of Medicine, Palo Alto, California, United States of America

Abstract

Niemann-Pick type C (NPC) disease is a rare autosomal-recessively inherited lysosomal storage disorder caused by mutations in *NPC1* (95%) or *NPC2*. Given the highly variable phenotype, diagnosis is challenging and particularly late-onset forms with predominantly neuropsychiatric presentations are likely underdiagnosed. Pathophysiologically, genetic alterations compromising the endosomal/lysosomal system are linked with age-related neurodegenerative disorders. We sought to examine a possible association of rare sequence variants in *NPC1* and *NPC2* with Parkinson's disease (PD), frontotemporal lobar degeneration (FTLD) and progressive supranuclear palsy (PSP), and to genetically determine the proportion of potentially misdiagnosed NPC patients in these neurodegenerative conditions. By means of high-resolution melting, we screened the coding regions of *NPC1* and *NPC2* for rare genetic variation in a homogenous German sample of patients clinically diagnosed with PD (n = 563), FTLD (n = 133) and PSP (n = 94), and 846 population-based controls. The frequencies of rare sequence variants in *NPC1/2* did not differ significantly between patients and controls. Disease-associated *NPC1/2* mutations were found in six PD patients (1.1%) and seven control subjects (0.8%), but not in FTLD or PSP. All rare variation was detected in the heterozygous state and no compound heterozygotes were observed. Our data do not support the hypothesis that rare *NPC1/2* variants confer susceptibility for PD, FTLD, or PSP in the German population. Misdiagnosed NPC patients were not present in our samples. However, further assessment of NPC disease genes in age-related neurodegeneration is warranted.

Citation: Zech M, Nübling G, Castrop F, Jochim A, Schulte EC, et al. (2013) Niemann-Pick C Disease Gene Mutations and Age-Related Neurodegenerative Disorders. PLoS ONE 8(12): e82879. doi:10.1371/journal.pone.0082879

Editor: Christian Wider, Centre Hospitalier Universitaire Vaudois (CHUV), Switzerland

Received October 21, 2013; **Accepted** October 28, 2013; **Published** December 30, 2013

Funding: Recruitment and analysis of 94 patients with progressive supranuclear palsy was funded by Actelion Pharmaceuticals Ltd (www.actelion.com). The remaining study as well as the recruitment of patients with Parkinson's disease and frontotemporal lobar degeneration was funded by in-house institutional funding from Technische Universität München and Helmholtz Zentrum München, Munich, Germany. Controls belong to the KORA research platform (KORA, Cooperative Research in the Region of Augsburg), which was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: MZ has received travel expenses form Actelion Pharmaceuticals Ltd. FC has received funding for travel from Medtronic. AJ has received funding for travel from Allergan and Ipsen Pharmaceuticals. BM received travel compensation from Novartis and Boehringer-Ingelheim, lecturing fees from Orion and Glaxo-Smith-Kline, grant support from GE Healthcare, Boehringer-Ingelheim, Desitin, TEVA-Pharma, and serves as a consultant to Bayer-Schering Pharma and the Michael J. Fox Foundation for Parkinson's Disease Research. MV has received travel expenses, carried out paid and unpaid consultancy work, and presentation honoraria from Actelion Pharmaceuticals Ltd., and has received travel expenses, presentation honoraria, and been invited to meetings funded and organized by Genzyme Corporation and Shire HGT. PLa has received presentation honoraria from Actelion Pharmaceuticals France. HK has received travel expenses, carried out paid consultancy work, and received presentation honoraria from Actelion Pharmaceuticals Ltd. CT serves on scientific advisory boards for Boehringer Ingelheim and UCB, has received speaker honoraria from Boehringer Ingelheim, UCB, and Mundipharma as well as travel compensation from UCB, Boehringer-Ingelheim, and Mundipharma. TMei was supported by the German Network for Mitochondrial Disorders (mitoNET 01GM0867), the European Commission Seventh Framework Program (N. 261123), Genetic European Variation in Disease Consortium, and German Ministry for Education and Research (01GR0804-4). BH serves on a scientific advisory board for Merz Pharmaceuticals; has received funding for travel from Biogen Idec, Ipsen, and Merz Pharmaceuticals; has received speaker honoraria from Allergan and Ipsen, and receives research support from Ipsen and the DFG. SL received grants from UCB Pharma, TEVA, Boehringer, Grünenthal, Orion. JW serves on a scientific advisory board for UCB; has received speaker honoraria from UCB and Boehringer Ingelheim; has filed a patent re: Winkelmann et al. Nat Genet 2007; and receives research support from the German RLS foundation, the Deutsche Forschungsgemeinschaft (DFG) and the Fritz

Thyssen Foundation. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: winkelmann@lrz.tum.de

Introduction

Niemann-Pick type C (NPC) disease (OMIM*257220 and OMIM*607625) is a neurovisceral lysosomal storage disorder, characterized biochemically by a lipid trafficking defect resulting in intracellular accumulation of unesterified cholesterol and other compounds. With incidence estimates of 1:120,000, it is a rare condition exhibiting an autosomal-recessive mode of inheritance. NPC is caused by homozygous or compound heterozygous mutations of either *NPC1* (95% of cases) or *NPC2* [1,2]. The diagnosis of NPC is established by a combination of genetic and biochemical testing, which involves *NPC1/2* gene sequencing and the demonstration of impaired intracellular cholesterol transport by filipin staining, respectively [3]. The disorder presents with an extensive phenotypic variability, ranging from fatal neonatal disease to chronic neurological deterioration in late adulthood. Besides the key clinical feature vertical supranuclear gaze palsy (VSGP), neurological symptoms encompass ataxia, early-onset cognitive decline, psychiatric disturbances, and movement disorders [1,2]. The majority of late-onset forms are diagnosed within the second or third decade, yet there are an increasing number of reported cases manifesting as late as 50 years or older, often mimicking common neurologic or psychiatric illnesses such as parkinsonian disorders or dementias [1,2,4–9]. As a result of its broad phenotypic spectrum, NPC is thought to be significantly under-diagnosed, which is momentous given that an orally applied enzyme inhibitor has proven to be an effective treatment option for slowing neurologic disease progression [1,2,10]. Recently, a remarkable proportion of NPC cases were found in adult patients with the concurrence of degenerative ataxia and presenile dementia (17% of 24 patients) [11]. Furthermore, corroborating the existence of an unrecognized pool of NPC, a multicentre study identified three NPC patients in 250 adult individuals (1.2%) suffering from psychosis and/or early-onset cognitive decline combined with neurological symptoms suggestive of NPC by *NPC1/2* gene sequencing [12]. To date, the prevalence of misdiagnosed NPC in populations with more common age-related neurodegenerative diseases is unknown.

On the molecular side, accruing evidence suggests that the group of lysosomal storage disorders or lysosomal dysfunction in general is linked with age-related neurodegenerative diseases such as Parkinson's disease (PD), frontotemporal lobar degeneration (FTLD), and progressive supranuclear palsy (PSP) [13–17]. Rare mutations in the lysosomal disorder genes *GBA* (Gaucher disease) and *SMPD1* (Niemann-Pick types A and B disease) were shown to represent susceptibility factors for PD [18,19], and the fundamental involvement of the lysosome in PD pathogenesis is supported by the observation that known PD genes such as *SNCA*, *LRRK2*, *parkin*, *PINK1*, and *ATP13A2* regulate lysosome-dependent pathways or lysosomal activity [20]. In FTLD, a critical role of impaired lysosomal function was pinpointed recently as *TMEM106B*, a gene discovered as a FTLD risk factor in genome-wide association studies, was found to influence lysosomal function and morphology [21]. Moreover, major genetic forms of FTLD such as *PRGN* and *CHMP2B* encode proteins affecting the integrity of lysosome-dependent cellular processes [22,23]. Finally, for the atypical parkinsonian disorder PSP, a recent genome-wide association study highlighted susceptibility at *STX6*, a gene implicated in the endosomal-lysosomal trafficking system, thus linking the disease to the lysosome as well [24].

Herein, in view of the clinical overlap and with regard to lysosomal dysfunction as a shared pathomechanistic feature, we screened for rare *NPC1* and *NPC2* sequence variants in patients clinically diagnosed with PD, FTLD and PSP, and a cohort of population-based controls. We first aimed to assess whether carriers of rare variants in *NPC1* and *NPC2* are at higher risk for developing PD, FTLD, or PSP. Second, based on genetic testing, we investigated the possibility of misdiagnosed NPC cases in the respective populations. Our analyses did not reveal any association between *NPC1/2* gene mutations and PD, FTLD, or PSP. Also, we could not identify any unrecognized NPC patients in our disease cohorts.

Materials and Methods

Standard protocol approvals, registrations, and patient consents

The study was approved by the ethics review board at the Technische Universität München, Munich, Germany, and the ethics review board of the Hessische Landesärztekammer in Frankfurt, Germany. All subjects provided a written consent form to participate in the study, which included detailed information about the genetic mutational screening and an authorization to publish the screening results. Subjects have been properly instructed and have indicated that they consent to participate by signing the appropriate informed consent paperwork. All potential participants who declined to participate or otherwise did not participate were eligible for treatment (if applicable) and were not disadvantaged in any other way by not participating in the study.

Participants

The study population was composed of 563 patients diagnosed with PD (32.9% female, 69.4±6.8 years), 133 patients with FTLD (41.4% female, 63.8±8.2 years), 94 patients with PSP (42.6% female, 69.7±6.8 years), and 846 general population controls (47.8% female, 75.9±6.6 years). All patients were enrolled in one of three German Medical Centers specializing in neurodegenerative disorders (Department of Neurology and Department of Psychiatry, Klinikum rechts der Isar, Technische Universität München, Munich, Germany; Paracelsus-Elena-Klinik, Kassel, Germany; Department of Neurology, Klinikum der Universität München, Ludwig-Maximilians-Universität München, Munich, Germany). The clinical diagnoses were established according to the consensus criteria for PD [25], FTLD [26], and PSP [27]. General population controls belong to the KORA-AGE cohort, a subset of the original KORA survey enriched for older individuals [28]. Individuals with known dopaminergic medication or signs of neurodegenerative disease were excluded from the control sample. All participants of the study were Caucasian and originated from the same geographic region.

Variant detection

Patients' blood samples were drawn and DNA was extracted from peripheral blood lymphocytes using standard protocols. PCR primers for the 25 exons and flanking intron regions of *NPC1* (RefSeq NM_000271) and the five exons and flanking intron regions of *NPC2* (RefSeq NM_006432) were designed with the ExonPrimer software (http://ihg.gsf.de). Primer sequences and PCR conditions are summarized in Tables S1 and S2. Variant screening was performed using Idaho®'s LightScanner™

high-resolution melting (HRM) curve analysis according to standard protocols (Idaho Technology Inc., Salt Lake City, UT) [29]. Samples with altered melting patterns were Sanger sequenced. In addition, Sanger sequencing of the entire *NPC1* or *NPC2* coding regions and flanking intron regions ensued when a known disease-associated mutation was identified, respectively. In our analyses, we focused on sequence variants with a minor allele frequency (MAF)<1% because NPC is known to be a rare condition caused by mutations with a very low frequency; synonymous substitutions were omitted since they are unlikely to be pathogenic.

Statistical analysis

Differences in variant frequencies between cases and controls were analyzed using Fisher's exact test and statistical significance levels were set at $p<0.05$.

In silico analysis of variants

PolyPhen2 [30], SIFT [31], and Mutation Taster [32] were used to evaluate the functional effect of single amino acid substitutions.

Biochemical and laboratory investigations

The filipin test was performed on fibroblasts cultured from patient skin biopsies as previously described [33]. The slides were examined on a Nikon Eclipse 80i epifluorescence microscope using an UV-1A filter (excitation 365/10, DM400, BA 400) with narrow pass. Tests to aid the diagnosis of NPC comprise measurement of plasma chitotriosidase activity as well as assessment of certain cholesterol oxidation products in plasma (oxysterols) [1,2]. Chitotriosidase activity was assayed using 4-methylumbelliferyl-β-D-N,N′,N″-triacetylchitotriose as a substrate [34]. Plasma levels of the oxysterol cholestane-3β,5α,6β-triol were quantified by gas chromatography-mass spectrometry as previously specified [35,36].

Results

In the present study, we identified rare sequence variants in *NPC1* and *NPC2* that had been previously found in patients with NPC and considered causative for the condition, henceforth referred to as "disease-associated", as well as rare sequence variants of unknown significance. Table 1 details all known disease-associated variants in *NPC1* and *NPC2* observed in patients with PD, FTLD, PSP, and controls. The screening revealed four different disease-associated *NPC1* missense variants (p.Asn222Ser, p.Arg518Trp, p.Ser1004Leu, p.Pro1007Ala) in five independent individuals with PD and one possibly disease-associated *NPC2* missense variant (p.Val30Met) in an additional subject with PD, all in the heterozygous status, giving an overall variant frequency of 1.1% among PD cases. In contrast, no disease-associated variants were seen in the groups of FTLD and PSP patients. In the control cohort, seven heterozygous carriers (0.8%) of six different disease-associated *NPC1* variants were detected, including one nonsense, one small insertion and four missense mutations (p.Asn222Ser, p.Arg348X, p.F779fsX9, p.Ser1004Leu, p.Asn1156Ser, p.Arg1186His). All rare *NPC1/2* sequence variants of unknown significance, as detected in addition to known disease-associated mutations, are listed in Table 2. These alterations comprised a total of 16 different missense and five different tentative splicing variants, eleven of them novel (*NPC1*: p.Tyr157Cys, p.Thr477Met, p.His497Tyr, p.Ala521Pro, p.Asp611Gly, p.Pro974Leu, p.Val1158Met, c.1655-1G>A, c.2131-1G>C, c.3042-5C>T; *NPC2*: p.Pro46His); two of the variants (p.As-

Table 1. Disease-associated *NPC1* and *NPC2* variants detected in individuals with PD, FTLD, PSP, and KORA-AGE controls.

Gene	Exon	Variation nucleotide	Variation amino acid	Mutation type	dbSNP137	Freq PD (n = 563)	Freq FTLD (n = 133)	Freq PSP (n = 94)	Freq controls (n = 846)	Freq NHLBI-ESP (EA)	NPC disease-association reported in
NPC1	6	c.665A>G	p.Asn222Ser	missense	rs55680026	1	0	0	1	C = 52/T = 8548	[51]
NPC1	8	c.1042C>T	p.Arg348X	nonsense	not found	0	0	0	1	not found	[52]
NPC1	9	c.1552C>T	p.Arg518Trp	missense	not found	1	0	0	0	A = 2/G = 8598	[53]
NPC1	15	c.2336_2337insT	p.F779fsX9	frameshift	not found	0	0	0	1	not found	[54]
NPC1	20	c.3011C>T	p.Ser1004Leu	missense	rs150334966	2	0	0	2	A = 7/G = 8593	[54]
NPC1	20	c.3019C>G	p.Pro1007Ala	missense	rs80358257	1	0	0	0	C = 2/G = 8598	[55]
NPC1	22	c.3467A>G	p.Asn1156Ser	missense	rs28942105	0	0	0	1	not found	[56]
NPC1	23	c.3557G>A	p.Arg1186His	missense	rs200444084	0	0	0	1	T = 1/C = 8599	[56]
NPC2	2	c.88G>A	p.Val30Met	missense	rs151220873	1	0	0	0	T = 25/C = 8575	[57,58]

Frequencies as found in the 4300 European American exomes of the NHLBI exome sequencing project (NHLBI-ESP, http://evs.gs.washington.edu/EVS/) are given for all identified variants. PD = Parkinson's disease; FTLD = frontotemporal lobar degeneration; PSP = progressive supranuclear palsy; Freq = frequency; EA = European American.

doi:10.1371/journal.pone.0082879.t001

Table 2. Rare *NPC1/2* sequence variants of unknown significance detected in individuals with PD, FTLD, PSP, and KORA-AGE controls.

Gene	Exon/ Intron	Genomic position (hg19)	Variation nucleotide	Variation amino acid	Mutation type	dbSNP137	Freq PD (n = 563)	Freq FTLD (n = 133)	Freq PSP (n = 94)	Freq controls (n = 846)	Freq NHLBI-ESP (EA)	PolyPhen2	SIFT	Mutation Taster
NPC1	2	chr18:21153416	c.180G>T	p.Gln60His	missense	rs145666943	0	0	1	2	A = 4/C = 8596	possibly damaging	N/A	disease causing
NPC1	5	chr18:21141485	c.470A>G	p.Tyr157Cys	missense		0	0	0	1	not found	probably damaging	damaging	disease causing
NPC1	5	chr18:21141414	c.541G>A	p.Ala181Thr	missense	rs199963560	0	0	0	1	not found	possibly damaging	damaging	disease causing
NPC1	9	chr18:21134845	c.1430C>T	p.Thr477Met	missense		0	0	0	1	not found	benign	tolerated	polymorphism
NPC1	9	chr18:21134795	c.1480G>A	p.Val494Met	missense	rs199812609	1	0	0	0	T = 1/C = 8599	benign	tolerated	disease causing
NPC1	9	chr18:21134786	c.1489C>T	p.His497Tyr	missense		0	0	0	1	not found	benign	tolerated	disease causing
NPC1	10	chr18:21131684	c.1561G>C	p.Ala521Pro	missense		0	0	0	1	not found	benign	tolerated	disease causing
NPC1	IVS11	chr18:21128073	c.1655-1G>A		(near-)splice		0	0	0	1	not found	N/A	N/A	disease causing
NPC1	11	chr18:21128055	c.1672G>T	p.Ala558Ser	missense	rs201156397	0	0	1	0	not found	probably damaging	damaging	disease causing
NPC1	12	chr18:21125039	c.1832A>G*	p.Asp611Gly	missense		1	0	0	0	not found	probably damaging	damaging	disease causing
NPC1	IVS14	chr18:21123534	c.2131-1G>C		(near-)splice		0	0	0	1	not found	N/A	N/A	disease causing
NPC1	16	chr18:21121118	c.2428G>T	p.Val810Phe	missense	rs145362908	1	1	0	1	A = 5/C = 8595	benign	tolerated	disease causing
NPC1	IVS20	chr18:21118640	c.2912-5G>A		(near-)splice		1	0	0	0	T = 1/C = 8599	N/A	N/A	polymorphism
NPC1	20	chr18:21118626	c.2921C>T	p.Pro974Leu	missense		0	0	0	1	not found	benign	tolerated	disease causing
NPC1	IVS21	chr18:21116845	c.3042-5C>T		(near-)splice		1	0	0	0	not found	N/A	N/A	polymorphism
NPC1	21	chr18:21116665	c.3217G>A	p.Gly1073Ser	missense	rs141440861	4	0	1	4	T = 16/C = 8584	benign	tolerated	disease causing
NPC1	22	chr18:21115438	c.3472G>A*	p.Val1158Met	missense		1	0	0	0	not found	probably damaging	damaging	disease causing
NPC2	2	chr14:74953085	c.137C>A	p.Pro46His	missense		0	1	0	0	not found	probably damaging	tolerated	disease causing
NPC2	3	chr14:74951269	c.212A>G	p.Lys71Arg	missense	rs142075589	0	1	0	1	C = 2/T = 8598	possibly damaging	tolerated	disease causing
NPC2	3	chr14:74951189	c.292A>C	p.Asn98His	missense	rs142858704	1	0	0	2	G = 10/T = 8590	benign	tolerated	polymorphism
NPC2	IVS4	chr14:74947404	c.441+1G>A		(near-)splice	rs140130028	2	1	0	1	T = 76/C = 8524	N/A	N/A	disease causing

Frequencies as found in the 4300 European American exomes of the NHLBI exome sequencing project (NHLBI-ESP, http://evs.gs.washington.edu/EVS/) are given for all identified variants. Additionally, in silico predictions of the damaging potential of all variants assessed by PolyPhen2, SIFT, and Mutation Taster are noted.

*identified in the same individual.

PD = Parkinson's disease; FTLD = frontotemporal lobar degeneration; PSP = progressive supranuclear palsy; Freq = frequency; EA = European American; N/A = not applicable.

doi:10.1371/journal.pone.0082879.t002

Table 3. *NPC1/2* variant frequencies by group[a].

	PD	FTLD	PSP	Controls
	No.(%)	No.(%)	No.(%)	No.(%)
	(n = 563)	(n = 133)	(n = 94)	(n = 846)
Disease-associated variants[b]	6 (1.1%)	0	0	7 (0.8%)
Fisher's exact test	$p = 0.78$	$p = 0.6$	$p = 1.0$	reference
all rare variants[c]	18 (3.2%)	4 (3.0%)	3 (3.2%)	26 (3.1%)
Fisher's exact test	$p = 0.88$	$p = 1.0$	$p = 1.0$	reference

PD = Parkinson's disease; FTLD = frontotemporal lobar degeneration; PSP = progressive supranuclear palsy.
[a]Absolute number of variant carriers, percentage of carriers within the group, *p* values.
[b]Variants previously described as disease-causing in a NPC patient.
[c]All rare (MAF<1%) variants detected in *NPC1* and *NPC2* (synonymous changes omitted).
doi:10.1371/journal.pone.0082879.t003

p611Gly, p.Val1158Met) were found in a single individual diagnosed with PD as described below. There were no significant differences in variant frequencies between patients with PD, FTLD, PSP and controls, neither for disease-associated *NPC1/2* variants alone nor for all rare variation found in the *NPC1* and *NPC2* genes (all $p>0.05$, Table 3).

The clinical characteristics of the six PD patients that were heterozygous for disease-associated *NPC1/2* variants are shown in Table 4. The age of disease onset ranged from 55 to 76 years, with an average onset at 65.8 years. All patients presented typical parkinsonian features responsive to dopaminergic agents, four patients exhibited a marked cognitive decline with disease progression, and two patients had a positive family history for PD. Over the course of disease, two patients were noted to have impaired eye movements, patient number 2 both vertical and horizontal and patient number 4 vertical. Patient number 4 further developed psychiatric symptoms at an early disease stage (Table 4). Sanger sequencing of the entire *NPC1* or *NPC2* coding regions and flanking intron regions detected no additional rare variants in these six individuals, respectively.

Overall, the screening disclosed no rare *NPC1/2* variants either in homozygosity or in compound heterozygosity, hence no NPC cases were recognized by our genetic analyses. One patient diagnosed with PD was found to carry two novel heterozygous missense variants in *NPC1* (p.Asp611Gly, p.Val1158Met), both with consistent pathogenic prediction by three prediction programs used (SIFT, PolyPhen2, Mutation Taster; Table 2). Segregation analysis demonstrated that the variants were not inherited independently but resided on the same chromosome. Thus, the patient was not compound heterozygous for the variants and did not meet the NPC diagnostic criteria [3]. Filipin test performed in cultured skin fibroblasts of this subject showed a pattern resembling the "variant" biochemical phenotype of NPC (Figure S1) [33]. Chitotriosidase activity and plasma oxysterol levels were in the normal range. Clinically, the 60-year-old man suffered from PD since the age of 55 years with markedly left-sided bradykinesia, rigidity, and rest tremor, and an excellent response to dopaminergic medication. There were no atypical signs for PD. Family history was positive for neurodegenerative disorders with his mother having been diagnosed with Alzheimer's disease, but negative for any movement disorders.

Table 4. Clinical characteristics of PD patients carrying disease-associated *NPC1/2* variants.

Patient No./Sex	Genotype	Age of onset, y	Age at sampling, y	IS	B	R	RT	PI	D	L-Dopa/DA	Family history	Additional information
1/M	p.Asn222Ser/wt (NPC1)	55	65	B	+	–	+	+	–	+	positive	none
2/M	p.Arg518Trp/wt (NPC1)	60	80	B	+	+	+	+	+	+	negative	vertical and horizontal gaze impaired
3/M	p.Ser1004Leu/wt (NPC1)	76	87	RT	+	+	+	–	–	+	negative	none
4/F	p.Ser1004Leu/wt (NPC1)	75	79	B	+	+	–	+	+	+	negative	vertical gaze impaired, hallucinations early in the disease course
5/M	p.Pro1007Ala/wt (NPC1)	61	72	B	+	+	+	–	+	+	positive	none
6/M	p.Val30Met/wt (NPC2)	68	71	B	+	+	–	+	+	+	negative	none

PD = Parkinson's disease; Sex: M = male, F = female; wt = wild type; IS = initial symptom; B = bradykinesia; R = rigor; RT = resting tremor; PI = postural instability; D = dementia; DA = dopamine agonist.
doi:10.1371/journal.pone.0082879.t004

Discussion

We investigated the possible role of rare sequence variants in the *NPC1* and *NPC2* genes, mutations in which are causative for the lysosomal storage disorder NPC, in three age-related neurodegenerative diseases (PD, FTLD, PSP). Dysfunction of the lysosomal degrading system has been implied in a variety of neurodegenerative processes and lysosomal storage disorders in particular have been strongly linked to parkinsonism [13–17,20]. Mutations in the *GBA* gene, which encodes the lysosomal enzyme deficient in Gaucher disease, are one of the commonest risk factors for PD, which was primarily shown in Ashkenazi Jewish individuals and subsequently in a number of other populations worldwide [37–41]. More recently, a founder mutation in *SMPD1*, the gene for Niemann-Pick types A and B disease (acid sphingomyelinase deficiencies), was recognized as a novel susceptibility factor for PD in the Ashkenazi Jewish population [19]. The same study failed to prove association between PD and founder mutations in the lysosomal enzyme genes *HEXA* (Tay-Sachs disease) and *MCOLN1* (mucolipidosis type IV) [19]. Now, our analyses generated evidence that mutations in the lysosomal storage disorder genes *NPC1* and *NPC2* are not associated with PD in a homogeneous sample of European descent. The proportion of PD patients positive for disease-associated *NPC1/2* variants (1.1%) was relatively low when compared to *GBA* mutation frequencies reported in non-Jewish PD cohorts (4–7%) [18,42]. Moreover, rare variants in *NPC1/2* appear not to be associated with FTLD and PSP in the German population. Notably, there are limitations to the present study. First, the study was powered at 80% to detect a significant association of rare *NPC1/2* variants with PD when modeling odds ratios $\geq$2.08 (significance level of 0.05, cumulative MAF of rare variants in the present study $\sim$1.6%). The sample size was not large enough to judge modest or small effects of rare *NPC1/2* variants on PD risk. Taking this further, for an association with FTLD or PSP odds ratios should have been $\geq$3.86 (power of 80%), considering that these patient cohorts were relatively small. Second, our control sample was composed of individuals from the general population without signs of neurodegenerative diseases or taking dopaminergic drugs. Nonetheless, there might be potential risks for PD, FTLD, and PSP later in life and these could have confounded our observations. The carrier frequency for disease-associated *NPC1* variants among control subjects (0.8%) is in line with the predicted frequency of 0.6% given a disease incidence of 1:120,000. Third, as we used HRM for variant detection and did not perform Sanger sequencing of the entire *NPC1/2* coding region, the frequencies of *NPC1/2* variants could be underestimated across all samples. However, this effect was likely to be small since previous investigations applying HRM yielded a diagnostic sensitivity of 100% for heterozygous variants and 93% for homozygous variants [29,43]. Ultimately, different results may be obtained by using more specific inclusion criteria for patients like an early disease onset or a positive family history, or by conducting the study in geographically and ethnically different populations.

Albeit the lack of a genetic association in this study, it cannot be fully excluded that heterozygous pathogenic variants in *NPC1* and *NPC2* represent a component of risk for age-related neurodegenerative disorders or might play a role in certain subsets of such patients. In two of six PD patients (33%) heterozygous for disease-associated *NPC1/2* variants impaired vertical gaze was found on clinical examination, an atypical sign for PD and the key feature of NPC, and one of these patients developed concomitant psychiatric symptoms early in the disease course. Findings from animal models highlight that heterozygous *NPC1* mutations affect neuronal function and neurodegenerative disease status, particularly in the context of aging [44,45]. Further, several studies suggest the possibility of symptomatic heterozygotes in human NPC: Josephs *et al.* proposed one mutant *NPC1* allele as the cause of parkinsonian tremor in a 75-year-old patient [46]. Harzer *et al.* report a *NPC1* heterozygote manifesting systemic signs of NPC during childhood [47]. And, a very recent manuscript described three independent adult relatives of NPC patients who were heterozygous *NPC1* mutation carriers and exhibited a parkinsonism syndrome [48].

NPC displays an extreme clinical heterogeneity, with a large number of possible differential diagnoses. The most common presentation in adult-onset cases is a psychiatric illness combined with cognitive decline and motor signs (parkinsonism in 10%), but mild clinical pictures with predominant motor dysfunction are also observed [5]. VSGP is a characteristic sign of NPC, but also evident in other neurological disorders. In the present study, we could not unveil any misdiagnosed NPC in 563 patients with PD, 133 patients with FTLD, and 94 patients with PSP by means of a mutational screen. This negative result notwithstanding, it seems important to note that NPC patients might be identified in adult neurologic disease cohorts, for example when testing larger numbers of patients or including individuals exhibiting more exceptional clinical presentations, as recently demonstrated [11,12]. Besides, it is possible that NPC diagnoses could have been missed because sensitivity of HRM is not 100%, and there was no exploration of large deletions or deep intronic mutations, which were shown to be rarely responsible for NPC [49,50] [Latour and Vanier, unpublished data]. Our study detected an individual with PD who carried two novel *NPC1* missense variants (p.Asp611Gly, p.Val1158Met) but was found not to be a compound heterozygote on segregation analysis. Notably, this case emphasizes the crucial need to check for independent allele segregation when establishing the diagnosis of NPC by gene sequencing. Biochemical characterization of the two novel variants by filipin staining revealed that at least one or the combination of the two variants is functionally relevant to the NPC1 protein since mild abnormalities resembling the "variant" biochemical phenotype were observed in patient skin fibroblasts. This pattern is seen in a subset of patients with NPC but is also well documented in heterozygote carriers of NPC [1,33]. An effect on plasma oxysterol levels has been described for heterozygous *NPC1* mutations [35], but was not seen in this case. Moreover, confirming the polymorphic nature of the NPC loci, our study disclosed eight additional novel variants in *NPC1* and one novel variant in *NPC2*, yet their functional significance with regard to the NPC1 and 2 proteins remains unknown.

In conclusion, our study indicates that rare variants in the *NPC1* and *NPC2* genes are not associated with PD, FTLD, and PSP in our populations and that, moreover, misdiagnosed NPC seems not to be frequent in these entities. Further *NPC* mutational screenings in larger and ethnically diverse cohorts of patients with PD and other neurodegenerative conditions should be undertaken to conclusively define the contribution of these lysosomal genes to the development of age-related neurodegeneration.

Supporting Information

Figure S1 Filipin test from a PD patient carrying in cis the *NPC1* variants p.Asp611Gly and p.Val1158Met. Fibroblasts cultured from skin biopsies of a healthy control subject (negative control, **A**), a classical NPC patient (positive control, **B**), and the PD patient (**C**), after staining of unesterified cholesterol by filipin. The

fibroblasts were maintained three days in a culture medium supplemented with 10% lipoprotein-deficient calf serum to maximize LDL-receptors expression. The cholesterol-starved fibroblasts were then challenged with human purified LDLs (50 µg/ml medium) for 24 h, and finally fixed with formalin and stained [33]. Cells were examined by epifluorescence microscopy (Nikon Eclipse 80i, UV-1A filter, ×20 Planfluor objective, DXM1200-C/NIS Elements imaging system). In C, the PD patient presents 30–50% of weakly positive cells. Original magnification ×200.
(TIF)

Table S1 Primers used for HRM and Sanger sequencing.
(DOC)

Table S2 Touchdown PCR protocol.
(DOC)

Acknowledgments

We are very grateful to Susanne Lindhof, Regina Feldmann, Jelena Golic, Katja Junghans, and Sybille Frischholz for expert technical assistance.

Author Contributions

Conceived and designed the experiments: MZ GN ECS BH SL JW. Performed the experiments: MZ GN ECS. Analyzed the data: MZ GN ECS BH SL JW. Contributed reagents/materials/analysis tools: MZ GN FC AJ ECS BM P. Lichtner AP CG T. Marquardt MV P. Latour HK CT JDS RP KO T. Meitinger BH SL JW. Wrote the paper: MZ GN JW.

References

1. Group N-CGW, Wraith JE, Baumgartner MR, Bembi B, Covanis A, et al. (2009) Recommendations on the diagnosis and management of Niemann-Pick disease type C. Mol Genet Metab 98: 152–165.
2. Vanier MT (2010) Niemann-Pick disease type C. Orphanet J Rare Dis 5: 16.
3. Patterson MC, Hendriksz CJ, Walterfang M, Sedel F, Vanier MT, et al. (2012) Recommendations for the diagnosis and management of Niemann-Pick disease type C: an update. Mol Genet Metab 106: 330–344.
4. Vanier MT, Wenger DA, Comly ME, Rousson R, Brady RO, et al. (1988) Niemann-Pick disease group C: clinical variability and diagnosis based on defective cholesterol esterification. A collaborative study on 70 patients. Clin Genet 33: 331–348.
5. Sevin M, Lesca G, Baumann N, Millat G, Lyon-Caen O, et al. (2007) The adult form of Niemann-Pick disease type C. Brain 130: 120–133.
6. Klunemann HH, Elleder M, Kaminski WE, Snow K, Peyser JM, et al. (2002) Frontal lobe atrophy due to a mutation in the cholesterol binding protein HE1/NPC2. Ann Neurol 52: 743–749.
7. Josephs KA, Van Gerpen MW, Van Gerpen JA (2003) Adult onset Niemann-Pick disease type C presenting with psychosis. J Neurol Neurosurg Psychiatry 74: 528–529.
8. Trendelenburg G, Vanier MT, Maza S, Millat G, Bohner G, et al. (2006) Niemann-Pick type C disease in a 68-year-old patient. J Neurol Neurosurg Psychiatry 77: 997–998.
9. Walterfang M, Fietz M, Fahey M, Sullivan D, Leane P, et al. (2006) The neuropsychiatry of Niemann-Pick type C disease in adulthood. J Neuropsychiatry Clin Neurosci 18: 158–170.
10. Patterson MC, Vecchio D, Prady H, Abel L, Wraith JE (2007) Miglustat for treatment of Niemann-Pick C disease: a randomised controlled study. Lancet Neurol 6: 765–772.
11. Schicks J, Muller Vom Hagen J, Bauer P, Beck-Wodl S, Biskup S, et al. (2013) Niemann-Pick type C is frequent in adult ataxia with cognitive decline and vertical gaze palsy. Neurology 80: 1169–1170.
12. Bauer P, Balding DJ, Klunemann HH, Linden DE, Ory DS, et al. (2013) Genetic screening for Niemann-Pick disease type C in adults with neurological and psychiatric symptoms: findings from the ZOOM study. Hum Mol Genet.
13. Zhang L, Sheng R, Qin Z (2009) The lysosome and neurodegenerative diseases. Acta Biochim Biophys Sin (Shanghai) 41: 437–445.
14. Settembre C, Fraldi A, Jahreiss L, Spampanato C, Venturi C, et al. (2008) A block of autophagy in lysosomal storage disorders. Hum Mol Genet 17: 119–129.
15. Shachar T, Lo Bianco C, Recchia A, Wiessner C, Raas-Rothschild A, et al. (2011) Lysosomal storage disorders and Parkinson's disease: Gaucher disease and beyond. Mov Disord 26: 1593–1604.
16. Tofaris GK (2012) Lysosome-dependent pathways as a unifying theme in Parkinson's disease. Mov Disord 27: 1364–1369.
17. Nixon RA (2013) The role of autophagy in neurodegenerative disease. Nat Med 19: 983–997.
18. Sidransky E, Nalls MA, Aasly JO, Aharon-Peretz J, Annesi G, et al. (2009) Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N Engl J Med 361: 1651–1661.
19. Gan-Or Z, Ozelius LJ, Bar-Shira A, Saunders-Pullman R, Mirelman A, et al. (2013) The p.L302P mutation in the lysosomal enzyme gene SMPD1 is a risk factor for Parkinson disease. Neurology 80: 1606–1610.
20. Dehay B, Martinez-Vicente M, Caldwell GA, Caldwell KA, Yue Z, et al. (2013) Lysosomal impairment in Parkinson's disease. Mov Disord.
21. Brady OA, Zheng Y, Murphy K, Huang M, Hu F (2013) The frontotemporal lobar degeneration risk factor, TMEM106B, regulates lysosomal morphology and function. Hum Mol Genet 22: 685–695.
22. Belcastro V, Siciliano V, Gregoretti F, Mithbaokar P, Dharmalingam G, et al. (2011) Transcriptional gene network inference from a massive dataset elucidates transcriptome organization and gene function. Nucleic Acids Res 39: 8677–8688.
23. Urwin H, Authier A, Nielsen JE, Metcalf D, Powell C, et al. (2010) Disruption of endocytic trafficking in frontotemporal dementia with CHMP2B mutations. Hum Mol Genet 19: 2228–2238.
24. Hoglinger GU, Melhem NM, Dickson DW, Sleiman PM, Wang LS, et al. (2011) Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nat Genet 43: 699–705.
25. Hughes AJ, Daniel SE, Kilford L, Lees AJ (1992) Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 55: 181–184.
26. Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, et al. (1998) Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology 51: 1546–1554.
27. Litvan I, Agid Y, Calne D, Campbell G, Dubois B, et al. (1996) Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop. Neurology 47: 1–9.
28. Wichmann HE, Gieger C, Illig T, Group MKS (2005) KORA-gen–resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 67 Suppl 1: S26–30.
29. van der Stoep N, van Paridon CD, Janssens T, Krenkova P, Stambergova A, et al. (2009) Diagnostic guidelines for high-resolution melting curve (HRM) analysis: an interlaboratory validation of BRCA1 mutation scanning using the 96-well LightScanner. Hum Mutat 30: 899–909.
30. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. (2010) A method and server for predicting damaging missense mutations. Nat Methods 7: 248–249.
31. Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. Genome Res 11: 863–874.
32. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 7: 575–576.
33. Vanier MT, Rodriguez-Lafrasse C, Rousson R, Gazzah N, Juge MC, et al. (1991) Type C Niemann-Pick disease: spectrum of phenotypic variation in disruption of intracellular LDL-derived cholesterol processing. Biochim Biophys Acta 1096: 328–337.
34. Hollak CE, van Weely S, van Oers MH, Aerts JM (1994) Marked elevation of plasma chitotriosidase activity. A novel hallmark of Gaucher disease. J Clin Invest 93: 1288–1292.
35. Porter FD, Scherrer DE, Lanier MH, Langmade SJ, Molugu V, et al. (2010) Cholesterol oxidation products are sensitive and specific blood-based biomarkers for Niemann-Pick C1 disease. Sci Transl Med 2: 56ra81.
36. Jiang X, Sidhu R, Porter FD, Yanjanin NM, Speak AO, et al. (2011) A sensitive and specific LC-MS/MS method for rapid diagnosis of Niemann-Pick C1 disease from human plasma. J Lipid Res 52: 1435–1445.
37. Aharon-Peretz J, Rosenbaum H, Gershoni-Baruch R (2004) Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. N Engl J Med 351: 1972–1977.
38. Sato C, Morgan A, Lang AE, Salehi-Rad S, Kawarai T, et al. (2005) Analysis of the glucocerebrosidase gene in Parkinson's disease. Mov Disord 20: 367–370.
39. Eblan MJ, Nguyen J, Ziegler SG, Lwin A, Hanson M, et al. (2006) Glucocerebrosidase mutations are also found in subjects with early-onset parkinsonism from Venezuela. Mov Disord 21: 282–283.
40. Tan EK, Tong J, Fook-Chong S, Yih Y, Wong MC, et al. (2007) Glucocerebrosidase mutations and risk of Parkinson disease in Chinese patients. Arch Neurol 64: 1056–1058.
41. Spitz M, Rozenberg R, Pereira Lda V, Reis Barbosa E (2008) Association between Parkinson's disease and glucocerebrosidase mutations in Brazil. Parkinsonism Relat Disord 14: 58–62.
42. Neumann J, Bras J, Deas E, O'Sullivan SS, Parkkinen L, et al. (2009) Glucocerebrosidase mutations in clinical and pathologically proven Parkinson's disease. Brain 132: 1783–1794.
43. Xiao J, Zhao Y, Bastian RW, Perlmutter JS, Racette BA, et al. (2010) Novel THAP1 sequence variants in primary dystonia. Neurology 74: 229–238.

44. Brown DE, Thrall MA, Walkley SU, Wurzelmann S, Wenger DA, et al. (1996) Metabolic abnormalities in feline Niemann-Pick type C heterozygotes. J Inherit Metab Dis 19: 319–330.
45. Yu W, Ko M, Yanagisawa K, Michikawa M (2005) Neurodegeneration in heterozygous Niemann-Pick type C1 (NPC1) mouse: implication of heterozygous NPC1 mutations being a risk for tauopathy. J Biol Chem 280: 27296–27302.
46. Josephs KA, Matsumoto JY, Lindor NM (2004) Heterozygous Niemann-Pick disease type C presenting with tremor. Neurology 63: 2189–2190.
47. Harzer K, Beck-Wodl S, Bauer P (2013) Niemann-Pick Disease Type C: New Aspects in a Long Published Family - Partial Manifestations in Heterozygotes. JIMD Rep.
48. Kluenemann HH, Nutt JG, Davis MY, Bird TD (2013) Parkinsonism syndrome in heterozygotes for Niemann-Pick C1. J Neurol Sci.
49. Rodriguez-Pascau L, Coll MJ, Vilageliu L, Grinberg D (2009) Antisense oligonucleotide treatment for a pseudoexon-generating mutation in the NPC1 gene causing Niemann-Pick type C disease. Hum Mutat 30: E993–E1001.
50. Rodriguez-Pascau L, Toma C, Macias-Vidal J, Cozar M, Cormand B, et al. (2012) Characterisation of two deletions involving NPC1 and flanking genes in Niemann-Pick type C disease patients. Mol Genet Metab 107: 716–720.
51. Tangemo C, Weber D, Theiss S, Mengel E, Runz H (2011) Niemann-Pick Type C disease: characterizing lipid levels in patients with variant lysosomal cholesterol storage. J Lipid Res 52: 813–825.
52. Sztolsztener ME, Strzelecka-Kiliszek A, Pikula S, Tylki-Szymanska A, Bandorowicz-Pikula J (2010) Cholesterol as a factor regulating intracellular localization of annexin A6 in Niemann-Pick type C human skin fibroblasts. Arch Biochem Biophys 493: 221–233.
53. Ribeiro I, Marcao A, Amaral O, Sa Miranda MC, Vanier MT, et al. (2001) Niemann-Pick type C disease: NPC1 mutations associated with severe and mild cellular cholesterol trafficking alterations. Hum Genet 109: 24–32.
54. Sun X, Marks DL, Park WD, Wheatley CL, Puri V, et al. (2001) Niemann-Pick C variant detection by altered sphingolipid trafficking and correlation with mutations within a specific domain of NPC1. Am J Hum Genet 68: 1361–1372.
55. Greer WL, Dobson MJ, Girouard GS, Byers DM, Riddell DC, et al. (1999) Mutations in NPC1 highlight a conserved NPC1-specific cysteine-rich domain. Am J Hum Genet 65: 1252–1260.
56. Vanier MT, Millat G (2003) Niemann-Pick disease type C. Clin Genet 64: 269–281.
57. Park WD, O'Brien JF, Lundquist PA, Kraft DL, Vockley CW, et al. (2003) Identification of 58 novel mutations in Niemann-Pick disease type C: correlation with biochemical phenotype and importance of PTC1-like domains in NPC1. Hum Mutat 22: 313–325.
58. Chikh K, Rodriguez C, Vey S, Vanier MT, Millat G (2005) Niemann-Pick type C disease: subcellular location and functional characterization of NPC2 proteins with naturally occurring missense mutations. Hum Mutat 26: 20–28.

3.2 Variants of Intermediate Frequency

Schulte EC*, Kousi M*, Tan P, Schormair B, Knauf F, Lichtner P, Trenkwalder C, Högl B, Frauscher B, Berger K, Fietze I, Gross N, Stiasny-Kolster K, Oertel W, Bachmann C.G, Paulus W, Zimprich A, Peters A, Gieger C, Meitinger T, Müller-Myshok B, Katsanis N, Winkelmann J. Targeted Resequencing and Systematic In Vivo Functional Testing Identifies Rare Variants in MEIS1 as Significant Contributors to Restless Legs Syndrome. Am J Hum Genet 2014, 95:85-95. PMID: 24995868 **(IF 2014: 10.9)**

During the search for rarer, non-common variants (MAF <5%; the cut-off for variants confidently detected by array-based methodologies like those used in GWAS at the time) contributing to RLS susceptibility and explaining some of the "missing heritability"(Schulte et al., 2014b), variants of intermediate frequency were also assessed if located within the exons, the bordering exon/intron boundaries or the 3' or 5' untranslated regions (UTRs) of *MEIS1*. In a sample consisting of 3,760 individuals with RLS and 3,542 general population controls, an overall excess of low-frequency and rare variants with MAF < 5% was observed (1,383 variant counts in case subjects versus 606 in control subjects; $p=1.04 \times 10^{-61}$)(Schulte et al., 2014b). This excess was primarily driven by a single low-frequency variant (rs11693221), located in the 3'UTR of the canonical isoform of *MEIS1* (MAF_{cases}=13.55%, $MAF_{controls}$=3.58%; $p=1.27x10^{-89}$; OR=4.42). This finding in conjunction with an enrichment of rare variants in the 5'UTR of *MEIS1* importantly implicates the UTRs in disease pathogenesis, possibly through a regulation of mRNA expression or stability.

The linkage disequilibrium (LD) (r^2) is low between this low-frequency variant in the 3'UTR and the most significantly associated common variant within *MEIS1*(Winkelmann et al., 2011; Winkelmann et al., 2007*)* in the 1000 Genomes data(Johnson et al., 2008), suggesting that this signal is independent of the one that initially placed *MEIS1* on the genetic map of RLS. Only one other low-frequency variant in a highly conserved non-coding region in intron 8 of *MEIS1* is in high linkage disequilibrium with r^2>0.8, had previously been discussed in the context of RLS in a Canadian case/control sample(Xiong et al., 2009), and has since also been shown to underly the strongest signal in the most recent GWAS of RLS(Schormair et al., 2017). Taken together, these data further substantiate that several independent genetic signals of different allele frequencies within a single genomic locus can contribute to shaping the genetic landscape of common, complex neuropsychiatric disorders or, more specifically, RLS.

Targeted Resequencing and Systematic In Vivo Functional Testing Identifies Rare Variants in *MEIS1* as Significant Contributors to Restless Legs Syndrome

Eva C. Schulte,[1,2,22] Maria Kousi,[3,22] Perciliz L. Tan,[3] Erik Tilch,[2,4] Franziska Knauf,[2] Peter Lichtner,[2,4] Claudia Trenkwalder,[5,6] Birgit Högl,[7] Birgit Frauscher,[7] Klaus Berger,[8] Ingo Fietze,[9] Magdolna Hornyak,[1,10,11] Wolfgang H. Oertel,[12] Cornelius G. Bachmann,[13,14] Alexander Zimprich,[15] Annette Peters,[16] Christian Gieger,[17] Thomas Meitinger,[2,4,18] Bertram Müller-Myhsok,[18,19,20] Nicholas Katsanis,[3,23] and Juliane Winkelmann[1,2,4,18,21,23,*]

Restless legs syndrome (RLS) is a common neurologic condition characterized by nocturnal dysesthesias and an urge to move, affecting the legs. RLS is a complex trait, for which genome-wide association studies (GWASs) have identified common susceptibility alleles of modest (OR 1.2–1.7) risk at six genomic loci. Among these, variants in *MEIS1* have emerged as the largest risk factors for RLS, suggesting that perturbations in this transcription factor might be causally related to RLS susceptibility. To establish this causality, direction of effect, and total genetic burden of *MEIS1*, we interrogated 188 case subjects and 182 control subjects for rare alleles not captured by previous GWASs, followed by genotyping of ~3,000 case subjects and 3,000 control subjects, and concluded with systematic functionalization of all discovered variants using a previously established in vivo model of neurogenesis. We observed a significant excess of rare *MEIS1* variants in individuals with RLS. Subsequent assessment of all nonsynonymous variants by in vivo complementation revealed an excess of loss-of-function alleles in individuals with RLS. Strikingly, these alleles compromised the function of the canonical *MEIS1* splice isoform but were irrelevant to an isoform known to utilize an alternative 3′ sequence. Our data link *MEIS1* loss of function to the etiopathology of RLS, highlight how combined sequencing and systematic functional annotation of rare variation at GWAS loci can detect risk burden, and offer a plausible explanation for the specificity of phenotypic expressivity of loss-of-function alleles at a locus broadly necessary for neurogenesis and neurodevelopment.

Introduction

Restless legs syndrome (RLS [MIM 102300]) is a common neurologic condition with an age-dependent prevalence of up to 10% in Europe and North America.[1] It is characterized by an irresistible urge to move the legs accompanied by disagreeable, often painful, sensations in the lower limbs at night. Moving the affected legs or walking leads to prompt but only temporary relief.[1] As a consequence, individuals suffer from persistent insomnia, leading to an impairment of quality of life and mental health. RLS is a highly familial trait but genetically complex, with estimates of narrow-sense heredity between 54% and 69% as derived from twin studies.[2,3]

Genome-wide association studies (GWASs) in large RLS case/control samples have identified common susceptibility alleles at six loci that together explain about 7% of the heritability.[4] Among many models that can explain some of the "missing heritability,"[5] we considered the possibility that a collection of rare variants of strong effect, which cannot be identified by means of GWASs,[6,7] might be a contributory factor. Although a single potentially causal rare variant has been described in *MEIS1* (MIM 601739),[8,9] to date no variants of strong effect have been established. Nonetheless, for some other complex genetic diseases, such as diabetes or chronic inflammatory bowel disease, rare variants have recently been identified within known GWAS loci,[10,11] supporting the concept of allelic series in complex genetic disorders. We therefore sought to assess the potential contribution of rare variants to disease burden both by using standard statistical methods and by assessing the incidence and contribution of

[1]Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, 81675 Munich, Germany; [2]Institut für Humangenetik, Helmholtz Zentrum München, 85764 Munich, Germany; [3]Center for Human Disease Modeling, Department of Cell Biology, Duke University, Durham, NC 27710, USA; [4]Institut für Humangenetik, Technische Universität München, 81675 Munich, Germany; [5]Paracelsus Elena Klinik, 34128 Kassel, Germany; [6]Klinik für Neurochirurgie, Georg August Universität, 37075 Göttingen, Germany; [7]Department of Neurology, Medizinische Universität Innsbruck, 6020 Innsbruck, Austria; [8]Institut für Epidemiologie und Sozialmedizin, Westfälische Wilhelms Universität Münster, 48149 Münster, Germany; [9]Zentrum für Schlafmedizin, Charité Universitätsmedizin, 10117 Berlin, Germany; [10]Interdisziplinäres Schmerzzentrum, Albert-Ludwigs Universität Freiburg, 79106 Freiburg, Germany; [11]Diakoniewerk München-Maxvorstadt, 80799 Munich, Germany; [12]Klinik für Neurologie, Philipps Universität Marburg, 35039 Marburg, Germany; [13]Abteilung für Neurologie, Paracelsus Klinikum Osnabrück, 49076 Osnabrück, Germany; [14]Klinische Neurophysiologie, Georg August Universität, 37075 Göttingen, Germany; [15]Department of Neurology, Medizinische Universität Wien, 1090 Vienna, Austria; [16]Institute of Epidemiology II, Helmholtz Zentrum München, 85764 Munich, Germany; [17]Institute of Genetic Epidemiology, Helmholtz Zentrum München, 85764 Munich, Germany; [18]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; [19]Max-Planck Institut für Psychiatrie München, 80804 Munich, Germany; [20]Institute of Translational Medicine, University of Liverpool, Liverpool L69 3BX, UK; [21]Department of Neurology and Neurosciences, Center for Sleep Sciences and Medicine, Stanford University, Palo Alto, CA 94304, USA
[22]These authors contributed equally to this work
[23]These authors contributed equally to this work
*Correspondence: winkelmann@stanford.edu
http://dx.doi.org/10.1016/j.ajhg.2014.06.005.

functionally annotated variants relevant to MEIS1 biology (Figure S1 available online). For this purpose, we exploited two major resources: a well-phenotyped, ethnically homogeneous RLS cohort and an experimentally tractable method to assay MEIS1 functionality grounded on previously defined in vivo observations on the roles of this protein in neurogenesis, wherein suppression of *meis1* in zebrafish embryos led to a quantitative reduction of the optic tectum, a major site of neurogenesis in the developing brain, and malformation of rhombomeres 3 and 5, which represent early hindbrain structures shown previously to be defective in the absence of *meis1*.[12,13] Our data showed a significant enrichment of rare variants across both *MEIS1* and also all seven RLS-associated genes. Based upon population statistics alone, only one single low-frequency variant in the 3′ UTR (rs11693221) showed significant association with RLS. However, the combinatorial usage of functional annotation and statistical analyses highlighted a major contribution of loss-of-function variants in *MEIS1* and suggested that rare alleles in this locus pose significant RLS risk to individuals.

Materials and Methods

Participants

Both case and control populations were entirely of German and Austrian descent. In the case subjects, diagnosis was based on the diagnostic criteria of the International RLS Study Group[1] as assessed in a personal interview conducted by an RLS expert. In keeping with the previous GWASs,[4,14] we excluded individuals with secondary RLS due to uremia, dialysis, or anemia resulting from iron deficiency. The presence of secondary RLS was determined by clinical interview, physical and neurological examination, blood chemistry, and nerve conduction studies whenever deemed clinically necessary. Participants' written informed consent was obtained prior to the initiation of the study. The institutional review boards of the contributing authors approved the study. The primary review board was located in Munich at the Bayerische Ärztekammer and Technische Universität München.

Genotyping by High-Resolution Melting Curve Analysis

In a first step, we used Idaho LightScanner high-resolution melting curve analysis (Biofire) to screen the coding regions and exon/intron boundaries of *PTPRD* (MIM 601598), *BC034767*, *TOX3* (MIM 611416), *BTBD9* (MIM 611237), *MEIS1*, and *MAP2K5* (MIM 602520) for variants. Due to the high GC content, the coding regions ±10 bp of *SKOR1* (MIM 611273) could not be subjected to LightScanner analysis and were Sanger sequenced instead. Included in the screening were 188 German RLS-affected case subjects and 182 general population control subjects belonging to the KORA cohort[13] based in Southern Germany. Where possible, the 188 case subjects used were half homozygous and half heterozygous for the published risk alleles.[4,14–16] The same set of control subjects was used for all screening experiments. *MEIS1*, *TOX3*, and *BC034767* variants identified in the 188 case subjects have already been published.[4,8] In the case of an altered melting pattern suggestive of variants, Sanger sequencing ensued to identify the underlying variant. The same method was used to screen the coding regions of *MEIS1* isoforms 1 and 2 ±10 bp as well as the 5′ and 3′ UTRs in 3,760 RLS case subjects of German and Austrian descent (62.2 ± 12.8 years; 30.8% male) and 3,542 general-population control subjects (55.1 ± 13.8 years; 40.1% male) belonging to the S4 and F4 surveys of KORA.[17] For the discovery sample, group comparisons between case and control subjects were performed in R[18] for each gene and each type of variant separately, with and without the common risk allele genotype as covariate, using logistic regression (logreg) of the phenotype on aggregate minor allele counts of variants of each type.[19] To account for a possible bias introduced by the comparison of different sets of risk-genotype-selected cases to a constant set of unselected controls, we performed a meta-analysis using rmeta to evaluate the contribution of rare coding variants across all seven genes.

Empirical p values were calculated with 1,000 permutations of the phenotype and assessing the ratio of p values equal to or smaller than the p value belonging to the original phenotype. For the large-scale screening of *MEIS1*, both logreg of the phenotype on number of variants as well as sequence kernel association tests (SKAT)[20] with and without Madsen-Browning weights[21] were performed. Empirical p values are based on 10,000 permutations of the phenotype, calculating the ratio of test statistics equal to or larger than the test statistic of the original phenotype.

Genotyping by Mass Spectrometry

Genotyping was carried out on the MassARRAY system by MALDI-TOF mass spectrometry with iPLEX Gold chemistry (Sequenom). Primers were designed with AssayDesign v.3.1.2.2 with iPLEX Gold default parameters. No assay could be designed for seven variants, largely those located in the extremely GC-rich gene *SKOR1*. Further, three assays failed two or more times and were, therefore, not pursued further. Automated genotype calling was carried out with SpectroTYPER v.3.4. Genotype clustering was visually checked by an experienced evaluator. SNPs with a call rate <90% were excluded. The genotyping sample consisted of 3,262 case subjects of German and Austrian descent (65.3 ± 11.3 years; 29.3% male) and 2,944 general population control subjects (56.1 ± 13.3 years; 48.7% male) from the KORA F4 survey.[17] For the most part, case and control subjects used in both genotyping approaches were drawn from the same samples. Both logreg of the phenotype on number of variants as well as SKATs[20] with and without Madsen-Browning weights[21] were performed. Empirical p values are based on 200 permutations of the phenotype, calculating the ratio of test statistics larger than the test statistic of the original phenotype.

In Vivo Complementation in Zebrafish Embryos, In Situ Hybridization, and Whole-Mount Immunostaining

Splice-blocking morpholinos (MOs) against *meis1* and *map2k5* were designed and obtained from Gene Tools. We injected 1 nl of diluted MO (4 ng for *meis1*_MO1, 3 ng for *meis1*_MO2, and 5 ng for *map2k5*) and/or RNA (75 pg for *meis1*, 100 pg for *map2k5*) into wild-type zebrafish embryos at the 1- to 2-cell stage. To evaluate hindbrain organization, injected embryos were raised until the 10- to 13-somite stage, corresponding to 14 to 16 hr post-fertilization (hpf); they were then dechorionated and fixed in 4% paraformaldehyde (PFA) overnight. Fixed embryos were transferred to 100% methanol at −20°C for at least 2 hr and were then processed after standard protocols[22] using a digoxygenin-labeled antisense riboprobe against *krox20*. For analysis of the

optic tectum, injected embryos were fixed overnight at 72 hpf in 4% PFA and stored in 100% methanol at −20°C. For acetylated tubulin staining, embryos were fixed in Dent's fixative (80% methanol, 20% DMSO) overnight at 4°C. The embryos were permeabilized with proteinase K followed by postfixation with 4% PFA. PFA-fixed embryos were washed first in PBS and subsequently in IF buffer (0.1% Tween-20, 1% BSA in PBS) for 10 min at room temperature. The embryos were incubated in blocking buffer (10% FBS, 1% BSA in PBS) for 1 hr at room temperature. After two washes in IF buffer for 10 min each, embryos were incubated in the primary antibody (anti-acetylated tubulin [T7451, mouse, Sigma-Aldrich], 1:1,000) in blocking solution, overnight at 4°C. After two additional washes in IF buffer for 10 min each, embryos were incubated in the secondary antibody solution (Alexa Fluor goat anti-mouse IgG [A21207, Invitrogen], 1:1,000) in blocking solution, for 1 hr at room temperature.

For RNA rescue and overexpression experiments, the human wild-type mRNAs of isoforms 1 (RefSeq accession number NM_002398.2/ENST00000272369) and 2 (no RefSeq ID/ENST00000381518) of *MEIS1* as well as the canonical isoform of *MAP2K5* (RefSeq NM_145160.2/ENST00000178640) were cloned into the pCS2+ vector and transcribed in vitro using the SP6 Message Machine kit (Ambion). All variants identified in isoform 1 as well as all additional variants only coding in isoform 2 plus two functional null variants from isoform 1 (p.Ser204Thr [c.610T>A] and p.Arg272His [c.815G>A]; both RefSeq NM_002398.2/ENST00000272369) were introduced with Phusion high-fidelity DNA polymerase (New England Biolabs) and custom-designed primers. Additionally, a non-naturally occurring homeobox domain-dead variant of *MEIS1* (p.Arg276Ala + p.Asn325Ala [c.826_827delinsGC + c.973_974delinsGC]; RefSeq NM_002398.2/ENST00000272369) was created. All the experiments were repeated in triplicate and significance of the morphant phenotype was judged with Student's t test.

Results

Variant Screening of Seven RLS GWAS Candidate Genes

To assess low-frequency (1% < minor allele frequency [MAF] < 5%) coding variation at the known RLS susceptibility loci encompassing *PTPRD*, *TOX3*, *BTBD9*, *MEIS1*, *MAP2K5*, *SKOR1*, and the noncoding RNA *BC034767*,[4,14–16] we screened the coding regions and exon-intron boundaries (±10 bp) in 188 German individuals with RLS and 182 control subjects belonging to the KORA population cohort[13] based in Southern Germany. A total of 49 variants with MAF < 5% were identified in case and control subjects together (Table S1). When collapsed across all seven genes, rare and low-frequency nonsynonymous variants showed a trend toward being more frequent in case than in control subjects (39 in case subjects versus 24 in control subjects; $p = 0.103$; logreg meta-analysis; odds ratio [OR] = 1.40). Nonsynonymous and synonymous variants combined, however, showed a stronger enrichment in individuals with RLS (77 case subjects versus 46 control subjects; $p = 0.023$, logreg meta-analysis; OR = 1.51). Addition of the common risk allele genotype[4,14–16] into the analysis as a covariate decreased the permuted p value (nonsynonymous only: $p = 0.079$; all variants: $p = 0.008$; n = 1,000 permutations), suggesting that the enrichment of rare variants across all loci is independent of the common risk allele genotype. However, the degree of interdependence between common and rare variants differed between genes (Table S2).

Within *MEIS1*, synonymous or nonsynonymous variants with a MAF below 5% were present in nine case subjects but only one control subject ($p = 0.021$) although no marked difference in the amount of variation was obvious in any of the other genes (Table S1). Here, the p value increased modestly after addition of the common risk variant (rs2300478) as covariate ($p = 0.080$; 1,000 permutations), indicating some interdependence of rare and common alleles at the *MEIS1* locus. Variants did not seem to cluster within specific regions of the examined genes (Figure S2).

Genotyping of Identified Low-Frequency and Rare Variants

To assess a possible association with the RLS phenotype, we next genotyped 39 of the 49 identified variants in 3,262 German and Austrian RLS case subjects (65.3 ± 11.3 years; 29.3% male) and 2,944 KORA control subjects (56.1 ± 13.3 years; 48.7% male). Of the 49 variants, 10 could not be included for technical reasons (see Material and Methods section above). Although variants (either altogether or only nonsynonymous ones) with MAF < 5% and MAF < 1% were not significantly enriched in RLS, we observed a distinct excess of very rare variants with a KORA-derived MAF < 0.1% (total: 57 case versus 16 control subjects, $p = 4.99 \times 10^{-4}$, OR = 2.50; nonsynonymous only: 23 case versus 5 control subjects, $p = 0.0019$; OR = 3.92; logreg; Figure 1).

We then went on to assess whether more sophisticated tests used to analyze rare variant associations of bidirectional effects (SKAT)[20] and differing allele frequency (SKAT with Madsen-Browning weights)[21] would change the association signal. However, although frequency-weighted results were still superior to unweighted results in most cases, overall SKAT analysis led to an increase in association p values both for the joined analysis across all loci as well as for *MEIS1* alone (Table S2).

We next asked whether this signal is distributed across all seven tested genes, or whether a specific subset of tested loci cause the apparent enrichment of rare nonsynonymous variation in individuals with RLS. A significant association with RLS was observed for *MEIS1*; the aggregate risk conferred by the variants showed a large increase as we transitioned from rare (MAF < 1%) to very rare (MAF < 0.1%) (Figure 1). More specifically, rare variants with MAF < 1% (total: 116 case subjects versus 67 control subjects, $p = 0.0064$, OR = 1.51; nonsynonymous only: 39 versus 14, $p = 0.0024$, OR = 2.46) and MAF < 0.1% (both total and nonsynonymous only: 9 versus 1: $p = 0.014$; OR = 8.14; Figure 1) were seen more frequently in cases than in controls. SKAT (all: $p = 0.049$;

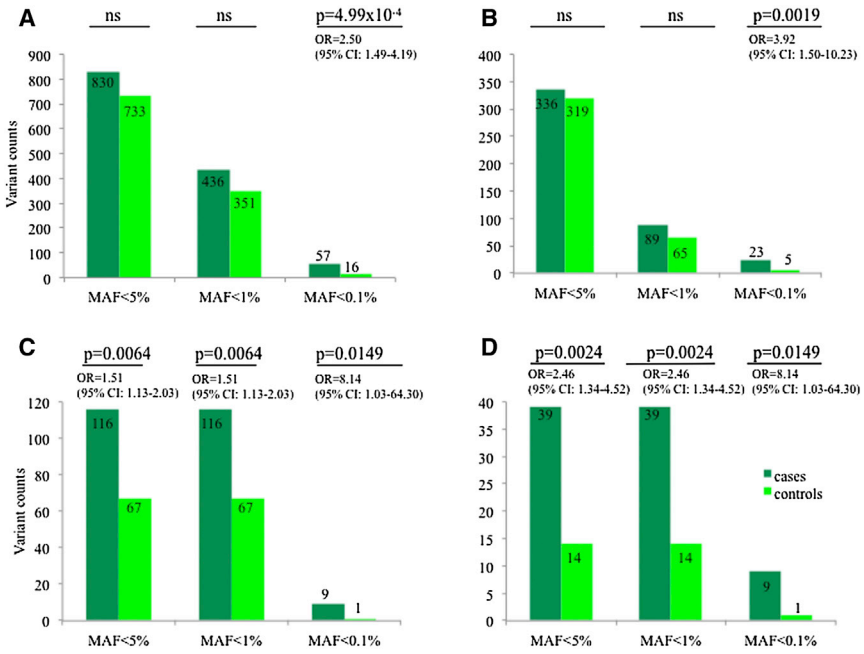


Figure 1. Excess of Rare Coding Variants at RLS-Associated GWAS Loci

Frequency assessment of 39 low-frequency and rare variants identified in the coding sequences of seven genes associated with RLS[4,14–16] in 3,262 case subject and 2,944 control subjects revealed an excess of both overall (A) and nonsynonymous (B) variants with MAF < 0.1% across all examined loci. The same held true for the overall (C) and nonsynonymous (D) variants at the *MEIS1* locus.

nonsynonymous only: p = 0.009) and Madsen-Browning-weighted SKAT (all: p = 0.019; nonsynonymous only: p = 0.029; 10,000 permutations each) substantiated this finding (Table S2). No low-frequency coding variants with MAF between 1% and 5% were found within the coding regions of *MEIS1*.

After exclusion of *MEIS1*, logreg (all variants: p = 0.009; nonsynoymous only: p = 0.024; 2,000 permutations) for variants with MAF < 0.1% and SKAT using Madsen-Browning weights (all variants: p = 0.019; nonsynonymous only: p = 0.019; 100 permutations) showed a nominally significant enrichment across all other six genes. In *PTPRD* alone, rare variants of all classes were also encountered more frequently in individuals with RLS (all variants with MAF < 0.1%: $p = 9.99 \times 10^{-4}$, logreg; all variants: p = 0.029, SKAT with MB) whereas this enrichment was not observed for nonsynonymous variants only. Of note, several individual rare variants in *MEIS1* (p.Arg272His [c.815G>A] and p.Met453Thr [c.1359T>C] [ENST00000381518]), *TOX3* (p.Ala233Ala [c.699T>C]; RefSeq NM_001080430.2/ENST00000219746), and *PTPRD* (c.551−4C>G and p.Pro278Pro [c.834T>G]; both RefSeq NM_002839.3/ENST00000381196) were associated nominally with RLS in the large case/control sample; however, associations did not withstand correction for multiple testing (Table S1).

Assessment of Rare Variation in *MEIS1*

The excess of low-frequency and rare variants at RLS-associated GWAS loci was most pronounced for *MEIS1*. Therefore, we sought to expand our analysis to a more comprehensive investigation of genetic variation with regard to frequency and location within *MEIS1* by screening the coding regions ±10 bp as well as the 5′ and 3′ UTRs for variants with MAF < 5% in 3,760 German and Austrian RLS case subjects (62.2 ± 12.8 years; 30.8% male) and 3,542 KORA control subjects (55.1 ± 13.8 years; 40.1% male). We identified a total of 75 such variants (Tables S3 and S4); 28 of these lie in either the canonical isoform 1 (ENST00000272369) or a longer isoform 2 (ENST00000398506) of *MEIS1* that encodes an alternate start site and an alternate C terminus. All synonymous and nonsynonymous variants identified in the initial screening of 188 case subjects and 182 control subjects were observed again.

Overall, we observed an excess of variants with MAF < 5% across all examined regions of *MEIS1* (1,383 variant counts in case subjects versus 606 in control subjects, $p = 1.04 \times 10^{-61}$; not permuted), which was driven primarily by a low-frequency variant, rs11693221, in the 3′ UTR ($MAF_{cases} = 13.55\%$, $MAF_{controls} = 3.58\%$; $p = 1.27 \times 10^{-89}$; OR = 4.42 [95% CI: 3.83–5.11]; not permuted; Figure 2, Tables S3 and S4). After exclusion of this variant, the remainder of individuals with low-frequency and rare variants across all regions of *MEIS1* was similar in case and control subjects (432 versus 396; p = 0.68). However, stratification of variants according to their localization showed an excess of rare variants with MAF < 5% in the 5′ UTR (16 case subjects versus 2 control subjects; logreg: p = 0.001, OR = 7.56; SKAT: p = 0.01; SKAT with MB: p = 0.006) and among nonsynonymous variants in isoform 2 (34 case subjects versus 15 control subjects; logreg: p = 0.007, OR = 2.31; SKAT: p = 0.004; SKAT with MB: p = 0.0005) (Figure 2, Table S3).

Functional Analysis of Rare Nonsynonymous Variants in *MEIS1* by In Vivo Complementation in Zebrafish Embryos

Resequencing of any locus is certain to reveal rare variants in both case and control subjects which, bereft of a means of preselecting for variants relevant to protein function, can dampen or extinguish bona fide association signals. We, therefore, considered a paradigm grounded on prior knowledge of *MEIS1* to test whether each of the rare discovered variants in our study have an effect on function

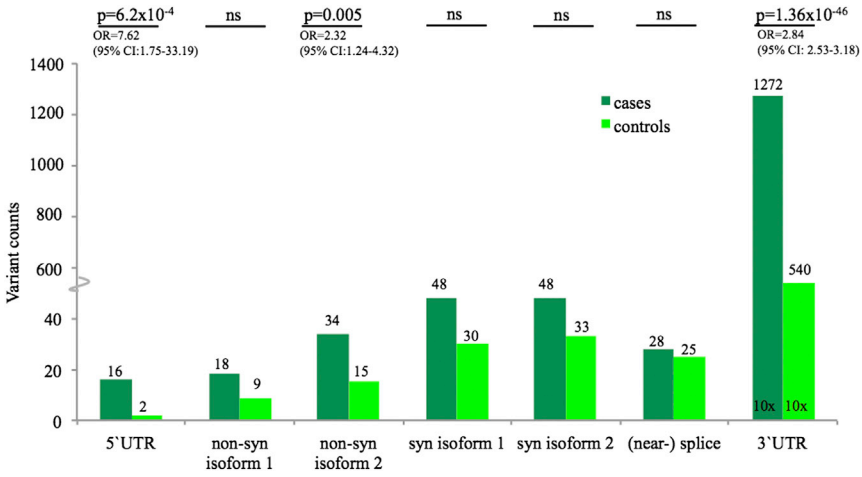


Figure 2. Variant Screening of the Coding Regions and UTRs of *MEIS1* in 3,760 Individuals with RLS and 3,542 KORA Control Subjects

Stratification according to variant localization shows an excess of rare variants in both the 5′ UTR and among nonsynonymous coding variants. Low-frequency and rare variants in the 3′ UTR were also more frequent in case subject than in control subjects. No difference was observed in the number of individuals carrying synonymous coding or (near-) splice variants.

and to then use this information to assess the burden of deleterious genetic lesions in our case/control RLS study. During zebrafish development, suppression of *meis1* has been shown to impact neurogenesis, a phenotype captured prominently by the quantitative reduction of the size of the optic tectum 72 hpf[12] as well as the disruption of hindbrain patterning at 14 hpf.[13] We first tested the ability of human *MEIS1* mRNA to rescue the optic tectum size phenotype and to thus establish a baseline assay for the evaluation of the identified nonsynonymous variants. To this end, we designed two independent MOs that target different exon-intron splice junctions of the endogenous zebrafish *meis1*. Both MOs gave rise to the same phenotype, bolstering our confidence in the specificity of the assay: we observed a reduction of the size of the optic tectum by ~30% when injecting 4 ng of *meis1*_MO1 targeting exon-intron boundary 2, and by ~20% when injecting 3 ng of a previously characterized MO (*meis1*_MO2) against the acceptor site of exon 2[23] (Figures S3 and S4). In both cases, the phenotype was rescued significantly and reproducibly ($p < 0.0001$; performed in triplicate, scored blind to injection cocktail) by coinjection with 75 pg of human capped *MEIS1* mRNA (Figures S3 and S4). The optic tectum phenotype could be rescued by either the canonical *MEIS1* isoform or by an isoform utilizing an alternative 3′ terminus. By contrast, injection of the "domain-dead" human mRNA (MEIS1_DD) bearing two variants engineered to ablate DNA binding ability (p.Arg276Ala+Asn325Ala) was indistinguishable from MO alone ($p < 0.68$). As a test of the relevance of the phenotype to RLS, MO suppression of *map2k5,* another GWAS-associated RLS gene, yielded a similar phenotype with regard to the size of the optic tectum, which was also specific as it could be rescued by coinjection with human *MAP2K5* mRNA (Figure S5).

Given these data, we proceeded to perform in vivo complementation assays on all 13 nonsynonymous coding variants identified in isoform 1, as well as in the 4 nonsynonymous variants that lie in the unique sequence of isoform 2 (Figure 3, Table S4), wherein human mRNA bearing one test variant was coinjected with MO and compared to the rescue ability of WT mRNA (n = at least 51 embryos tested per injection, Figure 3 and Tables S5 and S6). We classified variants as benign (rescue indistinguishable from WT), hypomorphs (mutant rescue significantly worse than WT but better than MO alone), or functionally null (indistinguishable from MO alone). Among the 13 variants of isoform 1, we identified three benign, four hypomorphic, and six null variants (Figures 3 and 4). Overexpression of *MEIS1* WT mRNA or mRNAs harboring each of the 13 variants had no effect on optic tectum size (Figure S6).

Layering these data over the incidence and distribution of these variants in our case/control data set, we found a significant excess of functionally null variants in individuals with RLS compared to control subjects (14 in case subjects versus 2 in control subjects; $p = 0.0012$; OR = 7.48 [95% CI: 1.68–33.40]) (Figures 3 and 5) whereas hypomorphic (2 in case subjects versus 4 in control subjects; $p > 0.05$) and benign (2 in cases subject versus 2 in control subjects; $p > 0.05$) variants showed similar distributions in case and control subjects.

To corroborate these data, we designed a second assay at an earlier developmental time point, grounded on the known requirement of *meis1* for the organization of the hindbrain; in triplicate experiments, ~30% of the embryos suppressed for *meis1* developed hindbrain defects that reproduced previously reported *meis1* phenotypes.[13] These consisted of significant widening of rhombomere 3 and/or 5 (r3 and r5); shortening of the distance between r3 and r5; thinning of either one of those structures; or absence of r3 and/or r5 altogether (Figure 4). This phenotype was rescued by coinjection with 75 pg of WT human *MEIS1* mRNA ($p = 0.045$). Assessment of six variants from our series (three variants scored as null and three variants scored as benign in the optic tectum assay) and blind triplicate scoring confirmed this result: p.Ser204Thr, p.His239Tyr (c.715C>T), and p.Arg272His were verified as null variants and p.Ala122Val (c.365C>T), p.Ser175Asn (c.524G>A), and p.Met366Leu (c.1096A>T) were validated as benign (all RefSeq NM_002398.2/ENST00000272369).

By contrast, all four nonsynonymous variants exclusive to isoform 2 scored benign in our assay. Strikingly, two

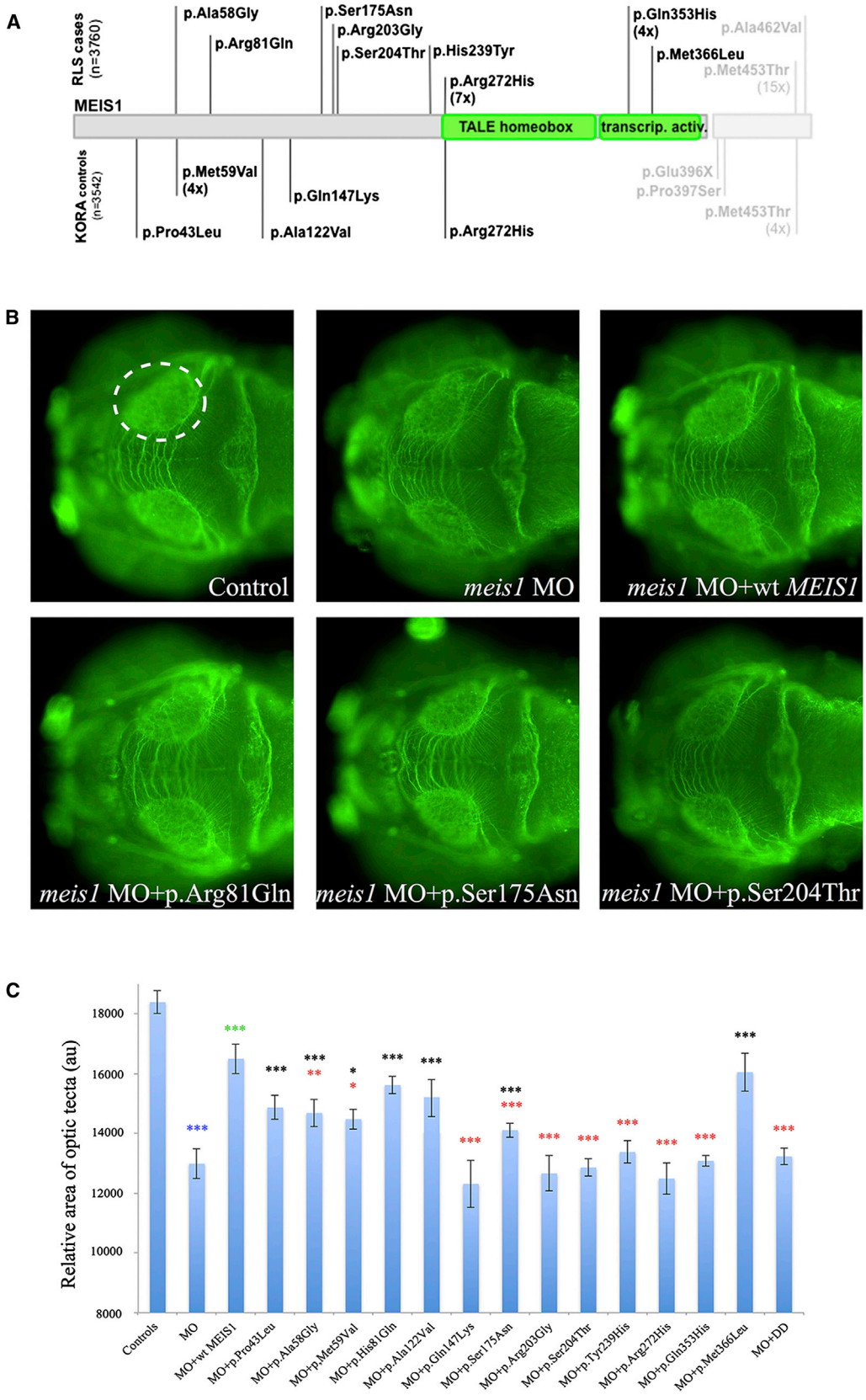


Figure 3. Functional Assessment of Rare Nonsynonymous Variants in *MEIS1* by In Vivo Complementation in Zebrafish Embryos
(A) Location and frequency of nonsynonymous *MEIS1* variants examined in zebrafish. Variants found in case subjects are given above the gene, those found in control subjects below. The short, canonical isoform 1 of *MEIS1* is given in dark gray (ENST00000272369); the additional amino acids in the longer isoform 2 in light gray (ENST00000398506).

(legend continued on next page)

variants found to be functional nulls in the isoform 1 complementation assay (p.Ser204Thr and p.Arg272His) were able to fully rescue the tectum size phenotype in isoform 2 (Figures 5 and S6), suggesting that the contribution of rare variants to RLS is mediated specifically by reduced activity of MEIS1 isoform 1-encoded protein.

Discussion

Previous GWASs have established the genomic locus encompassing *MEIS1* as the most significant susceptibility region for RLS.[4,14] The most likely candidate gene in this region is *MEIS1*, a TALE homeobox transcription factor known to be involved in specifying spinal motor neuron pool identity and connectivity[24] as well as proximo-distal limb patterning[25] and expressed in forebrain neurons and astrocytes[26] during embryonic development. A common RLS-linked intronic variant in *MEIS1* was also shown to induce differential forebrain enhancer activity during development.[27] Additional studies in the context of RLS have suggested a link between *MEIS1* and iron metabolism in the central nervous system.[28,29]

The excess of rare alleles of functional effect in RLS case subjects compared to control subjects shown here substantiates *MEIS1* as the causal genetic factor underlying the observed associations. Moreover, it implicates loss of function as the underlying mechanism, at least with regard to rare variants. We also observed a new association with a low-frequency variant (rs11693221) located in the 3′ UTR of the ENSEMBL-derived canonical isoform 1 of *MEIS1* (ENST00000272369) that represents the largest single-allele genetic risk factor for RLS identified to date. In the 1000 Genomes data,[30] the linkage disequilibrium (LD, r^2) between rs11693221 and the most significantly associated common variant within *MEIS1* (rs12469063)[4,14] is low ($r^2 = 0.080$). In the same data set,[30] rs11693221 is in high LD ($r^2 > 0.8$) solely with a single low-frequency variant (rs113851554, $r^2 = 0.83$). rs113851554 is located within a highly conserved noncoding region in intron 8 and has previously also been shown to be associated with RLS in a Canadian case/control sample.[29] Follow-up analyses are needed to fully dissect all functional effects underlying the *MEIS1* association signal.

Nonetheless, the finding of an associated low-frequency variant (rs11693221) in the 3′ UTR of isoform 1 in conjunction with the observed excess of rare variants in the 5′ UTR of *MEIS1* in individuals with RLS implicates the UTRs in disease pathogenesis, potentially through regulating expression and/or mRNA stability. The excess of rare noncoding variants in the 100 bp surrounding the exons of nine genes associated with asthma[31] and the fact that fine-mapping studies located about 22% of 36 GWAS association signals for celiac disease to either the 5′ or the 3′ noncoding regions (UTRs and several kilobases up- or downstream)[32] could indicate that these regions are indeed important in the context of complex genetic diseases and might be overlooked by the current surge of whole-exome sequencing studies.

The functional experiments conducted in zebrafish allowed us to differentiate between potentially benign and pathological sequence variation and thereby increased both effect size and significance levels observed in burden testing. Moreover, the identified rare nonsynonymous *MEIS1* variants, which showed an effect on optic tectum size in zebrafish embryos, were restricted to isoform 1 of *MEIS1* (ENST00000272369). Previous studies have demonstrated that rare variants can exhibit isoform-specific effects on a given phenotype, such as in the case of *DNAJB6* in limb-girdle muscular dystrophy.[33] Because the pathophysiology of RLS is just beginning to be elucidated, it will be of importance to see when and where this isoform of *MEIS1* is expressed as the temporal and spatial expression patterns of the different *MEIS1* transcripts are currently unknown. Given our observations, we speculate that understanding the differential biological roles of isoform 1 will help dissect the subset of MEIS1 functions relevant to the etiopathology of RLS.

Recent studies have implicated allelic series of variants of different frequency and effect sizes at loci identified in the context of GWASs in complex genetic diseases. In single cases, individual rare variants were shown to be associated with the phenotype[11,34,35] whereas in other cases it was the collective of rare variation either within a single gene[10] or across a number of GWAS-identified loci.[36,37] Our data substantiate this role of the whole of genetic variation, from common to low-frequency to rare variation, at a GWAS-identified locus in the genetic architecture of a complex genetic disease. Interestingly, the addition of synonymous variants to the burden analyses yielded a more significant enrichment of rare variants in many situations. Whether this is due to increased power, fine-scale population substructure, artificial signal amplification driven by high LD between the synonymous and the nonsynonymous or causal variants, or a true causal contribution of rare synonymous

(B) At 72 hpf, zebrafish larvae were stained as whole mounts using an antibody against acetylated tubulin and the size of the optic tecta was measured for phenotypic read out. Control, morpholino injection, and rescue by human WT mRNA are shown in the upper panels. The lower panels illustrate the effects of different alleles tested.
(C) Quantification of optic tectum area in zebrafish larvae at 72 hpf (n = at least 51 per genotype). Benign alleles show a significant difference with regard to the MO injection, hypomorphic alleles a significant difference with regard to both the MO injection and the rescue (MO plus WT) injection, and null alleles are significantly different from the rescue only. Asterisks denote significance levels as determined by Student's t test. Color of asterisks as follows: blue, MO versus control; green, rescue versus MO; black, allele versus MO; red, allele versus WT rescue. Abbreviations are as follows: MO, morpholino; WT, wild-type. Error bars represent standard deviations across all examined embryos.

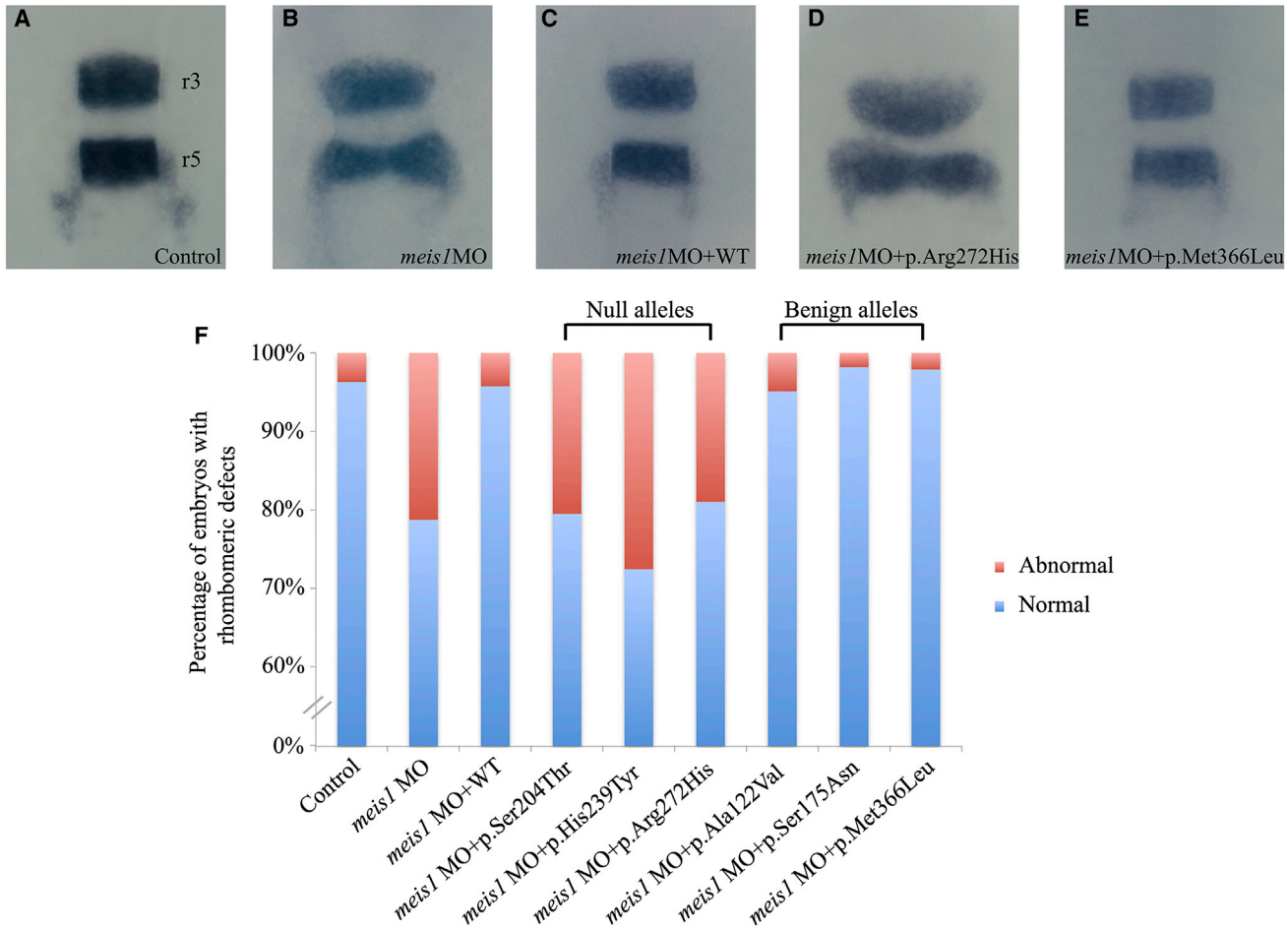


Figure 4. Functional Assessment of Null and Benign *MEIS1* Variants by In Vivo Complementation in Zebrafish Embryos and Evaluation of Hindbrain Patterning

(A–E) At 14–16 hpf, developing zebrafish embryos were evaluated for the integrity of rhombomeres 3 and 5 (r3 and r5) by in situ hybridization with a riboprobe against *krox20*. Upon disruption of *meis1*, we observed rhombomeric defects that involved widening of the evaluated structures (B and D) or shortening of the distance between r3 and r5 (D), as well as thinning or absence of the evaluated structures.

(F) Quantification showing that the aberrant phenotypes were especially pronounced in the morphant embryos and embryos coinjected with MO+null mRNA ($n \geq 26$ embryos per genotype). Abbreviations are as follows: MO, morpholino; WT, wild-type.

variation as was recently reported for sporadic Alzheimer disease[38] cannot be ascertained within the bounds of this study. We also note that we were unable to establish a single rare variant of large effect involved in RLS at the examined loci, possibly due to a lack of power in sight of the large amount of background rare genetic variation and the excess of singletons known to exist in the human genome.[39–41] However, systematic functional annotation of such singletons has improved our interpretative ability and has suggested that, in addition to the risk conferred by common and low-frequency alleles, rare variants contribute significantly to the genetic burden in RLS. Our data are consistent with previous observations wherein the rarer a genetic variant, the more likely it is to harbor a functional effect[42] and extends these to a disease context. It is particularly notable that, subsequent to functional tagging of alleles, we observed association with RLS only for functionally null variants, but not for hypomorphs, potentially intimating a threshold effect on total MEIS1 function necessary to drive pathology. Nonetheless, although our positive and negative controls for the in vivo complementation assay support previously reported high specificity and sensitivity for the approach,[43] and despite the fact that we achieved full concordance of allele effect tagging by two independent in vivo complementation assays, it will be important to validate our observations further in an independent model system; the evaluation of the two functionally null variants *MEIS1* p.Arg272His and p.Gln353His (c.1059G>C) (both RefSeq NM_002398.2/ENST00000272369) found primarily or exclusively in case subjects, located in the homeobox and transcription activation domains, respectively, in animal models could prove worthwhile to establish their relevance to the RLS phenotype and to further explore the pathophysiology of the disorder.

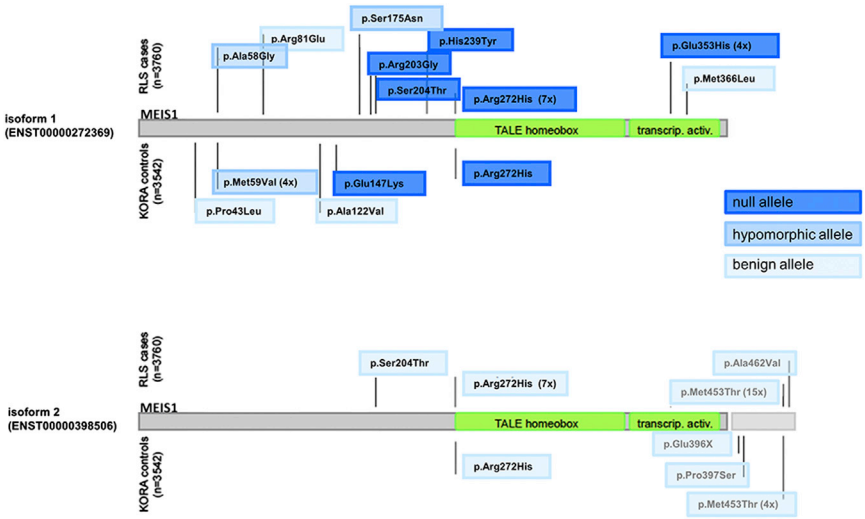


Figure 5. Functional Annotation of Rare, Nonsynonymous Variants in Isoforms 1 and 2 of *MEIS1* According to the Effect on Optic Tectum Size in Zebrafish Embryos

When tested in *MEIS1* isoform 1, an excess of rare null alleles was present among individuals with RLS. When tested in isoform 2, none of the variants show a functional effect, suggesting an isoform specificity with regard to a potential involvement in RLS. Numbers in parentheses behind variants indicate the total variant count in either case subjects or control subjects. If no number is given, the variant is a singleton.

Supplemental Data

Supplemental Data include six figures and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2014.06.005.

Acknowledgments

We are gratefully indebted to Katja Junghans, Susanne Lindhof, Jelena Golic, Sybille Frischholz, and Regina Feldmann at the Institut für Humangenetik, Helmholtz Zentrum München (Munich, Germany) for their expert technical assistance in performing Sequenom genotyping and Light Scanner analyses. Moreover, we thank Rene Rezsohazy at the Université de Louvain (Louvain, Belgium) for his advice in the construction of the domain-dead construct of *MEIS1*. This study was funded by in-house institutional funding from Technische Universität München and Helmholtz Zentrum München, a grant entitled "Functional Analysis of Rare Variants in Restless Legs Syndrome" from the Else Kröner-Fresenius-Stiftung (2013_A124), by seed funding from the Center for Human Disease Modeling, Duke University, and by P50 MH094268 to N.K. Recruitment of case and control cohorts was supported by institutional (Helmholtz Zentrum München) and government funding from the German Bundesministerium für Bildung und Forschung (03.2007-02.2011 FKZ 01ET0713). W.H.O. is a Senior Research Professor of the Charitable Hertie Foundation, Frankfurt/Main, Germany. N.K. is a Distinguished Brumley Professor.

Received: December 9, 2013
Accepted: June 10, 2014
Published: July 3, 2014

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://browser.1000genomes.org
CRAN – Package rmeta, http://cran.r-project.org/web/packages/rmeta/index.html
dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq

Accession Numbers

The dbSNP accession numbers of the variants reported in this paper are available in Table S7.

References

1. Allen, R.P., Picchietti, D., Hening, W.A., Trenkwalder, C., Walters, A.S., and Montplaisi, J.; Restless Legs Syndrome Diagnosis and Epidemiology workshop at the National Institutes of Health; International Restless Legs Syndrome Study Group (2003). Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. Sleep Med. *4*, 101–119.
2. Desai, A.V., Cherkas, L.F., Spector, T.D., and Williams, A.J. (2004). Genetic influences in self-reported symptoms of obstructive sleep apnoea and restless legs: a twin study. Twin Res. *7*, 589–595.
3. Xiong, L., Jang, K., Montplaisir, J., Levchenko, A., Thibodeau, P., Gaspar, C., Turecki, G., and Rouleau, G.A. (2007). Canadian restless legs syndrome twin study. Neurology *68*, 1631–1633.
4. Winkelmann, J., Czamara, D., Schormair, B., Knauf, F., Schulte, E.C., Trenkwalder, C., Dauvilliers, Y., Polo, O., Högl, B., Berger, K., et al. (2011). Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. PLoS Genet. *7*, e1002171.
5. Maher, B. (2008). Personal genomes: The case of the missing heritability. Nature *456*, 18–21.
6. Gibson, G. (2011). Rare and common variants: twenty arguments. Nat. Rev. Genet. *13*, 135–145.
7. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753.
8. Schulte, E.C., Knauf, F., Kemlink, D., Schormair, B., Lichtner, P., Gieger, C., Meitinger, T., and Winkelmann, J. (2011).

Variant screening of the coding regions of MEIS1 in patients with restless legs syndrome. Neurology *76*, 1106–1108.
9. Vilariño-Güell, C., Chai, H., Keeling, B.H., Young, J.E., Rajput, A., Lynch, T., Aasly, J.O., Uitti, R.J., Wszolek, Z.K., Farrer, M.J., and Lin, S.C. (2009). MEIS1 p.R272H in familial restless legs syndrome. Neurology *73*, 243–245.
10. Bonnefond, A., Clément, N., Fawcett, K., Yengo, L., Vaillant, E., Guillaume, J.L., Dechaume, A., Payne, F., Roussel, R., Czernichow, S., et al.; Meta-Analysis of Glucose and Insulin-Related Traits Consortium (MAGIC) (2012). Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat. Genet. *44*, 297–301.
11. Rivas, M.A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C.K., Boucher, G., Ripke, S., Ellinghaus, D., Burtt, N., et al.; National Institute of Diabetes and Digestive Kidney Diseases Inflammatory Bowel Disease Genetics Consortium (NIDDK IBDGC); United Kingdom Inflammatory Bowel Disease Genetics Consortium; International Inflammatory Bowel Disease Genetics Consortium (2011). Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat. Genet. *43*, 1066–1073.
12. Erickson, T., French, C.R., and Waskiewicz, A.J. (2010). Meis1 specifies positional information in the retina and tectum to organize the zebrafish visual system. Neural Dev. *5*, 22.
13. Waskiewicz, A.J., Rikhof, H.A., Hernandez, R.E., and Moens, C.B. (2001). Zebrafish Meis functions to stabilize Pbx proteins and regulate hindbrain patterning. Development *128*, 4139–4151.
14. Winkelmann, J., Schormair, B., Lichtner, P., Ripke, S., Xiong, L., Jalilzadeh, S., Fulda, S., Pütz, B., Eckstein, G., Hauk, S., et al. (2007). Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. Nat. Genet. *39*, 1000–1006.
15. Schormair, B., Kemlink, D., Roeske, D., Eckstein, G., Xiong, L., Lichtner, P., Ripke, S., Trenkwalder, C., Zimprich, A., Stiasny-Kolster, K., et al. (2008). PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. Nat. Genet. *40*, 946–948.
16. Stefansson, H., Rye, D.B., Hicks, A., Petursson, H., Ingason, A., Thorgeirsson, T.E., Palsson, S., Sigmundsson, T., Sigurdsson, A.P., Eiriksdottir, I., et al. (2007). A genetic risk factor for periodic limb movements in sleep. N. Engl. J. Med. *357*, 639–647.
17. Wichmann, H.E., Gieger, C., and Illig, T.; MONICA/KORA Study Group (2005). KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen *67* (*Suppl 1*), S26–S30.
18. Team, R.C. (2014). R: A language and environment for statistical computing (Vienna: R Foundation for Statistical Computing).
19. Morris, A.P., and Zeggini, E. (2010). An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genet. Epidemiol. *34*, 188–193.
20. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. *89*, 82–93.
21. Madsen, B.E., and Browning, S.R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet. *5*, e1000384.
22. Thisse, C., and Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat. Protoc. *3*, 59–69.
23. Cvejic, A., Serbanovic-Canic, J., Stemple, D.L., and Ouwehand, W.H. (2011). The role of meis1 in primitive and definitive hematopoiesis during zebrafish development. Haematologica *96*, 190–198.
24. Dasen, J.S., Tice, B.C., Brenner-Morton, S., and Jessell, T.M. (2005). A Hox regulatory network establishes motor neuron pool identity and target-muscle connectivity. Cell *123*, 477–491.
25. Mercader, N., Leonardo, E., Azpiazu, N., Serrano, A., Morata, G., Martínez, C., and Torres, M. (1999). Conserved regulation of proximodistal limb axis development by Meis1/Hth. Nature *402*, 425–429.
26. Barber, B.A., Liyanage, V.R., Zachariah, R.M., Olson, C.O., Bailey, M.A., and Rastegar, M. (2013). Dynamic expression of MEIS1 homeoprotein in E14.5 forebrain and differentiated forebrain-derived neural stem cells. Ann. Anat. *195*, 431–440.
27. Spieler, D., Kaffe, M., Knauf, F., Bessa, J., Tena, J.J., Giesert, F., Schormair, B., Tilch, E., Lee, H., Horsch, M., et al. (2014). Restless legs syndrome-associated intronic common variant in Meis1 alters enhancer function in the developing telencephalon. Genome Res. *24*, 592–603.
28. Catoire, H., Dion, P.A., Xiong, L., Amari, M., Gaudet, R., Girard, S.L., Noreau, A., Gaspar, C., Turecki, G., Montplaisir, J.Y., et al. (2011). Restless legs syndrome-associated MEIS1 risk variant influences iron homeostasis. Ann. Neurol. *70*, 170–175.
29. Xiong, L., Catoire, H., Dion, P., Gaspar, C., Lafrenière, R.G., Girard, S.L., Levchenko, A., Rivière, J.B., Fiori, L., St-Onge, J., et al. (2009). MEIS1 intronic risk haplotype associated with restless legs syndrome affects its mRNA and protein expression levels. Hum. Mol. Genet. *18*, 1065–1074.
30. Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J., and de Bakker, P.I. (2008). SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics *24*, 2938–2939.
31. Torgerson, D.G., Capurso, D., Mathias, R.A., Graves, P.E., Hernandez, R.D., Beaty, T.H., Bleecker, E.R., Raby, B.A., Meyers, D.A., Barnes, K.C., et al. (2012). Resequencing candidate genes implicates rare variants in asthma susceptibility. Am. J. Hum. Genet. *90*, 273–281.
32. Trynka, G., Hunt, K.A., Bockett, N.A., Romanos, J., Mistry, V., Szperl, A., Bakker, S.F., Bardella, M.T., Bhaw-Rosun, L., Castillejo, G., et al.; Spanish Consortium on the Genetics of Coeliac Disease (CEGEC); PreventCD Study Group; Wellcome Trust Case Control Consortium (WTCCC) (2011). Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat. Genet. *43*, 1193–1201.
33. Sarparanta, J., Jonson, P.H., Golzio, C., Sandell, S., Luque, H., Screen, M., McDonald, K., Stajich, J.M., Mahjneh, I., Vihola, A., et al. (2012). Mutations affecting the cytoplasmic functions of the co-chaperone DNAJB6 cause limb-girdle muscular dystrophy. Nat. Genet. *44*, 450–455, S1–S2.
34. Holm, H., Gudbjartsson, D.F., Sulem, P., Masson, G., Helgadottir, H.T., Zanon, C., Magnusson, O.T., Helgason, A., Saemundsdottir, J., Gylfason, A., et al. (2011). A rare variant in MYH6 is associated with high risk of sick sinus syndrome. Nat. Genet. *43*, 316–320.
35. Raychaudhuri, S., Iartchouk, O., Chin, K., Tan, P.L., Tai, A.K., Ripke, S., Gowrisankar, S., Vemuri, S., Montgomery, K., Yu, Y., et al. (2011). A rare penetrant mutation in CFH confers

high risk of age-related macular degeneration. Nat. Genet. *43*, 1232–1236.

36. Diogo, D., Kurreeman, F., Stahl, E.A., Liao, K.P., Gupta, N., Greenberg, J.D., Rivas, M.A., Hickey, B., Flannick, J., Thomson, B., et al.; Consortium of Rheumatology Researchers of North America; Rheumatoid Arthritis Consortium International (2013). Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis. Am. J. Hum. Genet. *92*, 15–27.
37. Johansen, C.T., Wang, J., Lanktree, M.B., Cao, H., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., et al. (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat. Genet. *42*, 684–687.
38. Cruchaga, C., Karch, C.M., Jin, S.C., Benitez, B.A., Cai, Y., Guerreiro, R., Harari, O., Norton, J., Budde, J., Bertelsen, S., et al.; UK Brain Expression Consortium; Alzheimer's Research UK Consortium (2014). Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature *505*, 550–554.
39. Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., et al.; NHLBI Exome Sequencing Project (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature *493*, 216–220.
40. Kiezun, A., Garimella, K., Do, R., Stitziel, N.O., Neale, B.M., McLaren, P.J., Gupta, N., Sklar, P., Sullivan, P.F., Moran, J.L., et al. (2012). Exome sequencing and the genetic basis of complex traits. Nat. Genet. *44*, 623–630.
41. Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., et al.; Broad GO; Seattle GO; NHLBI Exome Sequencing Project (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science *337*, 64–69.
42. Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., St Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S.A., Fraser, D., et al. (2012). An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science *337*, 100–104.
43. Zaghloul, N.A., Liu, Y., Gerdes, J.M., Gascue, C., Oh, E.C., Leitch, C.C., Bromberg, Y., Binkley, J., Leibel, R.L., Sidow, A., et al. (2010). Functional analyses of variants reveal a significant role for dominant negative and common alleles in oligogenic Bardet-Biedl syndrome. Proc. Natl. Acad. Sci. USA *107*, 10602–10607.

3.3 Common Variants

Winkelmann J, Schormair B, Czamara D, Knauf F, **Schulte EC**, Trenkwalder C, Dauvilliers Y, Polo O, Högl B, Berger K, Fuhs A, Gross N, Stiasny-Kolster K, Oertel W, Bachmann CG, Paulus W, Xiong L, Montplaisir J, Rouleau GA, Fietze I, Vavrova J, Kemlink D, Sonka K, Nevsimalova S, Lin SC, Wszolek Z, Vilarino-Guell C, Farrer MJ, Gschliesser V, Frauscher B, Falkenstetter T, Poewe W, Allen RP, Earley CJ, Ondo WG, Le WD, Spieler D, Kaffe M, Zimprich A, Kettunen J, Perola M, Silander K, Cournu-Rebeix I, Francavilla M, Fontenille C, Fontaine B, Vodicka P, Prokisch H, Lichtner P, Peppard P, Faraco J, Mignot E, Gieger C, Illig T, Wichmann HE, Müller-Myhsok B, Meitinger T. Genome-wide association study identifies novel Restless Legs Syndrome susceptibility loci on 2p14 and 16q12.1. PLoS Genetics 2011, 7: e1002171. PMID: 21779176 **(IF 2011: 10.2)**

Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, Adams MJ, Agerbo E, Air TM, Andlauer TMF, Bacanu SA, Baekvad-Hansen M, Beekman AFT, Bigdeli TB, Binder EB, Blackwood DRH, Bryois J, Buttenschon HN, Bybjerg-Grauholm J, Cai N, Castelao E, Christensen JH, Clarke TK, Coleman JIR, Colodro-Conde L, Couvy-Duchesne B, Craddock N, Crawford GE, Crowley CA, Dashti HS, Davies G, Deary IJ, Degenhardt F, Derks EM, Direk N, Dolan CV, Dunn EC, Fley TC, Eriksson N, Escott-Price V, Kideh FHF, Finucane HK, Forstner AJ, Frank J, Gaspar HA, Gill M, Giusti-Rodrigues P, Goes FS, Gordon SD, Grove J, Hall LS, Hannon E, Hansen CS, Hansen TF, Herms S, Hickie IB, Hoffmann P, Homuth G, Horn C, Hottenga JJ, Hougaard DM, Hu M, Hyde CL, Ising M, Jansen R, Jin F, Jorgenson E, Knowles JA, Kohane IS, Kraft J, Kretzschmar WW, Krogh J, Kutalik Z, Lane JM, Li Y, Li Y, Lind PA, Liu X, Lu L, MacIntyre DJ, MacKinnon DF, Maier RM, Maier W, Marchini J, Mbarek H, McGrath P, McGuffin P, Medland SE, Mehta D, Middeldorp CM, Mihailov E, Milaneschi Y, Milani L, Mill J, Mondimore FM, Montgomery GW, Mostafavi S, Mullins N, Nauck M, Ng B, Nivard MG, Nyholt DR, O'Reilly PF, Oskarsson H, Owen MJ, Painter JN, Pedersen CB, Pedersen MG, Peterson RE, Pettersson E, Peyrot WJ, Pistis G, Posthuma D, Purcell SM, Quiroz JA, Qvist P, Rice JP, Riley BP, Rivera M, Saeed Mirza S, Saxena R, Schoevers R, **Schulte EC**, Shen L, Shi J, Shyn SI, Sigurdsson E, Sinnamon GBC, Smit JH, Smith DJ, Stefansson H, Steinberg S, Stockmeier CA, Streit F, Strohmaier J, Tansey KE, Teismann H, Teumer A, Thompson W, Thomson PA, Thorgeirsson TE, Tian C, Traylor M, Treutlein J, Trubetskoy V, Uitterlinden AG, Umbricht D, Van der Auwera S, van Hemert AM, Viktorin A, Visscher PM, Wang Y, Webb BT, Weinsheimer SM, Wellmann J, Willemsen G, Witt SH, Wu Y, Xi HS, Yang J, Zhang F; eQTLGen; 23andMe, Arolt V, Baune BT, Berger K, Boomsma DI, Cichon S, Dannlowski U, de Geus ECJ, DePaulo JR, Domenici E, Domschke K, Esko T, Grabe HJ, Hamilton SP, Hayward C, Heath AC, Hinds DA, Kendler KS, Kloiber S, Lewis G, Li QS, Lucae S, Madden PFA, Magnusson PK, Martin NG, McIntosh AM, Metspalu A, Mors O, Mortensen PB, Müller-Myhsok B, Nordentoft M, Nöthen MM, O'Donovan MC, Paciga SA, Pedersen NL, Penninx BWJH, Perlis RH, Porteous DJ, Potash JB, Preisig M, Rietschel M, Schaefer C, Schulze TG, Smoller JW, Stefansson K, Tiemeier H, Uher R, Völzke H, Weissman MM, Werge T, Winslow AR, Lewis CM, Levinson DF, Breen G, Børglum AD, Sullivan PF; **Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium**. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 2018, 50: 668-681. PMID: 29700475. **(IF 2018: 25.5)**

Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, Als TD, Bigdeli TB, Børte S, Bryois J, Charney AW, Drange OK, Gandal MJ, Hagenaars SP, Ikeda M, Kamitaki N, Kim M, Krebs K, Panagiotaropoulou G, Schilder BM, Sloofman LG, Steinberg S, Trubetskoy V, Winsvold BS, Won HH, Abramova L, Adorjan K, Agerbo E, Al Eissa M, Albani D, Alliey-Rodriguez N, Anjorin A, Antilla V, Antoniou A, Awasthi S, Baek JH, Bækvad-Hansen M, Bass N, Bauer M, Beins EC, Bergen SE, Birner A, Bøcker Pedersen C, Bøen E, Boks MP, Bosch R, Brum M, Brumpton BM, Brunkhorst-Kanaan N, Budde M, Bybjerg-Grauholm J, Byerley W, Cairns M, Casas M, Cervantes P, Clarke TK, Cruceanu C, Cuellar-Barboza A, Cunningham J, Curtis D, Czerski PM, Dale AM, Dalkner N, David FS, Degenhardt F, Djurovic S, Dobbyn AL, Douzenis A, Elvsåshagen T, Escott-Price V, Ferrier IN, Fiorentino A, Foroud TM, Forty L, Frank J, Frei O, Freimer NB, Frisén L, Gade K, Garnham J, Gelernter J, Giørtz Pedersen M, Gizer IR, Gordon SD, Gordon-Smith K, Greenwood TA, Grove J, Guzman-Parra J, Ha K, Haraldsson M, Hautzinger M, Heilbronner U, Hellgren D, Herms S, Hoffmann P, Holmans PA, Huckins L, Jamain S, Johnson JS, Kalman JL, Kamatani Y, Kennedy JL, Kittel-Schneider S, Knowles JA, Kogevinas M, Koromina M, Kranz TM, Kranzler HR, Kubo M, Kupka R, Kushner SA, Lavebratt C, Lawrence J, Leber M, Lee HJ, Lee PH, Levy SE, Lewis C, Liao C, Lucae S, Lundberg M, MacIntyre DJ, Magnusson SH, Maier W, Maihofer A, Malaspina D, Maratou E, Martinsson L, Mattheisen M, McCarroll SA, McGregor NW, McGuffin P, McKay JD, Medeiros H, Medland SE, Millischer V, Montgomery GW, Moran JL, Morris DW, Mühleisen TW, O'Brien N, O'Donovan C, Olde Loohuis LM, Oruc L, Papiol S, Pardiñas AF, Perry A, Pfennig A, Porichi E, Potash JB, Quested D, Raj T, Rapaport MH, DePaulo JR, Regeer EJ, Rice JP, Rivas F, Rivera M, Roth J, Roussos P, Ruderfer DM, Sánchez-Mora C, **Schulte EC**, Senner F, Sharp S, Shilling PD, Sigurdsson E, Sirignano L, Slaney C, Smeland OB, Smith DJ, Sobell JL, Søholm Hansen C, Soler Artigas M, Spijker AT, Stein DJ, Strauss JS, Świątkowska B, Terao C, Thorgeirsson TE, Toma C, Tooney P, Tsermpini EE, Vawter MP, Vedder H, Walters JTR, Witt SH, Xi S, Xu W, Yang JMK, Young AH, Young H, Zandi PP, Zhou H, Zillich L; HUNT All-In Psychiatry, Adolfsson R, Agartz I, Alda M, Alfredsson L, Babadjanova G, Backlund L, Baune BT, Bellivier F, Bengesser S, Berrettini WH, Blackwood DHR, Boehnke M, Børglum AD, Breen G, Carr VJ, Catts S, Corvin A, Craddock N, Dannlowski U, Dikeos D, Esko T, Etain B, Ferentinos P, Frye M, Fullerton JM, Gawlik M, Gershon ES, Goes FS, Green MJ, Grigoroiu-Serbanescu M, Hauser J, Henskens F, Hillert J, Hong KS, Hougaard DM, Hultman CM, Hveem K, Iwata N, Jablensky AV, Jones I, Jones LA, Kahn RS, Kelsoe JR, Kirov G, Landén M, Leboyer M, Lewis CM, Li QS, Lissowska J, Lochner C, Loughland C, Martin NG, Mathews CA, Mayoral F, McElroy SL, McIntosh AM, McMahon FJ, Melle I, Michie P, Milani L, Mitchell PB, Morken G, Mors O, Mortensen PB, Mowry B, Müller-Myhsok B, Myers RM, Neale BM, Nievergelt CM, Nordentoft M, Nöthen MM, O'Donovan MC, Oedegaard KJ, Olsson T, Owen MJ, Paciga SA, Pantelis C, Pato C, Pato MT, Patrinos GP, Perlis RH, Posthuma D, Ramos-Quiroga JA, Reif A, Reininghaus EZ, Ribasés M, Rietschel M, Ripke S, Rouleau GA, Saito T, Schall U, Schalling M, Schofield PR, Schulze TG, Scott LJ, Scott RJ, Serretti A, Shannon Weickert C, Smoller JW, Stefansson H, Stefansson K, Stordal E, Streit F, Sullivan PF, Turecki G, Vaaler AE, Vieta E, Vincent JB, Waldman ID, Weickert TW, Werge T, Wray NR, Zwart JA, Biernacka JM, Nurnberger JI, Cichon S, Edenberg HJ, Stahl EA, McQuillin A, Di Florio A, Ophoff RA, Andreassen OA. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet 2021, 53:817-829. PMID: 34002096. **(IF 2021: 23.1)**

Common genetic risk factors for neuropsychiatric disorders identified by GWAS provide the backbone of the genetic architecture of most common neuropsychiatric disorders. And GWAS have been instrumental in identifying the broad biological underpinnings of neuropsychiatric disorders ranging from RLS to MDD, sometimes for the first time ever (e.g.(Winkelmann et al., 2007)). Yet, although extremely high polygenicity with contributions of hundreds, thousands, or tens of thousands of common variants of very small effect on the phenotype has been firmly established for many common, complex psychiatric disorders, here, too the conundrum that the currently known associated common variants only account for a minor portion of the heritability prevails, leaving a large portion of the heritability still "missing". For example, the most recent appraisals estimate that when taken together, all known common genetic risk factors for BPD only account for approximately 4.57 % of the phenotypic variance(Mullins et al., 2021). Beyond an inflection point specific to each disorder or trait, there is a linear relationship between the number of case and control individuals included in a GWAS and the number of novel loci that are discovered (Kendall et al., 2021).

Accordingly, continuing efforts to increase the size of well-phenotyped study samples, most prominently by the Psychiatric Genomics Consortium (PGC), are important to complete the picture of the contribution of common genetic variants to the genetic architecture of neuropsychiatric traits. Three such large international consortial efforts are presented here. For three neuropsychiatric disorders—MDD, BPD, and RLS—, GWAS identified a total of 60 novel genomic loci(Mullins et al., 2021; Winkelmann et al., 2011; Wray et al., 2018). The number of cases and controls included in the studies ranged from 4,857 (RLS) all the way up to 135,458 (MDD) cases and 7,380 (RLS) to 371,549 (BPD) controls. The GWAS also added important knowledge to our understanding of the biological basis of these disorders: RLS was established as a disorder with an important neurodevelopmental component(Winkelmann et al., 2011), surprising for a disorder with a clear age-dependent prevalence(Schulte, 2015); MDD loci highlighted a role of neurons—not microglia or astrocytes—in the prefrontal as well as the cingulate cortex as well as a cross-diagnostic genetic architecture(Wray et al., 2018); and synaptic signaling as well as neurons of the prefrontal cortex and the hippocampus emerged as important biological determinants of BPD(Mullins et al., 2021).

It is interesting to note that while all disorders in question are clearly common, complex genetic disorders, very different sizes of GWAS yield different numbers of associated genetic loci. For RLS, for example, the first loci of genome-wide significance were found in a case-control samples of 401 cases and 1644 controls(Winkelmann et al., 2007). For MDD, on the other hand, it took 5,303 cases and 5,337 controls and almost a decade longer to identify the first genome-wide significant loci(Converge, 2015). Similarly, BPD is a highly polygenic disorder with more than 8,600 variants estimated to contribute to its genetic makeup(Mullins et al., 2021). Yet, when compared to a schizophrenia (SCZ) GWAS of similar size(Pardinas et al., 2018), fewer genome-wide significant loci were identified for BPD(Mullins et al., 2021). There are many possible explanations for this phenomenon ranging from clinical heterogeneity all the way to different genetic architectures with different contributions from variants of different classifications (e.g. structural variants like CNVs versus non-structural variants like SNVs, different frequencies, or different ORs).

OPEN ACCESS Freely available online

PLoS GENETICS

Genome-Wide Association Study Identifies Novel Restless Legs Syndrome Susceptibility Loci on 2p14 and 16q12.1

Juliane Winkelmann[1,2,3]*, Darina Czamara[4,9], Barbara Schormair[1,3,9], Franziska Knauf[1,3], Eva C. Schulte[2], Claudia Trenkwalder[5], Yves Dauvilliers[6], Olli Polo[7,8], Birgit Högl[9], Klaus Berger[10], Andrea Fuhs[10], Nadine Gross[2], Karin Stiasny-Kolster[11,12], Wolfgang Oertel[12], Cornelius G. Bachmann[13], Walter Paulus[13], Lan Xiong[14], Jacques Montplaisir[15,16], Guy A. Rouleau[14], Ingo Fietze[17], Jana Vávrová[18], David Kemlink[18], Karel Sonka[18], Sona Nevsimalova[18], Siong-Chi Lin[19], Zbigniew Wszolek[19], Carles Vilariño-Güell[19], Matthew J. Farrer[19], Viola Gschliesser[9], Birgit Frauscher[9], Tina Falkenstetter[9], Werner Poewe[9], Richard P. Allen[20], Christopher J. Earley[20], William G. Ondo[21], Wei-Dong Le[21], Derek Spieler[1,3], Maria Kaffe[2,3], Alexander Zimprich[22], Johannes Kettunen[23,24], Markus Perola[23,24], Kaisa Silander[23,24], Isabelle Cournu-Rebeix[25,26,27], Marcella Francavilla[25,26,27], Claire Fontenille[25,26,27], Bertrand Fontaine[25,26,27], Pavel Vodicka[28], Holger Prokisch[1,3], Peter Lichtner[1,3], Paul Peppard[29], Juliette Faraco[30], Emmanuel Mignot[30], Christian Gieger[31], Thomas Illig[32], H.-Erich Wichmann[33,34,35], Bertram Müller-Myhsok[4], Thomas Meitinger[1,3]

1 Institute of Human Genetics, Technische Universität München, Munich, Germany, **2** Department of Neurology, Technische Universität München, Munich, Germany, **3** Institute of Human Genetics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, **4** Max Planck Institute of Psychiatry, Munich, Germany, **5** Paracelsus-Elena-Hospital, Kassel, Germany, **6** Unité du Sommeil, Service de Neurologie, Hôpital Gui-de-Chauliac, INSERM U1061, Montpellier, France, **7** Department of Pulmonary Medicine, Tampere University Hospital, Tampere, Finland, **8** Sleep Research Unit, University of Turku, Turku, Finland, **9** Department of Neurology, Innsbruck Medical University, Innsbruck, Austria, **10** Institute of Epidemiology and Social Medicine, University Münster, Münster, Germany, **11** Somnomar, Sleep Research Institute, Marburg, Germany, **12** Department of Neurology, Center of Nervous Diseases, Philipps University, Marburg, Germany, **13** Department of Clinical Neurophysiology, University of Göttingen, Göttingen, Germany, **14** Centre of Excellence in Neuromics, CHUM Research Centre and the Department of Medicine, University of Montreal, Montreal, Canada, **15** Laboratoire d'étude des maladies du cerveau, Centre de recherche du CHUM, Hôpital Notre-Dame, Université de Montréal, Montréal, Canada, **16** Centre d'étude du sommeil, Hôpital du Sacré-Coeur de Montréal, Montréal, Canada, **17** Charite – Universitätsmedizin Berlin Interdisciplinary Center of Sleep Medicine, Berlin, Germany, **18** Department of Neurology, 1st Faculty of Medicine, Charles University, Prague, Czech Republic, **19** Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, Canada, **20** Department of Neurology, Johns Hopkins University, Baltimore, Maryland, United States of America, **21** Department of Neurology, Baylor College of Medicine, Houston, Texas, United States of America, **22** Department of Neurology, Medical University of Vienna, Vienna, Austria, **23** Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland, **24** Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland, **25** INSERM, UMR_S975, Paris, France, **26** Centre de Recherche Institut du Cerveau et de la Moelle, CNRS 7225, Paris, France, **27** Fédération des maladies du système nerveux, Pitié – Salpêtrière Hospital, AP-HP, Paris, France, **28** Institute of Experimental Medicine, Czech Academy of Sciences, Prague, Czech Republic, **29** Department of Population Health Sciences, University of Wisconsin, Madison, Wisconsin, United States of America, **30** Center For Narcolepsy, Stanford University, Palo Alto, California, United States of America, **31** Institute of Genetic Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, **32** Unit for Molecular Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, **33** Institute of Epidemiology I, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, **34** Institute of Medical Informatics, Biometry, and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany, **35** Klinikum Grosshadern, Munich, Germany

Abstract

Restless legs syndrome (RLS) is a sensorimotor disorder with an age-dependent prevalence of up to 10% in the general population above 65 years of age. Affected individuals suffer from uncomfortable sensations and an urge to move in the lower limbs that occurs mainly in resting situations during the evening or at night. Moving the legs or walking leads to an improvement of symptoms. Concomitantly, patients report sleep disturbances with consequences such as reduced daytime functioning. We conducted a genome-wide association study (GWA) for RLS in 922 cases and 1,526 controls (using 301,406 SNPs) followed by a replication of 76 candidate SNPs in 3,935 cases and 5,754 controls, all of European ancestry. Herein, we identified six RLS susceptibility loci of genome-wide significance, two of them novel: an intergenic region on chromosome 2p14 (rs6747972, $P = 9.03 \times 10^{-11}$, OR = 1.23) and a locus on 16q12.1 (rs3104767, $P = 9.4 \times 10^{-19}$, OR = 1.35) in a linkage disequilibrium block of 140 kb containing the 5′-end of *TOX3* and the adjacent non-coding RNA *BC034767*.

Citation: Winkelmann J, Czamara D, Schormair B, Knauf F, Schulte EC, et al. (2011) Genome-Wide Association Study Identifies Novel Restless Legs Syndrome Susceptibility Loci on 2p14 and 16q12.1. PLoS Genet 7(7): e1002171. doi:10.1371/journal.pgen.1002171

Editor: Mark I. McCarthy, University of Oxford, United Kingdom

Received December 7, 2010; **Accepted** May 24, 2011; **Published** July 14, 2011

Funding: The replication phase was supported by a grant from the US RLS Foundation. Part of this work was financed by the National Genome Research Network (NGFN). The KORA study group consists of H-E Wichmann (speaker), R Holle, J John, T Illig, C Meisinger, A Peters, and their coworkers, who are responsible for the design and conduction of the KORA studies. The KORA research platform (KORA, Cooperative Research in the Region of Augsburg) was initiated and financed by the Helmholtz Zentrum München, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. The collection of sociodemographic and clinical data in the Dortmund Health Study was supported by the German Migraine & Headache Society (DMKG) and by unrestricted grants of equal share from Astra Zeneca, Berlin Chemie, Boots Healthcare, Glaxo-Smith-Kline, McNeil Pharma (former Woelm Pharma), MSD Sharp & Dohme, and Pfizer to the University of Muenster. Blood collection in the Dortmund Health Study was done through funds from the Institute of Epidemiology and Social Medicine, University of Muenster. Data collection in the COR-Study was supported by unrestricted grants of the German RLS Society (Deutsche Restless Legs Vereinigung e.V.) and Axxonis Pharma, Boehringer Ingelheim Pharma, Mundipharma Research, Roche Pharma, and UCB to the University of Muenster. CG Bachmann was supported by grants of the German RLS Society, Deutsche Restless Legs Vereinigung, e.V. H Prokisch and T Meitinger were supported by the German Federal Ministry of Education and Research (BMBF) project Systems Biology of Metabotypes (SysMBo #0315494A). RP Allen and CJ Earley were supported by the grant PO1-AG21190 National Institute of Health, USA; the Canadian part of the study was supported by a Canadian Institutes of Health Research (CIHR) grant to GA Rouleau and J Montplaisir. D Kemlink and S Nevsimalova were supported by an ESRS grant MSM0021620849; Jávrová and K Sonka were supported by grant MSM0021620816. Recruitment of Czech controls was funded by grant IGA NR 8563-5, Ministry of Health of the Czech Republic. B Frauscher, I Cournu-Rebeix, M Francavilla, and C Fontenille are co-authors on behalf of BRC-REFGENSEP, which is supported by INSERM, AFM (Généthon), ARSEP, and GIS-IBISA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: winkelmann@lrz.tu-muenchen.de

◔ These authors contributed equally to this work.

Introduction

Restless legs syndrome (RLS) is a common neurological disorder with a prevalence of up to 10 %, which increases with age [1]. Affected individuals suffer from an urge to move due to uncomfortable sensations in the lower limbs present in the evening or at night. The symptoms occur during rest and relaxation, with walking or moving the extremity leading to prompt relief. Consequently, initiation and maintenance of sleep become defective [1]. RLS has been associated with iron deficiency, and is pharmacologically responsive to dopaminergic substitution. Increased cardiovascular events, depression, and anxiety count among the known co-morbidities [1].

Genome-wide association studies (GWAs) identified genetic risk factors within *MEIS1*, *BTBD9*, *PTPRD*, and a locus encompassing *MAP2K5* and *SKOR1* [2–4]. To identify additional RLS susceptibility loci, we undertook an enlarged GWA in a German case-control population, followed by replication in independent case-control samples originating from Europe, the United States of America, and Canada. In doing so, we identified six RLS susceptibility loci with genome-wide significance in the joint analysis, two of them novel: an intergenic region on chromosome 2p14 and a locus on 16q12.1 in close proximity to *TOX3* and the adjacent non-coding RNA *BC034767*.

Results/Discussion

We enlarged our previously reported [2,4] GWA sample to 954 German RLS cases and 1,814 German population-based controls from the KORA-S3/F3 survey and genotyped them on Affymetrix 5.0 (cases) and 6.0 (controls) arrays. To correct for population stratification, as a first step, we performed a multidimensional scaling (MDS) analysis, leading to the exclusion of 18 controls as outliers. In a second step, we conducted a variance components analysis to identify any residual substructure in the remaining samples, resulting in an inflation factor λ of 1.025 (Figures S1 and S2). The first four axes of variation from the MDS analysis were included as covariates in the association analysis of the genome-wide stage and all P-values were corrected for the observed λ.

Prior to statistical analysis, genotyping data was subjected to extensive quality control. We excluded a total of 302 DNA samples due to a genotyping call rate <98 %. For individual SNP quality control, we adopted a stringent protocol in order to account for the complexity of an analysis combining 5.0 and 6.0 arrays. We excluded SNPs with a minor allele frequency (MAF) <5%, a callrate <98%, or a significant deviation from Hardy-Weinberg Equilibrium (HWE) in controls ($P<0.00001$). In addition, we dropped SNPs likely to be false-positive associations due to differential clustering between 5.0 and 6.0 arrays by adding a second set of cases of an unrelated phenotype and discarding SNPs showing association in this setup (see Materials and Methods). Finally, we tested 301,406 SNPs for association in 922 cases and 1,526 controls. Based on a threshold level of a nominal λ-corrected $P_{GWA}<10^{-4}$, a total of 47 SNPs distributed over 26 loci were selected for follow-up in the replication study (Figure 1, Table S1).

We genotyped these 47 SNPs together with 29 adjacent SNPs in strong linkage disequilibrium (LD, $r^2 = 0.5$–0.9) using the Sequenom iPLEX platform in seven case-control populations of European descent, comprising a total of 3,935 cases and 5,754 controls. Eleven SNPs with a call rate <95%, MAF<5%, and $P<0.00001$ for deviation from HWE in controls as well as 432 samples with a genotyping call rate <90% were excluded. A set of 47 SNPs, genotyped in 186 samples on both platforms (Affymetrix and Sequenom), was used to calculate an average concordance rate of 99.24 %.

The combined analysis of all replication samples confirmed the known four susceptibility loci and, in addition, identified two novel association signals on chromosomes 2p14 and 16q12.1 (Table 1). To address possible population stratification within the combined replication sample, we performed a fixed-effects meta-analysis. For four of the replication case-control populations, we included λ inflation factors which were available from a genomic controls experiment in a previous study in these populations [4]. These were used to correct the estimates for the standard error. Joint analysis of GWA and all replication samples showed genome-wide significance for these two novel loci as well as for the known RLS loci in *MEIS1*, *BTBD9*, *PTPRD*, and *MAP2K5/SKOR1* with a nominal λ -corrected $P_{JOINT} <5\times10^{-8}$ (Table 1). Depending on the variable power to detect the effects, the separate analyses of individual subsamples in the replication either confirmed the association after correction for multiple testing or yielded nominally significant results (Tables S2 and S3). The differing relevance of the risk loci in the individual samples is illustrated in

Author Summary

Restless legs syndrome (RLS) is one of the most common neurological disorders. Patients with RLS suffer from an urge to move the legs and unpleasant sensations located mostly deep in the calf. Symptoms mainly occur in resting situations in the evening or at night. As a consequence, initiation and maintenance of sleep become defective. Here, we performed a genome-wide association study to identify common genetic variants increasing the risk for disease. The genome-wide phase included 922 cases and 1,526 controls, and candidate SNPs were replicated in 3,935 cases and 5,754 controls, all of European ancestry. We identified two new RLS–associated loci: an intergenic region on chromosome 2p14 and a locus on 16q12.1 in a linkage disequilibrium block containing the 5′-end of *TOX3* and the adjacent non-coding RNA *BC034767*. *TOX3* has been implicated in the development of breast cancer. The physiologic role of TOX3 and BC034767 in the central nervous system and a possible involvement of these two genes in RLS pathogenesis remain to be established.

forest plots (Figure 2). There was no evidence of epistasis between any of the six risk loci ($P_{Bonferroni} > 0.45$).

The association signal on 2p14 (rs6747972: nominal λ-corrected $P_{JOINT} = 9.03 \times 10^{-11}$, odds ratio (OR) = 1.23) is located in an LD block of 120 kb within an intergenic region 1.3 Mb downstream of *MEIS1* (Figure 3). Assuming a long-range regulatory function of the SNP-containing region, *in silico* analysis for clusters of highly conserved non-coding elements using the ANCORA browser (http://ancora.genereg.net) identified *MEIS1* as well as *ETAA1* as potential target genes [5,6].

The second locus on chromosome 16q12.1 (rs3104767: nominal λ-corrected $P_{JOINT} = 9.4 \times 10^{-19}$, OR = 1.35) is located within an LD block of 140 kb (Figure 3), which contains the 5′UTR of *TOX3* (synonyms *TNRC9* and *CAGF9*) and the non-coding RNA *BC034767* (synonym *LOC643714*). *TOX3* is a member of the high mobility box group family of non-histone chromatin proteins which interacts with *CREB* and *CBP* and plays a critical role in mediating calcium-dependent transcription in neurons [7]. GWAs have identified susceptibility variants for breast cancer in the identical region [8]. The best-associated breast cancer SNP, rs3803662, is in low LD ($r^2 \sim 0.1$, HapMap CEU data) with rs3104767, but showed association to RLS (λ-corrected nominal $P_{GWA} = 7.29 \times 10^{-7}$). However, logistic regression analysis conditioned on rs3104767 demonstrated that this association is dependent on rs3104767 (rs3803662: $P_{GWA/conditioned} = 0.2883$).

BC034767 is represented in GenBank by two identical mRNA transcripts, BC034767 and BC029912. According to the gene model information of the UCSC and Ensembl genome browsers (http://genome.ucsc.edu and http://www.ensembl.org/index.html), these mRNAs are predicted to be non-coding. Additional *in silico* analysis using the Coding Potential Calculator (http://cpc.cbi.pku.edu.cn) supported this by attributing only a weak coding potential to this RNA, suggesting a regulatory function instead [9]. We also searched for rare alleles with strong effects and performed a mutation screening by sequencing all coding and non-coding

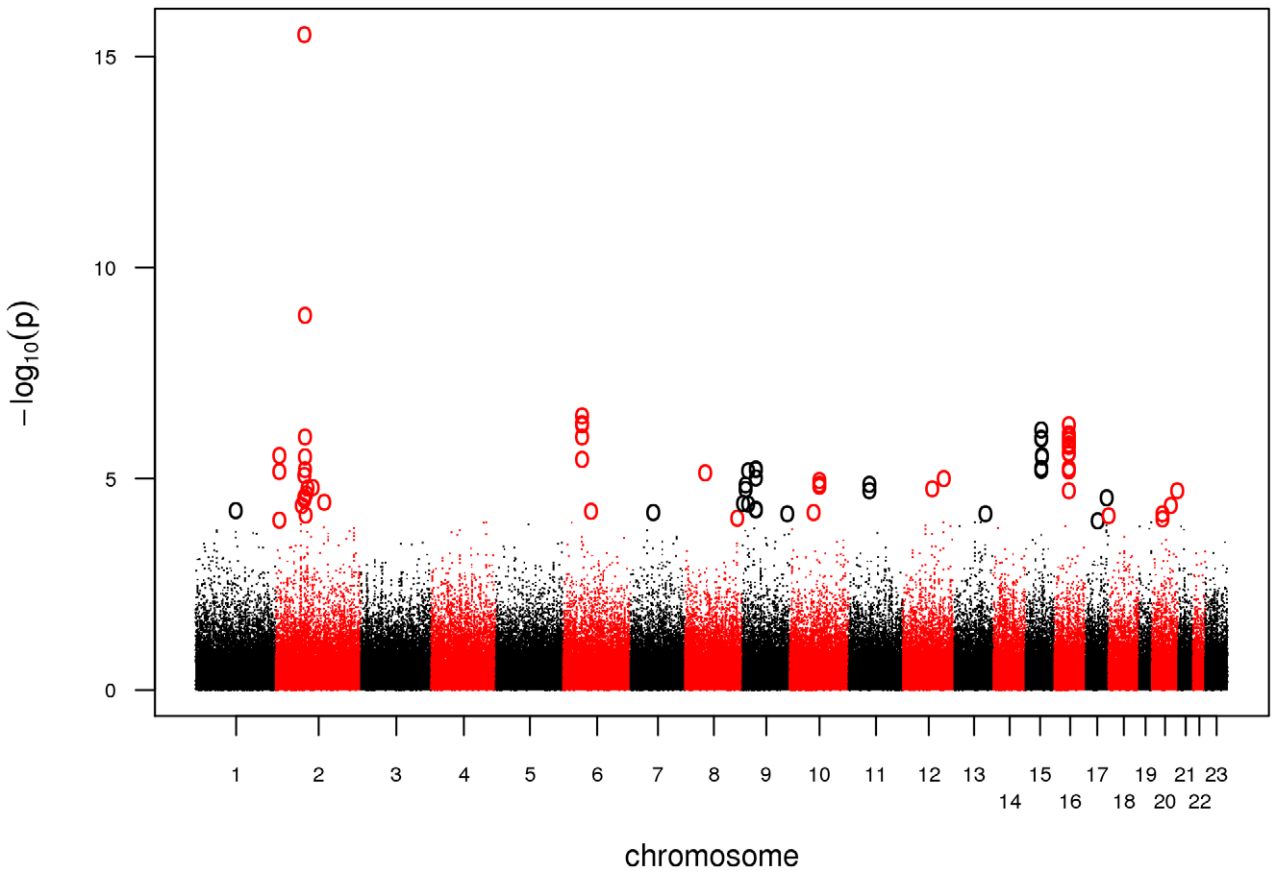


Figure 1. Manhattan plot of the GWA. Association results of the GWA stage. The x-axis represents genomic position along the 22 autosomes and the x-chromosome, the y-axis shows -log10(P) for each SNP assayed. SNPs with a nominal λ-corrected $P < 10^{-4}$ are highlighted as circles.
doi:10.1371/journal.pgen.1002171.g001

exons of *TOX3* and *BC034767* in 188 German RLS cases (Table S4). In *TOX3*, a total of nine variants not listed in dbSNP (Build 130) were found, three of which are non-synonymous. Only one of these is also annotated in the 1000 Genomes project (November 2010 data release). Three additional new variants were located in putative exons 1 and 2 of *BC034767*. Analysis of the frequency of these variants as well as all known non-synonymous, frameshift, and splice-site coding SNPs in *TOX3* in a subset of one of the replication samples (726 cases and 735 controls from the GER1 sample) did not reveal any association to RLS. For a power of >80%, however, variants with an OR above 4.5 and a MAF ≥0.01 would be required. For even lower MAFs, ORs ≥10 would be necessary for sufficient power. Furthermore, the described CAG repeat within exon 7 of *TOX3* was not polymorphic as shown by fragment analysis in 100 population-based controls.

According to publicly available expression data (http://genome.ucsc.edu), in humans, *BC034767* is expressed in the testes only, while *TOX3* expression has been shown in the salivary glands, the trachea, and in the CNS. Detailed in-depth real time PCR profiling of *TOX3* showed high expression levels in the frontal and occipital cortex, the cerebellum, and the retina [10]. To assess a putative eQTL function of rs6747972 or rs3104767, we studied the SNP-genotype-dependent expression of *TOX3* and *BC034767* as well as of genes known to directly interact with *TOX3* (*CREB-1/CREBBP/CITED1*) and potential target genes of long-range regulatory elements at the locus on chromosome 2 (*MEIS1/ETAA1*) in RNA expression microarray data from peripheral blood in 323 general population controls [11]. No differential genotype-dependent expression variation was found.

To assess the potential for genetic risk prediction, we split our GWA sample in a training and a test set and determined classifiers for case-control status in the training set to predict case-control status in the test set. Training and test set were independent of each other – not only with respect to included individuals but also with respect to the genotyping procedure as we used genotypes generated on different genotyping platforms. As training set, we used those cases of the current GWA which had been genotyped on 500K arrays in a previous GWA and the corresponding control set [2], in total, 326 cases and 1,498 controls. The test set comprised 583 cases and 1,526 controls, genotyped on 5.0/6.0 arrays as part of the current study. Prior to the analysis, we removed the six known risk loci and performed LD-pruning to limit the analysis to SNPs not in LD with each other. In the end, a total of 76,532 SNPs were included in the pruned dataset. We conducted logistic regression with age and sex as covariates. Based on these association results, the sum score of SNPs showing the most significant effects (i.e. the number of risk alleles over all SNPs) weighted by the ln(OR) of these effects was chosen as predictor variable in the test set. We then varied the P-value threshold for SNPs included in the sum score. For a P-value <0.6, we observed a maximum area under the curve (AUC) of 63.9% and an explained genetic variance of 6.6% (Nagelkerke's R), values comparable to estimates obtained for other complex diseases such as breast cancer or diabetes (Table S5) [12–14]. Inclusion of the six known risk loci in this analysis resulted in a maximum AUC of 64.2% and an explained genetic variance of 6.8%.

Additionally, we performed risk prediction in the combined GWA and replication sample including only the six established RLS risk loci. For this purpose, we used the weighted risk allele score resulting in ORs of up to 8.6 (95% CI: 2.46–46.25) and an AUC of 65.1% (Figures S3 and S4).

By increasing the size of our discovery sample, we have identified two new RLS susceptibility loci. The top six loci show effect sizes between 1.22 and 1.77 and risk allele frequencies between 19 and 82%, and reveal genes in neuronal transcription pathways not previously suspected to be involved in the disorder.

Materials and Methods

Study population and phenotype assessment

Ethics statement. Written informed consent was obtained from each participant in the respective language. The study has

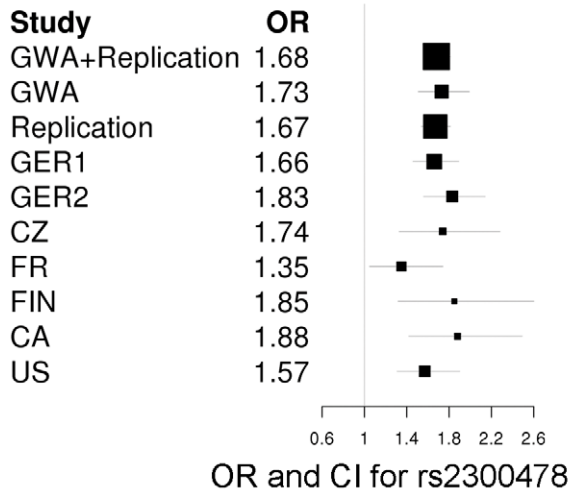
Table 1. Association results of GWA and joint analysis of GWA and replication.

Chr	Locus	LD block (Mb)	SNP	Position (bp)	Risk allele	Risk allele frequency cases/controls	P_{GWA}	$P_{REPLICATION}$	P_{JOINT}	Odds ratio (95% CI)
Known risk loci (1 SNP per locus)										
2	MEIS1	66.57–66.64	rs2300478	66634957	G	0.35/0.24	7.77×10^{-16}	4.39×10^{-35}	3.40×10^{-49}	1.68 (1.57–1.81)
6	BTBD9	37.82–38.79	rs9357271	38473851	T	0.82/0.76	6.74×10^{-7}	2.01×10^{-16}	7.75×10^{-22}	1.47 (1.35–1.47)
9	PTPRD	8.80–8.88	rs1975197	8836955	A	0.19/0.16	4.94×10^{-5}	1.07×10^{-6}	3.49×10^{-10}	1.29 (1.19–1.40)
15	MAP2K5/SKOR1	65.25–65.94	rs12593813	65823906	G	0.75/0.68	1.49×10^{-6}	1.54×10^{-17}	1.37×10^{-22}	1.41 (1.32–1.52)
New genome-wide significant loci ($P_{JOINT} < 5.2\times10^{-8}$)										
2	intergenic region	67.88–68.00	rs6747972	67923729	A	0.47/0.44	1.37×10^{-6}	3.73×10^{-6}	9.03×10^{-11}	1.23 (1.16–1.31)
			rs2116050	67926267	G	0.49/0.47	7.84×10^{-6}	4.85×10^{-6}	4.83×10^{-10}	1.22 (1.15–1.30)
16	TOX3/BC034767	51.07–51.21	rs3104767	51182239	G	0.65/0.58	7.38×10^{-7}	2.16×10^{-13}	9.40×10^{-19}	1.35 (1.27–1.43)
			rs3104788	51196004	T	0.65/0.58	1.19×10^{-6}	2.42×10^{-13}	1.63×10^{-18}	1.33 (1.25–1.43)

RLS-associated SNPs with genome-wide significance. P_{GWA}, λ-corrected nominal P-value of GWA stage. $P_{REPLICATION}$, nominal P-value obtained from meta-analysis of the replication stage samples. P_{JOINT}, nominal P-value of the joint meta-analysis of GWA and replication stage, λ-corrected in samples where λ-values were available. Nominal P-values in GWA were calculated using logistic regression with sex, age, and the first four components from the MDS analysis of the IBS matrix as covariates. For nominal $P_{REPLICATION}$ and P_{JOINT} -values, a fixed-effects inverse-variance meta-analysis was performed. Risk allele frequencies and odds ratios were calculated in the joint sample. LD blocks were defined by D′ using Haploview 4.2 based on HapMap CEU population data from HapMap release #27. CI, 95% confidence interval. Genome positions refer to the Human March 2006 (hg18) assembly.

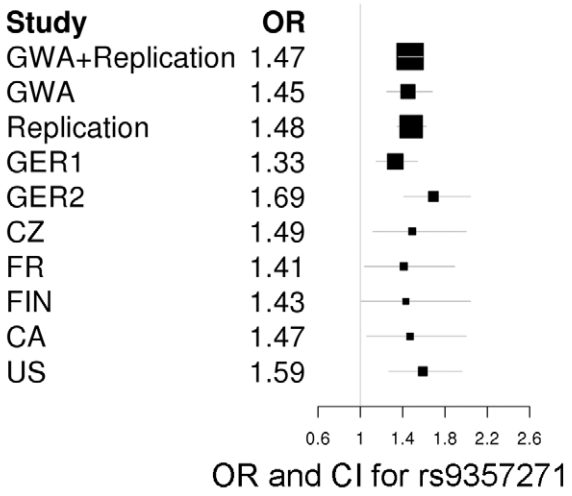
doi:10.1371/journal.pgen.1002171.t001

A



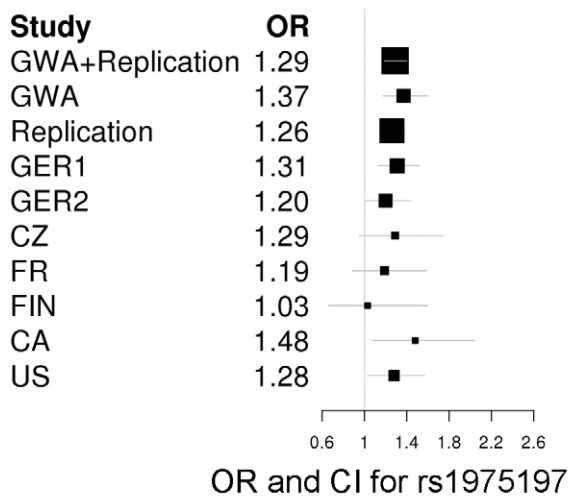
OR and CI for rs2300478

B



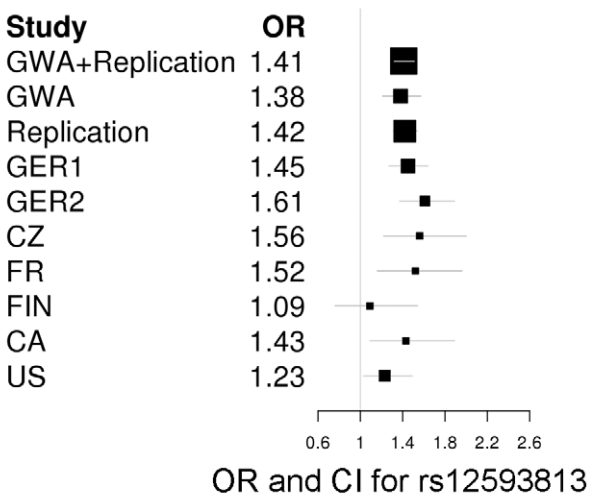
OR and CI for rs9357271

C



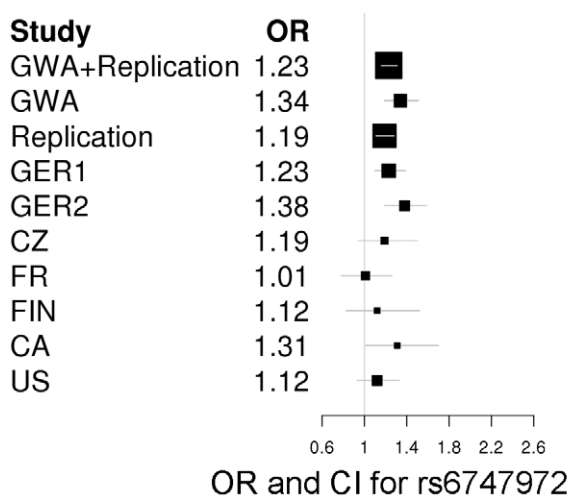
OR and CI for rs1975197

D



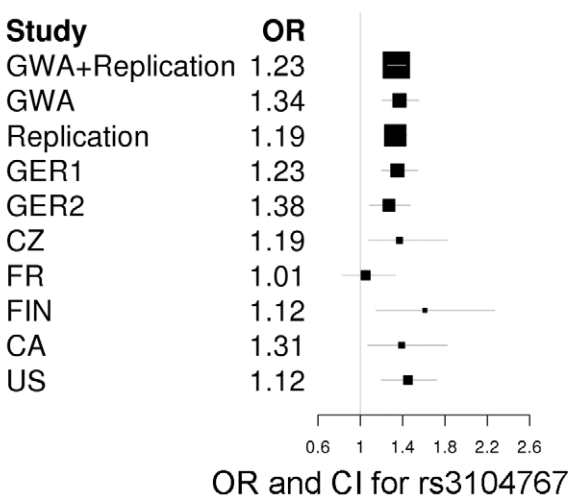
OR and CI for rs12593813

E



OR and CI for rs6747972

F



OR and CI for rs3104767

Figure 2. Forest plots of the RLS risk loci (1 SNP per locus). OR and corresponding confidence interval for the GWA sample, all individual replication samples, the combined replication sample as well as the combined GWA and replication sample are depicted. ORs are indicated by squares with the size of the square corresponding to the sample size for the individual populations. (A) rs2300478 in *MEIS1*; (B) rs9357271 in *BTBD9*; (C) rs1975197 in *PTPRD*; (D) rs12593813 in *MAP2K5/SKOR1*; (E) rs6747972 in intergenic region on chromosome 2; (F) rs3104767 in *TOX3/BC034767*.
doi:10.1371/journal.pgen.1002171.g002

been approved by the institutional review boards of the contributing authors. The primary review board was located in Munich, Bayerische Ärztekammer and Technische Universität München.

RLS patients (GWA and replication phase). A total of 2,944 cases (GWA = 954, replication = 1,990) of European descent were recruited in two cycles via specialized outpatient clinics for RLS. German and Austrian cases for the GWA (GWA) and the replication sample (GER1) were recruited in Munich, Marburg, Kassel, Göttingen, Berlin (Germany, n in GWA = 830, n in GER1 = 1,028), Vienna, and Innsbruck (Austria, n in GWA = 124, n in GER1 = 288). The additional replication samples originated from Prag (Czech Republic (CZ), n = 351), Montpellier (France (FR), n = 182), and Turku (Finland (FIN), n = 141). In all patients, diagnosis was based upon the diagnostic criteria of the International RLS Study Group [1] as assessed in a personal interview conducted by an RLS expert. A positive family history was based on the report of at least one additional family member affected by RLS. We excluded patients with secondary RLS due to uremia, dialysis, or anemia due to iron deficiency. The presence of secondary RLS was determined by clinical interview, physical and neurological examination, blood chemistry, and nerve conduction studies whenever deemed clinically necessary.

In addition, 1,104 participants (GER2) of the "Course of RLS (COR-) Study", a prospective cohort study on the natural course of disease in members of the German RLS patient organizations, were included as an additional replication sample. After providing informed consent, study participants sent their blood for DNA extraction to the Institute of Human Genetics, Munich, Germany. A limited validation of the RLS diagnosis among the majority of members was achieved through a diagnostic questionnaire. Five percent had also received a standardized physical examination and interview in one of the specialized RLS centers in Germany prior to recruitment. To avoid doublets, we checked these subjects against those recruited through other German RLS centers and excluded samples with identical birth date and sex.

556 cases (US) were recruited in the United States at Departments of Neurology at Universities in Baltimore, Miami, Houston, and Palo Alto. Diagnosis of RLS was made as mentioned above.

285 cases (CA) were recruited and diagnosed as above in Montréal, Canada. All subjects were exclusively of French-Canadian ancestry as defined by having four grandparents of French-Canadian origin.

Detailed demographic data of all samples are provided in Table S6.

Control populations (GWA and replication phase). Controls for German and Austrian cases were of European descent and recruited from the KORA S3/F3 and S4 surveys, general population-based controls from southern Germany. KORA procedures and samples have been described [15]. For the GWA phase, we included 1,814 subjects from S3/F3, and, for the replication stage, 1,471 subjects from S4.

For replication of the GER2 sample, we used controls from the Dortmund Health Study (DHS), a population-based survey conducted in the city of Dortmund with the aim of determining the prevalence of chronic diseases and their risk factors in the general population. Sampling for the study was done randomly from the city's population register stratified by five-year age group and gender [16]. 597 subjects selected at random from the Czech blood and bone marrow donor registry served as Czech controls [17]. French controls included 768 parents of multiple sclerosis patients recruited from the French Group of Multiple Sclerosis Genetics Study (REFGENSEP) [18]. Finnish controls comprised 360 participants of the National FINRISK Study, a cross-sectional population survey on coronary risk factors collected every five years. The current study contains individuals recruited in 2002. Detailed description of the FINRISK cohorts can be found at www.nationalbiobanks.fi.

French-Canadian controls were 285 unrelated individuals recruited at the same hospital as the cases.

1,200 participants of the Wisconsin Sleep Cohort (WSC), an ongoing longitudinal study on the causes, consequences, and natural course of disease of sleep disorders, functioned as US controls [19].

None of the controls were phenotyped for RLS. All studies were approved by the institutional review boards in Germany, Austria, Czech Republic, France, Finland, the US, and Canada. Written informed consent was obtained from each participant. Detailed demographic data of all samples are provided in Table S6.

Genotyping

GWA. Genotyping was performed on Affymetrix Genome-Wide Human SNP Arrays 5.0 (cases) and 6.0 (controls) following the manufacturer's protocol. The case sample included 628 cases from previous GWAs [2,4] and 326 new cases. After genotype-calling using the BRLMM-P clustering algorithm [20], a total of 475,976 overlapping SNPs on both Affymetrix arrays were subjected to quality control. We added 655 cases of a different phenotype unrelated to RLS, genotyped on 5.0 arrays, to the analysis and excluded those SNPs which showed a significant difference of allele frequencies in cases (RLS and unrelated phenotype on 5.0) and controls (6.0) (n = 92). Thereby, we filtered out SNPs likely to be false-positive associations. We excluded SNPs with a minor allele frequency (MAF) <5% (n = 88,582), a callrate <98% (n = 65,906) or a significant deviation from Hardy-Weinberg Equilibrium (HWE) in controls (P<0.00001) (n = 20,060). Cluster plots of the GWA genotyping data for the best-associated SNPs in Table 1 are shown in Figure S5. Genotypes of these SNPs are available in Table S7.

Replication. We selected all SNPs with a λ-corrected $P_{nominal}<10^{-4}$ in the GWA for replication. These SNPs clustered in 26 loci (defined as the best associated SNP ±150 kb of flanking sequence). We genotyped a total of three SNPs in each of the 26 regions. These were either further associated neighbouring SNPs with a λ-corrected $P_{nominal}<10^{-3}$ or, in case of singleton SNPs, additional neighbouring SNPs from HapMap with the highest possible r^2 (at least >0.5) with the best-associated SNP. We also genotyped the best-associated SNPs identified in the previous GWAs [2,4].

Genotyping was performed on the MassARRAY system using MALDI-TOF mass spectrometry with the iPLEX Gold chemistry (Sequenom Inc, San Diego, CA, USA). Primers were designed using AssayDesign 3.1.2.2 with iPLEX Gold default parameters. Automated genotype calling was done with SpectroTYPER 3.4. Genotype clustering was visually checked by an experienced evaluator.

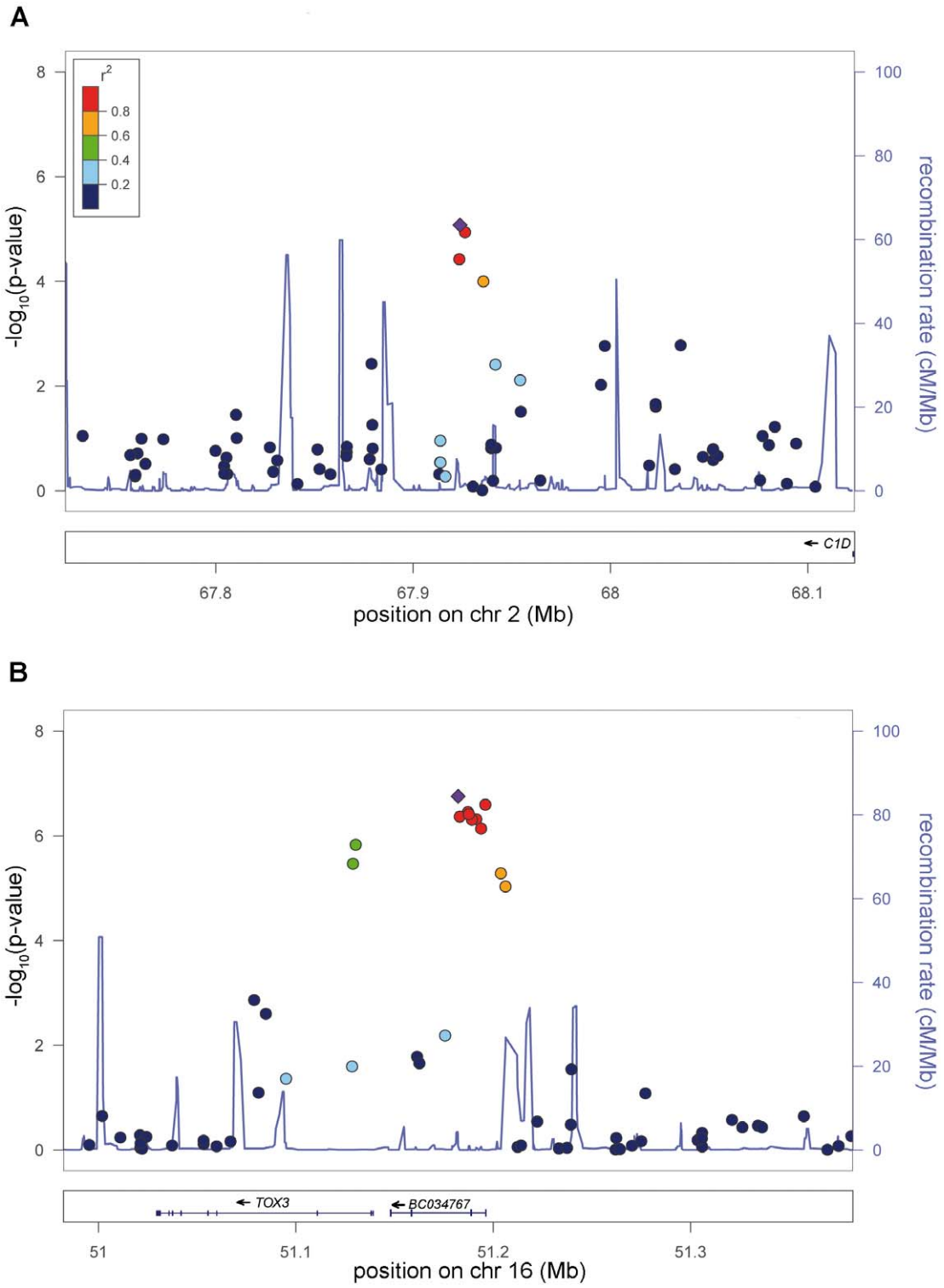


Figure 3. New genome-wide significant RLS loci. a) Risk locus on chromosome 2p14, showing the best-associated SNP rs6747972 and ±200 kb of surrounding sequence. b) Risk locus on chromosome 16p21, showing the best-associated SNP rs3104767 and ±200 kb of surrounding sequence. The left-hand x-axis shows the negative log10 of the nominal λ-corrected P-values of the GWA stage for all SNPs genotyped in the respective region. The right-hand x-axis shows the recombination frequency in cM/Mb. The y-axis shows the genomic position in Mb based on the hg18 assembly. The r^2-based LD between SNPs is colour-coded, ranging from red ($r^2>0.8$) to dark blue ($r^2<0.2$) and uses the best-associated SNP as reference. This SNP is depicted as a violet diamond. Recombination frequency and r^2 values are calculated from the HapMap II (release 22) CEU population. Plots were generated with LocusZoom 1.1 (http://csg.sph.umich.edu/locuszoom/).
doi:10.1371/journal.pgen.1002171.g003

SNPs with a call rate<95%, MAF<5%, and P<0.00001 for deviations from HWE in controls were excluded. DNA samples with a call rate<90% were also excluded.

Population stratification analysis

GWA. To identify and correct for population stratification, we performed an MDS analysis as implemented in PLINK 1.07 (http://pngu.mgh.harvard.edu/~purcell/plink, [21]) on the IBS matrix of our discovery sample. After excluding outliers by plotting the main axes of variation against each other, we performed logistic regression with age, sex, and the values of the MDS components as covariates. Using the Genomic Control approach [22], we obtained an inflation factor λ of 1.11.

Additionally, we performed a variance components analysis using the EMMAX software (http://genetics.cs.ucla.edu/emmax, [23]) and, again, calculated the inflation factor with Genomic Control, now resulting in a λ of 1.025. EMMAX uses a mixed linear model and does not only correct for population stratification but also for hidden relatedness. We, therefore, decided to base correction for population substructure on the EMMAX results.

Replication. Correction for population stratification was performed for the German, Czech, and the Canadian subsamples. The λ-values of 1.1032, 1.2286, and 1.2637 were derived from a previous Genomic Control experiment within the same samples using 176 intergenic or intronic SNPs [4]. Here, we had applied the expanded Genomic Control method GCF developed by Devlin and Roeder [24]. In the meta-analysis of all replication samples, the λ-corrected standard errors were included for the German, Czech, and Canadian samples. For the other replication samples from France, Finland, and the USA, no such data was available and, therefore, no correction factor was included in the analysis.

Statistical analysis

Statistical analysis was performed using PLINK 1.07 (http://pngu.mgh.harvard.edu/~purcell/plink, [21]). In the GWA sample, we applied logistic regression with age, sex, and the first four axes of variation resulting from an MDS analysis as covariates.

P-values were λ-corrected with the λ of 1.025 from the EMMAX analysis. In the individual analysis of the single replication samples, we tested for association using logistic regression and correcting for gender and age as well as for population stratification where possible (see Population Stratification). Each replication sample was Bonferroni-corrected using the number of SNPs which passed quality control for the respective sample.

For the combined analysis of all replication samples, we performed a fixed-effects inverse-variance meta-analysis. Where available, we used λ-corrected standard errors in this analysis. Bonferroni-correction was performed for 74 SNPs, i.e. the number of SNPs which passed quality control in at least one replication sample.

For the joint analysis of the GWA and the replication samples, we also used a fixed-effects inverse-variance meta-analysis and again included λ-corrected values as far as possible. For the conditioned analysis, the SNP to be conditioned on was included as an additional covariate in the logistic regression analysis as implemented in PLINK.

Interaction analysis was performed using the –epistasis option in PLINK. Significance was determined via Bonferroni-correction (i.e. 0.05/28, as 28 SNP combinations were tested for interaction).

Power calculation

Power calculation was performed using the CaTS power calculator [25] using a prevalence set of 0.08 and an additive genetic model (Table S3). The significance level was set at 0.05/74 for replication stage analysis and at 0.05/301,406 for genome-wide significance in the joint analysis of GWA and replication. For the rare variants association study, the significance level was set at 0.05/12.

Mutation screening of *TOX3* and *BC034767*

All coding and non-coding exons including adjacent splice sites of *TOX3* (reference sequence NM_001146188) and *BC034767* (reference sequence IMAGE 5172237) were screened for mutations in 188 German RLS cases.

Mutation screening was performed with high resolution melting curve analysis using the LightScanner technology and standard protocols (IDAHO Technology Inc.). DNAs were analyzed in doublets. Samples with aberrant melting pattern were sequenced using BigDyeTerminator chemistry 3.1 (ABI) on an ABI 3730 sequencer. Sequence analysis was performed with the Staden package [26]. Primers were designed using ExonPrimer (http://ihg.gsf.de) or Primer3plus (www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). All identified variants were then genotyped in 735 RLS cases and 735 controls of the general population (KORA cohort) on the MassARRAY system, as described above.

In addition, fragment analysis of exon 7 of *TOX3* was performed to screen for polymorphic CAG trinucleotide repeats. DNA of 100 controls (50 females, 50 males) was pooled and analyzed on an ABI 3730 sequencer with LIZ-500 (ABI) as a standard. Primers were designed using Primer3plus, the forward Primer contains FAM for detection. Analysis was performed using GeneMapper v3.5.

Expression analyses

Associations between *MEIS1/ETAA1* RNA expression and rs6747972 and between *TOX3/BC034767/CREB-1/CREBBP/CITED1* expression and rs3104767 were assessed using genome-wide SNP data (Affymetrix 6.0 chip) in conjunction with microarray data for human blood samples (n = 323 general population controls from the KORA cohort, Illumina Human WG6 v2 Expression BeadChip) [11]. A linear regression model conditioned on expression and controlling for age and sex was used to test for association.

Prediction of genetic risk

Based on the performance of P-value-threshold selected SNPs in a training and a test sample. As training sample, we used those GWA-cases which had also been genotyped for our previous study [2]. We also included the control samples from this study. As a first quality control step, we carried out an association analysis comparing the Affymetrix 500K genotypes of these GWA-cases to the Affymetrix 5.0 genotypes of the same cases. Significant P-values would indicate systematic differences in the genotyping between the different chips. For further analysis, we only used those 259,302 SNPs with P-values >0.10. We performed a second quality control step in which IDs with a callrate below 98% and SNPs with a callrate below 98%, a MAF lower than 5%, or a P-value for deviation from HWE<0.00001 were removed.

Further, we excluded the four already known risk loci as well as the two newly identified loci and performed LD-pruning to limit the analysis to SNPs not in LD with each other. This was performed using a window-size of 50 SNPs. In each step, this window was shifted 5 SNPs. We used a threshold of 2 for the VIF (variance inflation factor). 76,532 SNPs, 326 cases, and 1,498 controls were included in the final training dataset. We conducted logistic regression with age and sex as covariates. Based on these association results, the sum score of SNPs showing the most

significant effects (i.e. the number of risk alleles over all SNPs) weighted by the ln(OR) of these effects was chosen as predictor variable in the test set, comprising the remaining 583 cases of the GWA sample and 1,526 controls. None of these cases/controls were included in the training-sample, i.e. the test-sample constitutes a completely independent sample. Based on this sum score, we calculated the ROC curve and Nagelkerke's R to measure the explained variance.

Based on a weighted risk allele score. To evaluate the predictive value in our sample, we calculated a weighted sum score of risk alleles in the combined GWA and replication sample. To this end, we used one SNP from each RLS risk region and also included markers from the two newly identified regions on chromosome 16q12 and 2p14 (*MEIS1:* rs2300478, *2p14*: rs6747972, *BTBD9*: rs9296249, *PTPRD*: rs1975197, *MAP2K5:* rs11635424, *TOX3/BC034767*: rs3104767). At each SNP, the number of risk alleles was weighted with the corresponding ln(OR) for this SNP. The corresponding distribution of the score in cases and controls is illustrated in Figure S3. Employing this score for risk prediction resulted in an AUC of 0.651 (Figure S4).

Supporting Information

Figure S1 MDS analysis plot for GWA. Distribution of cases (red) and controls (black) along the two main axes of variation identified in the MDS analysis. The three visible clouds are due to a common 3.8 Mb inversion polymorphism on chromosome 8 (described in: Tian C, Plenge RM, Ransom M, Lee A, Villoslada P, et al. (2008) Analysis and Application of European Genetic Substructure Using 300 K SNP Information. PLoS Genet 4: e4. doi:10.1371/journal.pgen.0040004).
(TIFF)

Figure S2 QQ-plot of GWA results. QQ-plot showing the P-value distribution before (red) and after (blue) correction for population stratification using Genomic Control.
(TIFF)

Figure S3 Weighted risk allele score analysis. Histogram of the weighted risk allele scores for cases and controls. The corresponding OR and CI for each category against the median category is depicted in green. The left y-axis refers to the number of individuals (in %), the right-axis refers to the OR values.
(TIFF)

Figure S4 ROC curve for weighted risk score analysis. Receiver operating characteristic (ROC) curve for the weighted risk allele score approach of risk prediction. The area under the curve (AUC) is 65.1%.
(TIFF)

Figure S5 Cluster plots of GWA genotyping for the six risk loci. For the best-associated SNPs at each risk locus, clusterplots were generated for cases and controls. Intensities of the A and B allele (based on the Affymetrix annotation of the SNPs) are given on the x- and y-axes and the respective genotypes are indicated in blue, green, and orange.
(PDF)

Table S1 GWA results for SNPs with λ-corrected $P_{GWA}<10–4$ and additional SNPs selected for replication. A star (*) indicates SNPs which had been identified in previous RLS GWAs [2–4]. P-values of the GWA phase are given as λ-corrected nominal P-values. Two different methods for λ correction were applied, multi-dimensional-scaling (MDS)-analysis using PLINK and variance components (VC)-analysis using the EMMAX software with the P-values listed in the respective columns "MDS λ-corrected P_{GWA}" and "VC λ-corrected P_{GWA}". The selection of SNPs for replication was based on the MDS λ-corrected P-values. r^2-values based on Hapmap CEU data are given for those SNPs which were selected for replication based on their LD with the best-associated SNP in each region. Genomic position and gene annotation refer to the hg18 genome.
(DOC)

Table S2 Replication stage association results for individual replication samples. P-values are derived from logistic regression and correcting for gender and age as well as for population stratification where possible (see Materials and Methods). Each replication sample was Bonferroni-corrected using the number of SNPs which passed quality control for the respective sample. The OR refers to the minor allele. NA; SNP could not be analysed due to failing quality control in the respective sample.
(DOC)

Table S3 Power analysis for GWA, replication and joint analysis of GWA and replication. Power calculation was performed using the CaTS power calculator [25] using a prevalence set of 0.08 and an additive genetic model. The significance level α was set at 0.05/74 for replication stage analysis and at 0.05/301,406 for genome-wide significance in the joint analysis of GWA and replication.
(DOC)

Table S4 Results of TOX3 and BC034767 mutation screening. * "A" refers to the mutant allele, "B" to the reference allele. Position refers to hg18 genome annotation. Codon numbering refers to the reference sequence NM_001146188. Data of the 1000 genomes project was obtained from the November 2010 release via the 1000 genomes browser (http://browser.1000genomes.org/index.html).
(DOC)

Table S5 Prediction of genetic risk; training- and test-set approach. Inclusion threshold P-values were derived from a logistic regression with age and sex as covariates in the training sample. # SNPs indicates the number of SNPs passing the inclusion threshold. Based on these association results, the sum score of SNPs showing the most significant effects (i.e. the number of risk alleles over all SNPs) weighted by the ln(OR) of these effects was chosen as predictor variable in the test set. Based on this sum score, an AUC and Nagelkerke's R were calculated.
(DOC)

Table S6 Demographic data of GWA and replication samples. Mean age, mean age of onset and respective standard deviations and ranges are given in years. N: number of individuals; SD: standard deviation; AAO: age of onset. GWA: Genome-wide association study; CZ: Czechia; FR: France; FIN: Finland; CA: Canada; US: United States. - indicates that this information is not applicable for the respective sample.
(DOC)

Table S7 Genotype data of GWA samples. Genotypes of the GWA samples are given for the eight best-associated SNPs (see Table 1). SNP alleles are ACGT-coded. Phenotype information includes gender (1 = male, 2 = female) and disease status (1 = unaffected, 2 = affected).
(XLS)

Acknowledgments

We are grateful to all patients who participated in this study. We thank Jelena Golic, Regina Feldmann, Sibylle Frischholz, Susanne Lindhof,

Katja Junghans, Milena Radivojkov-Blagojevic, and Bianca Schmick for excellent technical assistance.

Author Contributions

Study design: J Winkelmann, B Müller-Myhsok, T Meitinger. Recruitment and biobanking of German/Austrian RLS cases: J Winkelmann, C Trenkwalder, B Högl, K Berger, N Gross, K Stiasny-Kolster, W Oertel, CG Bachmann, W Paulus, I Fietze, V Gschliesser, B Frauscher, T Falkenstetter, W Poewe, D Spieler, M Kaffe, A Zimprich, T Meitinger. Recruitment and biobanking of KORA controls: C Gieger, T Illig, H-E Wichmann. Recruitment and biobanking of Canadian RLS cases and controls: L Xiong, J Montplaisir, GA Rouleau. Czech RLS cases and controls: Jávrová, D Kemlink, K Sonka, S Nevsimalova, P Vodicka. US cases and controls: S-C Lin, Z Wszolek, C Vilariño-Güell, MJ Farrer, RP Allen, CJ Earley, WG Ondo, W-D Le, P Peppard, J Faraco, E Mignot. Finnish cases and controls: O Polo, J Kettunen, M Perola, K Silander. French cases and controls: Y Dauvilliers, I Cournu-Rebeix, M Francavilla, C Fontenille, B Fontaine. Affymetrix genotyping: B Schormair, P Lichtner. Sequenom genotyping: B Schormair, F Knauf, EC Schulte, P Lichtner. Sequencing and Fragment analysis: F Knauf. Expression analysis: EC Schulte, H Prokisch. Supervision of all markers typed: J Winkelmann, P Lichtner. Statistical analysis: D Czamara, B Müller-Myhsok. Clustering of Affymetrix genotypes: D Czamara, B Müller-Myhsok. Wrote the manuscript: J Winkelmann, D Czamara, B Schormair, B Müller-Myhsok, T Meitinger.

References

1. Allen RP, Picchietti D, Hening WA, Trenkwalder C, Walters AS, et al. (2003) Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. Sleep Med 4: 101–119.
2. Winkelmann J, Schormair B, Lichtner P, Ripke S, Xiong L, et al. (2007) Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. Nat Genet 39: 1000–1006.
3. Stefansson H, Rye DB, Hicks A, Petursson H, Ingason A, et al. (2007) A genetic risk factor for periodic limb movements in sleep. N Engl J Med 357: 639–647.
4. Schormair B, Kemlink D, Roeske D, Eckstein G, Xiong L, et al. (2008) PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. Nat Genet 40: 946–948.
5. Engstrom PG, Fredman D, Lenhard B (2008) Ancora: a web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes. Genome Biol 9: R34.
6. Kikuta H, Laplante M, Navratilova P, Komisarczuk AZ, Engstrom PG, et al. (2007) Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. Genome Res 17: 545–555.
7. Yuan SH, Qiu Z, Ghosh A (2009) TOX3 regulates calcium-dependent transcription in neurons. Proc Natl Acad Sci U S A 106: 2909–2914.
8. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, et al. (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447: 1087–1093.
9. Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. Nat Rev Genet 10: 155–159.
10. Dittmer S, Kovacs Z, Yuan SH, Siszler G, Kögl M, et al. (2011) TOX3 is a neuronal survival factor that induces transcription depending on the presence of CITED1 or phosphorylated CREB in the transcriptionally active complex. J Cell Sci 124: 252–60.
11. Meisinger C, Prokisch H, Gieger C, Soranzo N, Mehta D, et al. (2009) A genome-wide association study identifies three loci associated with mean platelet volume. Am J Hum Genet 84(1): 66–71.
12. Wacholder S, Hartge P, Prentice R, Garcia-Closas M, Feigelson HS, et al. (2010) Performance of common genetic variants in breast-cancer risk models. N Engl J Med 362: 986–993.
13. Lango H, Palmer CN, Morris AD, Zeggini E, Hattersley AT, et al. (2008) Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. Diabetes 57: 3129–3135.
14. van Hoek M, Dehghan A, Witteman JC, van Duijn CM, Uitterlinden AG, et al. (2008) Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. Diabetes 57: 3122–3128.
15. Wichmann HE, Gieger C, Illig T (2005) KORA-gen–resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 67 Suppl 1: S26–30.
16. Happe S, Vennemann M, Evers S, Berger K (2008) Treatment wish of individuals with known and unknown restless legs syndrome in the community. J Neurol 255: 1365–1371.
17. Pardini B, Naccarati A, Polakova V, Smerhovsky Z, Hlavata I, et al. (2009) NBN 657del5 heterozygous mutations and colorectal cancer risk in the Czech Republic. Mutat Res 666: 64–67.
18. Cournu-Rebeix I, Genin E, Leray E, Babron MC, Cohen J, et al. (2008) HLA-DRB1*15 allele influences the later course of relapsing remitting multiple sclerosis. Genes Immun 9: 570–574.
19. Young T, Palta M, Dempsey J, Peppard PE, Nieto FJ, et al. (2009) Burden of sleep apnea: rationale, design, and major findings of the Wisconsin Sleep Cohort study. Wmj 108: 246–249.
20. Affymetrix Inc. (2007) BRLMM-P: a Genotype Calling Method for the SNP 5.0 Array. http://www.affymetrix.com/support/technical/whitepapers.affx. Accessed 03. December 2010.
21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.
22. Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55: 997–1004.
23. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, et al. (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42: 348–54.
24. Devlin B, Bacanu SA, Roeder K (2004) Genomic controls to the extreme. Nat Genet 36: 1129–1130.
25. Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 38: 209–213.
26. Staden R, Beal KF, Bonfield JK (2000) The Staden package, 1998. Methods Mol Biol 132: 115–130.

ARTICLES
https://doi.org/10.1038/s41588-018-0090-3
nature genetics

Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression

Major depressive disorder (MDD) is a common illness accompanied by considerable morbidity, mortality, costs, and heightened risk of suicide. We conducted a genome-wide association meta-analysis based in 135,458 cases and 344,901 controls and identified 44 independent and significant loci. The genetic findings were associated with clinical features of major depression and implicated brain regions exhibiting anatomical differences in cases. Targets of antidepressant medications and genes involved in gene splicing were enriched for smaller association signal. We found important relationships of genetic risk for major depression with educational attainment, body mass, and schizophrenia: lower educational attainment and higher body mass were putatively causal, whereas major depression and schizophrenia reflected a partly shared biological etiology. All humans carry lesser or greater numbers of genetic risk factors for major depression. These findings help refine the basis of major depression and imply that a continuous measure of risk underlies the clinical phenotype.

MDD is a notably complex and common illness[1]. It is often chronic or recurrent and is thus accompanied by considerable morbidity, disability, excess mortality, substantial costs, and heightened risk of suicide[2–8]. Twin studies attribute approximately 40% of the variation in liability to MDD to additive genetic effects (phenotype heritability, h^2)[9], and h^2 may be greater for recurrent, early-onset, and postpartum MDD[10,11]. Genome-wide association studies (GWAS) of MDD have had notable difficulties in identifying individual associated loci[12]. For example, there were no significant findings in the initial Psychiatric Genomics Consortium (PGC) MDD mega-analysis (9,240 cases)[13] or in the CHARGE meta-analysis of depressive symptoms ($n = 34{,}549$)[14]. More recent studies have proven modestly successful. A study of Han Chinese women (5,303 recurrent MDD cases) identified significant loci[15], a meta-analysis of depressive symptoms (161,460 individuals) identified 2 loci[16], and an analysis of self-reported major depression identified 15 loci (75,607 cases).

There are many reasons why identifying causal loci for MDD has proven difficult[12]. MDD is probably influenced by many genetic loci, each with small effects[17], as are most common diseases[18], including psychiatric disorders[19,20]. Estimates of the proportion of variance attributable to genome-wide SNPs (SNP heritability, h^2_{SNP}) indicate that around one-quarter of the h^2 for MDD is due to common genetic variants[21,22] and demonstrate that a genetic signal is detectable in genome-wide association data, implying that larger sample sizes are needed to detect specific loci given their effect sizes. Such a strategy has been proven in studies of schizophrenia, the flagship adult psychiatric disorder in genomics research. We thus accumulated clinical, population, and volunteer cohorts[23]. This pragmatic approach takes the view that sample size can overcome heterogeneity to identify risk alleles that are robustly associated with major depression. Potential concerns about combining carefully curated research cohorts with volunteer cohorts were considered; multiple lines of evidence suggest that the results are likely to be applicable to clinical MDD. As discussed below, our analyses have neurobiological, clinical, and therapeutic relevance for major depression.

Results

Cohort analyses: phenotype validation. We identified seven cohorts that used a range of methods to ascertain cases with major depression (Table 1 and Supplementary Tables 1–3). The methods used by these cohorts were thoroughly reviewed, drawing on the breadth of expertise in the PGC, and we assessed the comparability of the cohorts using genomic data. We use 'MDD' to refer to directly evaluated subjects meeting standard criteria for major depressive disorder and use 'major depression' where case status was determined using alternative methods as well as to the phenotype from the full meta-analysis.

We evaluated the comparability of the seven cohorts by estimating the common variant genetic correlations (r_g) between them. These analyses supported the comparability of the seven cohorts (Supplementary Table 3), as the weighted mean r_g was 0.76 (s.e. = 0.03). The high genetic correlations between the 23andMe and other cohorts are notable. While there was no statistical evidence of heterogeneity in the r_g estimates for pairs of cohorts ($P = 0.13$), the estimate was statistically different from 1, which may reflect etiological heterogeneity. This estimate can be benchmarked against the slightly larger weighted mean r_g between schizophrenia cohorts of 0.84 (s.e. = 0.05)[21].

Given the positive evidence of the genetic comparability of these cohorts, we completed a genome-wide association meta-analysis of 9.6 million imputed SNPs in 135,458 MDD and major depression cases and 344,901 controls (Fig. 1). There was no evidence of residual population stratification[24] (LD score regression intercept = 1.018, s.e. = 0.009). We estimated h^2_{SNP} to be 8.7% (s.e. = 0.004, liability scale, assuming lifetime risk of 0.15; Supplementary Fig. 1 and Supplementary Table 3b), and note that this is about one-quarter of the h^2 estimated from twin or family studies[9]. This fraction is somewhat lower than that of other complex traits[18] and is plausibly due to etiological heterogeneity (reflecting the mean $r_g < 1$ between cohorts).

To evaluate the impact of combining major depression cohorts that used different ascertainment methods, we undertook a series of genetic risk score (GRS) prediction analyses to demonstrate the validity of our genome-wide association results for clinical MDD (Fig. 2). Notably, the variance explained in out-of-sample prediction increased with the size of the genome-wide association discovery cohort (Fig. 2a), with the GRS from the full discovery

A full list of authors and affiliations appears at the end of the paper.

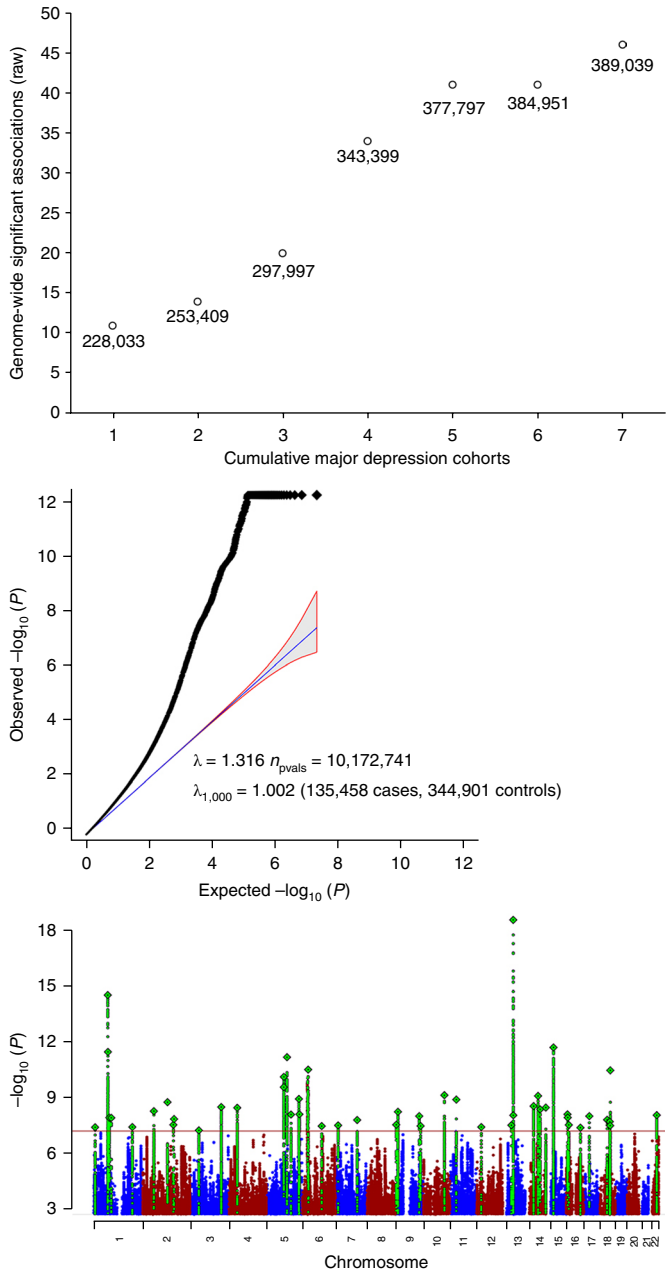


Fig. 1 | Results of genome-wide association meta-analysis of seven cohorts for major depression. a, Relationship between adding cohorts and the number of genome-wide significant genomic regions (before the vetting used to define the final 44 regions). Beginning with the largest cohort (cohort 1 on the *x* axis), we added the next largest cohort (cohort 2) until all cohorts were included (7 cohorts). The number next to each point is the total effective sample size, equivalent to the sample size where the numbers of cases and controls are equal. **b**, Association test quantile–quantile plot showing a marked departure from the null model of no associations (*y* axis truncated at 10^{-12}). **c**, Manhattan plot for association tests from meta-analysis of 135,458 major depression cases and 344,901 controls, with the *x* axis showing genomic position (chromosomes 1–22 plus the X chromosome) and the *y* axis showing statistical significance as $-\log_{10}(P)$ *z* statistics; the threshold for significance accounting for multiple testing is shown by the red horizontal line ($P = 5 \times 10^{-8}$).

sample meta-analysis explaining 1.9% of variance in liability (Fig. 2a, Supplementary Fig. 2, and Supplementary Table 4). For any randomly selected case and control, GRS ranked cases higher than controls with probability of 0.57 and the odds ratio of MDD for those in the tenth versus first GRS decile (OR = 10) was 2.4 (Fig. 2b and Supplementary Table 4). GRS analyses in other disorders (for example, schizophrenia[25]) have shown that the mean GRS increases with clinical severity in cases. We found a significantly higher major depression GRS in those with more severe MDD, as measured in different ways (Fig. 2c). Last, because around one-half of the major depression cases were identified by self-report (i.e., diagnosis or treatment for clinical depression by a medical professional), we further evaluated the comparability of the 23andMe cohort with the other cohorts (full meta-analysis excluding 23andMe, 'FMex23andMe') as detailed in Fig. 2c, Supplementary Table 5, and the Supplementary Note. Taken together, we interpret these results as supporting this meta-analysis of GWAS for these seven cohorts.

Implications for the biology of major depression. Our meta-analysis of seven MDD and major depression cohorts identified 44 independent loci that were statistically significant ($P < 5 \times 10^{-8}$), statistically independent of any other signal[26], and supported by multiple SNPs. This number supports our prediction that genome-wide association discovery in major depression would require about five times more cases than for schizophrenia (lifetime risk ~1% and h^2 ~0.8) to achieve approximately similar power[27]. Of these 44 loci, 30 are new and 14 were significant in a prior study of MDD or depressive symptoms. The overlap of our findings with prior reports was as follows: 1 of 1 with CHARGE depressive symptom[14], 1 of 2 overlap with SSGAC depressive symptom[16], and 12 of 15 overlap with Hyde et al.[28]. There are few trans-ancestry comparisons for major depression, so we contrasted these European results with the Han Chinese CONVERGE study[15] (Supplementary Note). The loci identified in CONVERGE are uncommon in Europeans (rs12415800, 0.45 versus 0.02; rs35936514, 0.28 versus 0.06) and were not significant in our analysis.

Table 2 lists genes in or near the lead SNP in each region, regional plots are in Supplementary Data 1, and Supplementary Tables 6 and 7 provide summaries of available information about the biological functions of the genes in each region. In the Supplementary Note, we review four key genes in more detail: *OLFM4* and *NEGR1* (notable for reported associations with obesity and body mass index[29–34]), *RBFOX1* (notable for independent associations at both the 5′ and 3′ ends, a splicing regulator[35,36], with a functional role that may be consistent with chronic hypothalamic–pituitary–adrenal axis hyperactivation reported in MDD[37]), and *LRFN5* (notable for its role in presynaptic differentiation[38,39] and neuroinflammation[40]).

Gene-wise analyses identified 153 significant genes after controlling for multiple comparisons (Supplementary Table 7). Many of these genes were in the extended major histocompatibility complex (MHC) region (45 of 153), and their interpretation is complicated by high linkage disequilibrium (LD) and gene density. In addition to the genes discussed above, other notable and significant genes outside of the MHC region include multiple potentially 'druggable' targets that suggest connections of the pathophysiology of MDD to neuronal calcium signaling (*CACNA1E* and *CACNA2D1*), dopaminergic neurotransmission (*DRD2*, a principal target of antipsychotics), glutamate neurotransmission (*GRIK5* and *GRM5*), and presynaptic vesicle trafficking (*PCLO*).

Finally, comparison of the major depression loci with 108 loci for schizophrenia[19] identified 6 shared loci. Many SNPs in the extended MHC region are strongly associated with schizophrenia, but implication of the MHC region is new for major depression. Another example is *TCF4* (transcription factor 4), which is strongly associated with schizophrenia but was not previously associated with MDD. TCF4 is essential for normal brain development, and rare mutations in *TCF4* cause Pitt–Hopkins syndrome, which includes autistic features[41]. The GRS calculated from the schizophrenia

Table 1 | Brief description of the seven MDD or major depression cohorts used in the meta-analysis

Sample	Country	Case ascertainment	Cases	Controls
PGC29[13,a]	Various	Structured diagnostic interviews[b]	16,823	25,632
deCODE[13]	Iceland	National inpatient electronic records	1,980	9,536
GenScotland[78,79]	UK	Structured diagnostic interview	997	6,358
GERA[80]	USA	Kaiser Permanente Northern California Healthcare electronic medical records (1995–2013)	7,162	38,307
iPSYCH[81]	Denmark	National inpatient electronic records	18,629	17,841
UK Biobank[82] (pilot data release)	UK	From self-reported MDD symptoms or treatment or electronic records[69]	14,260	15,480
23andMe[28] (discovery sample)[c]	USA	Self-reported diagnosis or treatment for clinical depression by a medical professional	75,607	231,747
Total			135,458	344,901

[a]Nineteen samples in addition to the ten samples published in the first PGC-MDD paper[13]. [b]One sample used natural language processing of electronic medical records followed by expert diagnostic review. [c]In Hyde et al.[28], SNPs in 15 genomic regions met genome-wide significance in the combined discovery and replication samples and 11 regions achieved genome-wide significance in the discovery sample made available to the research community and used here. More details are provided in Supplementary Tables 1–3.

genome-wide association results explained 0.8% of the variance in liability of MDD (Fig. 2c).

Implications from integration of functional genomic data. Results from 'omic' studies of functional features of cells and tissues are necessary to understand the biological implications of the results of GWAS for complex disorders[42]. To further elucidate the biological relevance of the major depression findings, we integrated the results with functional genomic data. First, using enrichment analyses, we compared the major depression GWAS findings to bulk tissue mRNA-seq from Genotype-Tissue Expression (GTEx) data[43]. Only brain samples showed significant enrichment (Fig. 3a), and the three tissues with the most significant enrichments were all cortical. Prefrontal cortex and anterior cingulate cortex are important for higher-level executive functions and emotional regulation, which are often impaired in MDD. Both of these regions were implicated in a large meta-analysis of brain magnetic resonance imaging (MRI) findings in adult MDD cases[44]. Second, given the predominance of neurons in the cortex, we confirmed that the major depression genetic findings connect to genes expressed in neurons but not oligodendrocytes or astrocytes (Fig. 3b)[45]. Given the different methods used by the seven MDD/major depression cohorts in this study, demonstration of enrichment of association signals in the brain regions expected to be most relevant to MDD provides independent support for the validity of our approach.

Third, we used partitioned LD score regression[46] to evaluate the enrichment of the major depression genome-wide association findings in over 50 functional genomic annotations (Fig. 3c and Supplementary Table 8). The major finding was the significant enrichment of h^2_{SNP} in genomic regions conserved across 29 Eutherian mammals[47] (20.9-fold enrichment, $P = 1.4 \times 10^{-15}$). This annotation was also the most enriched for schizophrenia[46]. We could not evaluate regions conserved in primates or human 'accelerated' regions, as there were too few for confident evaluation[47]. The other enrichments implied regulatory activity and included open chromatin in human brain and an epigenetic mark of active enhancers (H3K4me1). Notably, exonic regions did not show enrichment, suggesting that, as with schizophrenia[17], genetic variants that change exonic sequences may not have a large role in major depression. We found no evidence that Neanderthal introgressed regions were enriched for major depression genome-wide association findings[48].

Fourth, we applied methods to integrate genome-wide association SNP results with those from gene expression and methylation quantitative trait locus (eQTL and meQTL) studies. SMR[49] analysis identified 13 major depression–associated SNPs with strong evidence that they control local gene expression in one or more tissues and 9 with strong evidence that they control local DNA methylation (Supplementary Table 9 and Supplementary Data 2). A transcriptome-wide association study[50] applied to data from the dorsolateral prefrontal cortex[51] identified 17 genes where major depression–associated SNPs influenced gene expression (Supplementary Table 10). These genes included *OLFM4* (discussed above).

Fifth, we added additional data types to attempt to improve understanding of individual loci. For the intergenic associations, we evaluated total-stranded RNA-seq data from human brain and found no evidence for unannotated transcripts in these regions. A particularly important data type is assessment of DNA–DNA interactions, which can localize a genome-wide association finding to a specific gene that may be nearby or hundreds of kilobases away[52–54]. We integrated the major depression results with 'easy Hi-C' data from brain cortical samples (3 adult, 3 fetal, >1 billion reads each). These data clarified three associations. The statistically independent associations in *NEGR1* (rs1432639, $P = 4.6 \times 10^{-15}$) and over 200 kb away (rs12129573, $P = 4.0 \times 10^{-12}$) both implicate *NEGR1* (Supplementary Fig. 3a), the former likely due to the presence of a reportedly functional copy number polymorphism (see Supplementary Note) and the presence of intergenic loops. The latter association has evidence of DNA looping interactions with *NEGR1*. The association in *SOX5* (rs4074723) and the two statistically independent associations in *RBFOX1* (rs8063603 and rs7198928, $P = 6.9 \times 10^{-9}$ and 1.0×10^{-8}) had only intragenic associations, suggesting that the genetic variation in the regions of the major depression associations act locally and can be assigned to these genes. In contrast, the association in *RERE* (rs159963, $P = 3.2 \times 10^{-8}$) could not be assigned to *RERE* as it may contain super-enhancer elements given its many DNA–DNA interactions with many nearby genes (Supplementary Fig. 3b).

Implications based on the roles of sets of genes. A parsimonious explanation for the presence of many significant associations for a complex trait is that the different associations are part of a higher-order grouping of genes[55]. These could be a biological pathway or a collection of genes with a functional connection. Multiple methods allow evaluation of the connection of major depression genome-wide association results to sets of genes grouped by empirical or predicted function (pathway or gene set analysis).

Full pathway analyses are in Supplementary Table 11, and 19 pathways with false discovery rate q values <0.05 are summarized in Fig. 4. The major groupings of significant pathways were as follows: RBFOX1, RBFOX2, RBFOX3, or CELF4 regulatory networks; genes whose mRNAs are bound by FMRP; synaptic genes; genes

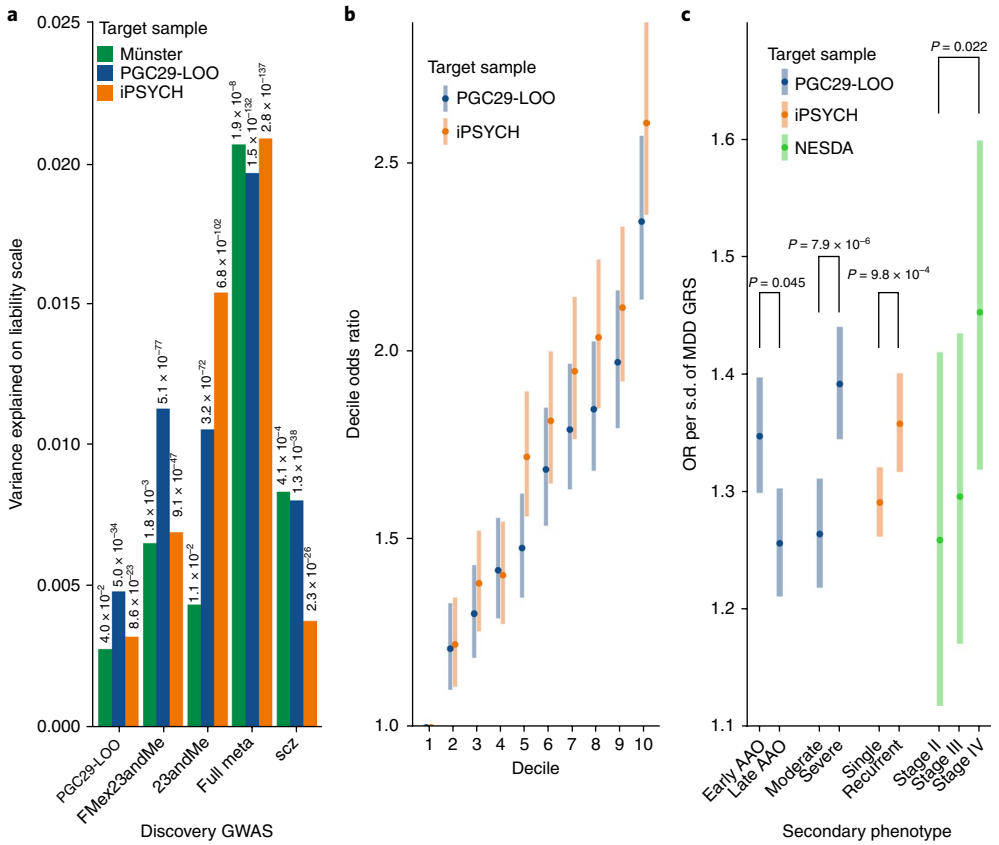


Fig. 2 | Genetic risk score prediction analyses into PGC29 MDD target samples. a, Variance explained (liability scale) based on different discovery samples for three target samples: PGC29 (16,823 cases, 25,632 controls), iPSYCH (a nationally representative sample of 18,629 cases and 17,841 controls), and a clinical cohort from Münster not included in the genome-wide association analysis (845 MDD inpatient cases, 834 controls). For PGC29-LOO, the target sample was one of the PGC29 samples with the discovery sample being the remaining 28 PGC29 samples, hence representing leave-one-out (LOO) analysis. **b**, Odds ratios of major depression per GRS decile relative to the first decile for the iPSYCH and PGC29 target samples. **c**, Odds ratios of major depression in GRS s.d.: PGC29-LOO, 3,950 early-onset versus 3,950 late-onset cases; PGC29-LOO, 4,958 severe versus 3,976 moderate cases defined by count of endorsed MDD symptom criteria; iPSYCH, 5,574 cases of recurrent MDD versus 12,968 single-episode cases; and NESDA from PGC29, severity defined as chronic/unremitting MDD, 610 'stage IV' cases versus 499 'stage II' or 332 first-episode MDD cases[77] . Error bars represent 95% confidence intervals. Logistic regression association test *P* values are shown in the target sample for the GRS generated from SNPs with $P < 0.05$ in the discovery sample. FMex23andMe, full meta-analysis excluding 23andMe; scz, schizophrenia[19].

involved in neuronal morphogenesis; genes involved in neuron projection; genes associated with schizophrenia (at $P < 10^{-4}$)[19]; genes involved in central nervous system (CNS) neuron differentiation; genes encoding voltage-gated calcium channels; genes involved in cytokine and immune response; and genes known to bind to the retinoid X receptor. Several of these pathways are implicated by GWAS of schizophrenia and by rare exonic variation of schizophrenia and autism[56,57] and immediately suggest shared biological mechanisms across these disorders.

A key issue for common variant GWAS is their relevance for pharmacotherapy. We conducted gene set analysis that compared the major depression genome-wide association results to targets of antidepressant medications defined by pharmacological studies[58] and found that 42 sets of genes encoding proteins bound by antidepressant medications were highly enriched for smaller major depression association *P* values than expected by chance (42 drugs, rank enrichment test $P = 8.5 \times 10^{-10}$). This finding connects our major depression genomic findings to MDD therapeutics and suggests the salience of these results for new lead compound discovery for MDD[59].

Implications based on relationships with other traits. Prior epidemiological studies associated MDD with many other diseases and traits. Because of limitations inherent to observational studies, understanding whether a phenotypic correlation is potentially causal or if it results from reverse causation or confounding is generally difficult. Genetic studies now offer complementary strategies to assess whether a phenotypic association between MDD and a risk factor or a comorbidity is mirrored by a nonzero r_g (common variant genetic correlation) and, for some of these, evaluate the potential causality of the association given that exposure to genetic risk factors begins at conception.

We used LD score regression to estimate the r_g of major depression with 221 psychiatric disorders, medical diseases, and human traits[22,60]. Supplementary Table 12 contains the full results, and Table 3 shows the r_g values with false discovery rates <0.01. First, the r_g values were very high between our major depression genome-wide association results and those from two studies of current depressive symptoms. Both correlations were close to 1 (the samples in one report overlapped partially with this meta-analysis[16], but the samples from the other did not[14]).

Second, we found significant positive genetic correlations between major depression and every psychiatric disorder assessed along with smoking initiation. This is a comprehensive and well-powered evaluation of the relationship of MDD with other psychiatric disorders, and these results indicate that the common genetic

Table 2 | 44 significantly associated genomic regions in meta-analysis of 135,458 major depression cases and 344,901 controls

Chr.	Region (Mb)	SNP	Location (bp)	*P*	A1/A2	OR (A1)	s.e. (log(OR))	Freq.	Prev.	Gene context
1	8.390–8.895	rs159963	8,504,421	3.2×10^{-8}	A/C	0.97	0.0049	0.56	H,S	[*RERE*]; *SLC45A1*, 100,194
1	72.511–73.059	rs1432639	72,813,218	4.6×10^{-15}	A/C	1.04	0.0050	0.63	H	*NEGR1*, –64,941
1	73.275–74.077	rs12129573	73,768,366	4.0×10^{-12}	A/C	1.04	0.0050	0.37	S	*LINC01360*, –3,486
1	80.785–80.980	rs2389016	80,799,329	1.0×10^{-8}	T/C	1.03	0.0053	0.28	H	
1	90.671–90.966	rs4261101	90,796,053	1.0×10^{-8}	A/G	0.97	0.0050	0.37		
1	197.343–197.864	rs9427672	197,754,741	3.1×10^{-8}	A/G	0.97	0.0058	0.24		*DENND1B*, –10,118
2	57.765–58.485	rs11682175	57,987,593	4.7×10^{-9}	T/C	0.97	0.0048	0.52	H,S	*VRK2*, –147,192
2	156.978–157.464	rs1226412	157,111,313	2.4×10^{-8}	T/C	1.03	0.0059	0.79		[*LINC01876*]; *NR4A2*, 69,630; *GPD2*, –180,651
3	44.222–44.997	chr3_44287760_I	44,287,760	4.6×10^{-8}	I/D	1.03	0.0051	0.34	T	[*TOPAZ1*]; *TCAIM*, –91,850; *ZNF445*, 193,501
3	157.616–158.354	rs7430565	158,107,180	2.9×10^{-9}	A/G	0.97	0.0048	0.58	H	[*RSRC1*]; *LOC100996447*, 155,828; *MLF1*, –181,772
4	41.880–42.189	rs34215985	42,047,778	3.1×10^{-9}	C/G	0.96	0.0063	0.24		[*SLC30A9*]; *LINC00682*, –163,150; *DCAF4L1*, 59,294
5	87.443–88.244	chr5_87992715_I	87,992,715	7.9×10^{-11}	I/D	0.97	0.0050	0.58	H	*LINC00461*, –12,095; *MEF2C*, 21,342
5	103.672–104.092	chr5_103942055_D	103,942,055	7.5×10^{-12}	I/D	1.03	0.0048	0.48	C	
5	124.204–124.328	rs116755193	124,251,883	7.0×10^{-9}	T/C	0.97	0.0050	0.38		*LOC101927421*, –120,640
5	164.440–164.789	rs11135349	164,523,472	1.1×10^{-9}	A/C	0.97	0.0048	0.48	H	
5	166.977–167.056	rs4869056	166,992,078	6.8×10^{-9}	A/G	0.97	0.0050	0.63		[*TENM2*]
6	27.738–32.848	rs115507122	30,737,591	3.3×10^{-11}	C/G	0.96	0.0063	0.18	S	Extended MHC
6	99.335–99.662	rs9402472	99,566,521	2.8×10^{-8}	A/G	1.03	0.0059	0.24		*FBXL4*, –170,672; *C6orf168*, 154,271
7	12.154–12.381	rs10950398	12,264,871	2.6×10^{-8}	A/G	1.03	0.0049	0.41		[*TMEM106B*]; *VWDE*, 105,637
7	108.925–109.230	rs12666117	109,105,611	1.4×10^{-8}	A/G	1.03	0.0048	0.47		
9	2.919–3.009	rs1354115	2,983,774	2.4×10^{-8}	A/C	1.03	0.0049	0.62	H	*PUM3*, –139,644; *LINC01231*, –197,814
9	11.067–11.847	rs10959913	11,544,964	5.1×10^{-9}	T/G	1.03	0.0057	0.76		
9	119.675–119.767	rs7856424	119,733,595	8.5×10^{-9}	T/C	0.97	0.0053	0.29		[*ASTN2*]
9	126.292–126.735	rs7029033	126,682,068	2.7×10^{-8}	T/C	1.05	0.0093	0.07		[*DENND1A*]; *LHX2*, –91,820
10	106.397–106.904	rs61867293	106,563,924	7.0×10^{-10}	T/C	0.96	0.0061	0.20	H	[*SORCS3*]
11	31.121–31.859	rs1806153	31,850,105	1.2×10^{-9}	T/G	1.04	0.0059	0.22		[*DKFZp686K1684*]; [*PAUPAR*]; *ELP4*, 44,032; *PAX6*, –10,596
12	23.924–24.052	rs4074723	23,947,737	3.1×10^{-8}	A/C	0.97	0.0049	0.41		[*SOX5*]
13	44.237–44.545	rs4143229	44,327,799	2.5×10^{-8}	A/C	0.95	0.0091	0.92		[*ENOX1*]; *LACC1*, –125,620; *CCDC122*, 82,689
13	53.605–54.057	rs12552	53,625,781	6.1×10^{-19}	A/G	1.04	0.0048	0.44	H	[*OLFM4*]; *LINC01065*, 80,099
14	41.941–42.320	rs4904738	42,179,732	2.6×10^{-9}	T/C	0.97	0.0049	0.57		[*LRFN5*]
14	64.613–64.878	rs915057	64,686,207	7.6×10^{-10}	A/G	0.97	0.0049	0.42		[*SYNE2*]; *MIR548H1*, –124,364; *ESR2*, 7,222
14	75.063–75.398	chr14_75356855_I	75,356,855	3.8×10^{-9}	D/I	1.03	0.0049	0.49		[*DLST*]; *PROX2*, –26,318; *RPS6KL1*, 13,801
14	103.828–104.174	rs10149470	104,017,953	3.1×10^{-9}	A/G	0.97	0.0049	0.49	S	*BAG5*, 4,927; *APOPT1*, –11,340
15	37.562–37.929	rs8025231	37,648,402	2.4×10^{-12}	A/C	0.97	0.0048	0.57	H	

Continued

Table 2 | 44 significantly associated genomic regions in meta-analysis of 135,458 major depression cases and 344,901 controls (continued)

Chr.	Region (Mb)	SNP	Location (bp)	*P*	A1/A2	OR (A1)	s.e. (log(OR))	Freq.	Prev.	Gene context
16	6.288–6.347	rs8063603	6,310,645	6.9×10^{-9}	A/G	0.97	0.0053	0.65		[*RBFOX1*]
16	7.642–7.676	rs7198928	7,666,402	1.0×10^{-8}	T/C	1.03	0.0050	0.62		[*RBFOX1*]
16	13.022–13.119	rs7200826	13,066,833	2.4×10^{-8}	T/C	1.03	0.0055	0.25		[*SHISA9*]; *CPPED1*, –169,089
16	71.631–72.849	rs11643192	72,214,276	3.4×10^{-8}	A/C	1.03	0.0049	0.41		*PMFBP1*, –7,927; *DHX38*, 67,465
17	27.345–28.419	rs17727765	27,576,962	8.5×10^{-9}	T/C	0.95	0.0088	0.92		[*CRYBA1*]; *MYO18A*, –69,555; *NUFIP2*, 5,891
18	36.588–36.976	rs62099069	36,883,737	1.3×10^{-8}	A/T	0.97	0.0049	0.42		[*MIR924HG*]
18	50.358–50.958	rs11663393	50,614,732	1.6×10^{-8}	A/G	1.03	0.0049	0.45	O	[*DCC*]; *MIR4528*, –148,738
18	51.973–52.552	rs1833288	52,517,906	2.6×10^{-8}	A/G	1.03	0.0054	0.72		[*RAB27B*]; *CCDC68*, 50,833
18	52.860–53.268	rs12958048	53,101,598	3.6×10^{-11}	A/G	1.03	0.0051	0.33	S	[*TCF4*]; *MIR4529*, –44,853
22	40.818–42.216	rs5758265	41,617,897	7.6×10^{-9}	A/G	1.03	0.0054	0.28	H,S	[*L3MBTL2*]; *EP300-AS1*, –24,392; *CHADL*, 7,616

Chr. (chromosome) and region (boundaries in Mb; hg19) are shown, defined by the locations of SNPs with $P < 1 \times 10^{-5}$ and LD $r^2 > 0.1$ with the most strongly associated SNP (logistic regression; the lowest *P* value in the region listed is not corrected for multiple testing), whose location is given in base pairs. In three regions, a second SNP fulfilled the filtering criteria and was followed up with conditional analyses: chr. 1, conditional analysis selected only rs1432639 as significant, with $P = 2.0 \times 10^{-4}$ for rs12134600 after fitting rs1432639; chr. 5, conditional analysis showed two independent associations selecting rs247910 and rs10514301 as the most strongly associated SNPs; chr. 10, conditional analysis selected only rs61867293 with $P = 8.6 \times 10^{-5}$ for rs1021363 after conditioning on rs61867293. For each of the 47 SNPs, there was at least one additional genome-wide significant SNP in the cluster of surrounding SNPs with low *P* values. Chromosome X was analyzed but had no findings that met genome-wide significance. Entries in the "Prev." column indicate which of four previous studies identified genome-wide significant associations in a region: H, Hyde et al.[28], 23andMe genome-wide association of self-reported clinical depression (the discovery sample overlaps with this paper); O, Okbay et al.[16], meta-analysis of genome-wide association of MDD, depressive symptoms, psychological well-being, and neuroticism (includes many PGC29 samples); S, PGC report on 108 schizophrenia-associated loci[19]; C, CHARGE Consortium meta-analysis of depressive symptoms[14]. The "Gene context" column shows the distances between the peak SNP and the closest genes; brackets indicate that the peak SNP was within that gene. The closest genes upstream (taking strand into account; a negative number indicates distance in base pairs between the peak SNP and the nearest gene boundary) and downstream (positive distance in base pairs) are also shown if there was a flanking gene within 200 kb. The name of the closest gene is in brackets. Note that it is generally not known whether associated SNPs have biological effects on these or other more distant genes. A1/A2, the two alleles (or insertion–deletion) at each SNP; A1 was tested for association and its OR (odds ratio) and s.e. are shown; Freq., the frequency of A1 in controls across all cohorts.

variants that predispose to MDD overlap substantially with those for adult- and childhood-onset psychiatric disorders, although they remain substantially distinct as well.

Third, the common variant genetic architecture of major depression was positively correlated with multiple measures of sleep quality (daytime sleepiness, insomnia, and tiredness). The first two of these correlations used UK Biobank data with people endorsing major depression or other major psychiatric disorders, shift workers, and those taking hypnotics excluded. This pattern of correlations combined with the importance of sleep and fatigue in major depression (two criteria for MDD) suggests a close and potentially profound mechanistic relationship. Major depression also had a strong genetic correlation with neuroticism (a personality dimension assessing the degree of emotional instability); this is consistent with the literature showing a close interconnection of MDD and this personality trait. The strong negative r_g with subjective well-being underscores the capacity of major depression to impact human health.

Finally, major depression had significant negative genetic correlations with data from two studies of educational attainment, which while often considered at the genetic level as proxy measures of intelligence also likely includes more complex personality constructs. With this in mind, it is relevant to note that the r_g between major depression and IQ[61] was not significantly different from zero, despite the r_g between years of education and IQ being 0.7, implying complex relationships between these traits worthy of future investigation. We also found significant positive correlations with multiple measures of adiposity, relationship to female reproductive behavior (decreased age at menarche, age at first birth, and increased number of children), and positive correlations with coronary artery disease and lung cancer.

We used bidirectional Mendelian randomization (MR) to investigate the relationships between four traits genetically correlated with major depression: years of education (EDY)[62], body mass index (BMI)[29], coronary artery disease (CAD)[63], and schizophrenia[19]. These traits were selected because all of the following were true: they were phenotypically associated with MDD, had significant r_g with MDD, and had >30 independent genome-wide significant associations from large GWAS. We report GSMR[64] results but obtained qualitatively similar results with other MR methods (Supplementary Fig. 4 and Supplementary Table 13). MR analyses provided evidence for a 1.12-fold increase in major depression per s.d. of BMI ($P_{GSMR} = 1.2 \times 10^{-7}$) and a 0.84-fold decrease in major depression per s.d. of EDY ($P_{GSMR} = 2.3 \times 10^{-6}$). There was no evidence of reverse causality of major depression for BMI ($P_{GSMR} = 0.53$) or EDY ($P_{GSMR} = 0.11$). For BMI, there was some evidence of pleiotropy, as six BMI-associated SNPs were excluded by the HEIDI outlier test including SNPs near *OLFM4* and *NEGR1*. Thus, these results are consistent with EDY and BMI as either causal risk factors or correlated with causal risk factors for major depression. These results provide hypotheses for future research to understand these potentially directional relationships.

For CAD, the MR analyses were not significant when considering major depression as an outcome ($P_{GSMR} = 0.30$) or as an exposure ($P_{GSMR} = 0.12$); however, the high standard error of the estimates using MDD SNP instruments implies that this analysis should be revisited when more major genome-wide significant SNP instruments for depression become available from future GWAS.

We used MR to investigate the relationship between major depression and schizophrenia. Although major depression had positive r_g with many psychiatric disorders, only schizophrenia had

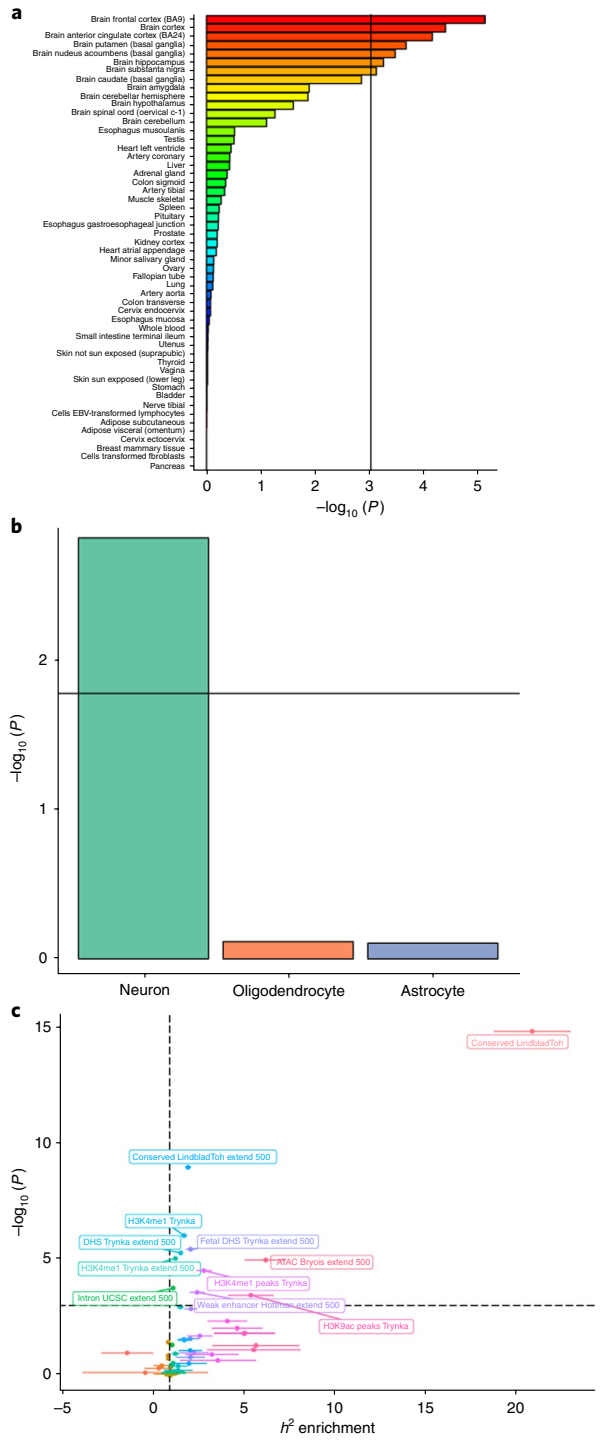


Fig. 3 | Analyses exploring enrichment of major depression association results based on different SNP annotations. a, Enrichment in bulk tissue mRNA-seq data from GTEx shown as *t* statistics; sample sizes in the GTEx database range from $n = 75$–564. The threshold for significance accounting for multiple testing is shown by the vertical line. **b**, Major depression results and enrichment in three major brain cell types shown as *t* statistics; the threshold for significance accounting for multiple testing is shown by the horizontal line. Sample sizes vary as these data were aggregated from many different sources. **c**, Partitioned LDSC to evaluate enrichment of the major depression genome-wide association findings in over 50 functional genomic annotations (Supplementary Table 8) shown as enrichment statistics; the threshold for significance accounting for multiple testing is shown by the horizontal dashed line. Sample sizes vary as these data were aggregated from many different sources.

Table 3 | LDSC genetic correlations of MDD with other disorders, diseases, and human traits

Trait	r_g	s.e.	FDR	h^2_{SNP}	Ref.
Depressive symptoms, CHARGE	0.91	0.123	3.2×10^{-12}	0.04	14
Depressive symptoms, SSGAC	0.98	0.034	1.3×10^{-176}	0.05	16
ADHD (iPSYCH-PGC)	0.42	0.033	6.1×10^{-36}	0.24	83
Anorexia nervosa	0.13	0.028	7.1×10^{-5}	0.55	84
Anxiety disorders	0.80	0.140	2.0×10^{-7}	0.06	85
Autism spectrum disorders (iPSYCH-PGC)	0.44	0.039	8.4×10^{-28}	0.20	86
Bipolar disorder	0.32	0.034	3.3×10^{-19}	0.43	20
Schizophrenia	0.34	0.025	7.7×10^{-40}	0.46	19
Smoking, ever versus never	0.29	0.038	7.0×10^{-13}	0.08	87
Daytime sleepiness[a]	0.19	0.048	5.7×10^{-4}	0.05	–
Insomnia[a]	0.38	0.038	4.0×10^{-22}	0.13	–
Tiredness	0.67	0.037	6.2×10^{-72}	0.07	88
Subjective well-being	−0.65	0.035	7.5×10^{-76}	0.03	16
Neuroticism	0.70	0.031	2.5×10^{-107}	0.09	16
College completion	−0.17	0.034	6.7×10^{-6}	0.08	89
Years of education	−0.13	0.021	1.6×10^{-8}	0.13	62
Body fat	0.15	0.038	6.5×10^{-4}	0.11	90
Body mass index	0.09	0.026	3.6×10^{-3}	0.19	32
Obesity class 1	0.11	0.029	1.6×10^{-3}	0.22	30
Obesity class 2	0.12	0.033	3.0×10^{-3}	0.18	30
Obesity class 3	0.20	0.053	1.6×10^{-3}	0.12	30
Overweight	0.13	0.030	1.4×10^{-4}	0.11	30
Waist circumference	0.11	0.024	8.2×10^{-5}	0.12	91
Waist-to-hip ratio	0.12	0.030	2.9×10^{-4}	0.11	91
Triglycerides	0.14	0.028	1.0×10^{-5}	0.17	92
Age at menarche	−0.14	0.023	6.3×10^{-8}	0.20	93
Age of first birth	−0.29	0.029	6.1×10^{-22}	0.06	94
Father's age at death	−0.28	0.058	3.0×10^{-5}	0.04	95
Number of children ever born	0.13	0.036	2.4×10^{-3}	0.03	94
Coronary artery disease	0.12	0.027	8.2×10^{-5}	0.08	63
Squamous cell lung cancer	0.26	0.075	3.6×10^{-3}	0.04	96

All genetic correlations (r_g) estimated using bivariate LDSC applied to major depression genome-wide association results are in Supplementary Table 12. Shown above are the r_g values of major depression with false discovery rate (FDR) < 0.01 (FDR estimated for 221 genetic correlations; H_0: $r_g = 0$). Thematically related traits are indicated by bold font. iPSYCH is a nationally representative cohort based on blood spots collected at birth. Within iPSYCH, r_g with ADHD was 0.58 (s.e. = 0.050) and with ASD was 0.51 (s.e. = 0.07); these values are larger than those listed above and are inconsistent with artifactual correlations. h^2_{SNP} is shown to aid interpretation as high r_g in the context of high h^2_{SNP} is more noteworthy than when h^2_{SNP} is low. [a]Self-reported daytime sleepiness and insomnia from UK Biobank excluding subjects with major depression, other psychiatric disorders (bipolar disorder, schizophrenia, autism, intellectual disability), shift workers, and those taking hypnotics.

sufficient associations for MR analyses. We found significant bidirectional correlations in SNP effect sizes for schizophrenia loci in major depression ($P_{GSMR} = 1.1 \times 10^{-40}$) and for major depression loci in schizophrenia ($P_{GSMR} = 1.5 \times 10^{-11}$). These results suggest that the major depression–schizophrenia r_g of 0.34 is consistent with

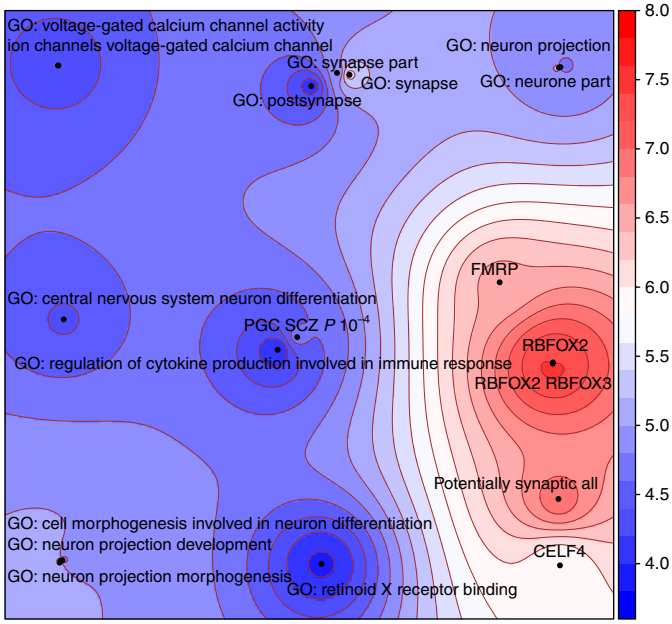


Fig. 4 | Generative topographic mapping of the 19 significant pathway results. The average position of each pathway on the map is represented by a point. The map is colored by the $-\log_{10} P$ value obtained using MAGMA. The *x* and *y* coordinates were obtained from a kernel generative topographic mapping algorithm (GTM) that reduces high-dimensional gene sets to a 2D scatterplot by accounting for gene overlap between gene sets. Each point represents a gene set. Nearby points are more similar in gene overlap than more distant points. The color surrounding each point (gene set) corresponds to significance according to the scale on the right. The significant pathways (Supplementary Table 11) fall into nine main clusters as described in the text.

partially shared biological pathways being causal for both disorders. Although it is plausible that diagnostic misclassification/ambiguity (for example, misdiagnosis of MDD as schizoaffective disorder) could contaminate these analyses, levels of misclassification would need to be implausibly high (30% unidirectional, 15% bidirectional) to result in an r_g of ~0.3[65].

All MR analyses were repeated after excluding the 23 and Me cohort, and the pattern of the results was the same (Supplementary Table 13).

Discussion

The nature of severe depression has been discussed for millennia[66]. This genome-wide association meta-analysis is among the largest ever conducted in psychiatric genetics and provides a body of results that helps refine the fundamental basis of major depression.

In conducting this meta-analysis of major depression, we employed a pragmatic approach by including cohorts that met empirical criteria for sufficient genetic and phenotypic similarity. Our approach was cautious, clinically informed, guided by empirical data, and selective (for example, we did not include cohorts with bipolar disorder (which requires MDD), depressive symptoms, neuroticism, or well-being). Approximately 44% of all major depression cases were assessed using traditional methods (PGC29, GenScot), treatment registers (iPSYCH, GERA; such approaches have been extensively used to elucidate the epidemiology of major depression), or a combination of methods (deCODE, UK Biobank), whereas ~56% of cases were from 23andMe (via self-report)[28]. Multiple lines of genetic evidence supported conducting meta-analysis of these seven cohorts (for example, out-of-sample prediction, sign tests, and genetic correlations).

However, our approach may be controversial to some readers given the unconventional reliance on self-report of major depression. We would reframe the issue: we hypothesize that brief methods of assessing major depression are informative for the genetics of MDD. We present a body of results that is consistent with this hypothesis. Even if unconventional, our hypothesis is testable and falsifiable, and we invite and welcome empirical studies to further support or refute this hypothesis.

Our results lead us to draw some broad conclusions. First, major depression is a brain disorder. Although this is not unexpected, some past models of MDD have had little or no place for heredity or biology. The genetic results best match gene expression patterns in the prefrontal and anterior cingulate cortex, anatomical regions that show differences between MDD cases and controls. The genetic findings implicated neurons (not microglia or astrocytes), and we anticipate more detailed cellular localization when sufficient single-cell and single-nucleus RNA-seq datasets become available[67].

Second, the genetic associations for major depression (as with schizophrenia)[46] tend to occur in genomic regions conserved across a range of placental mammals. Conservation suggests important functional roles. Notably, our analyses did not implicate exons or coding regions.

Third, the results also implicated developmental gene regulatory processes. For instance, the genetic findings pointed at the splicing regulator *RBFOX1* (the presence of two independent genetic associations in *RBFOX1* strongly suggests that it is the relevant gene). Gene set analyses implicated genes containing binding sites to the protein product of *RBFOX1*, and this gene set is also significantly enriched for rare exonic variation in autism and schizophrenia[56,57]. These analyses highlight the potential importance of splicing to generate alternative isoforms; risk for major depression may be mediated not by changes in isolated amino acids but rather by changes in the proportions of isoforms coming from a gene, given that isoforms often have markedly different biological functions[68,69]. These convergent results provide possible clues to a biological mechanism common to multiple severe psychiatric disorders that merits future research.

Fourth, in the most extensive analysis of the genetic 'connections' of major depression with a wide range of disorders, diseases, and human traits, we found significant positive genetic correlations with measures of body mass and negative genetic correlations with years of education, while showing no evidence of genetic correlation with IQ. MR analysis results are consistent with both BMI and years of education being causal, or correlated with causal, risk factors for major depression, and our results provide hypotheses and motivation for more detailed prospective studies, as currently available data may not provide insight about the fundamental driver or drivers of causality. The underlying mechanisms are likely more complex, as it is difficult to envision how genetic variation in educational attainment or body mass alters risk for MDD without invoking an additional mechanistic component. While the significant MR analyses need further investigations to fully understand, the negative MR results provide important evidence that there is not a direct causal relationship between MDD and subsequent changes in body mass or education years. If such associations are observed in epidemiological or clinical samples, then other factors must drive the association.

Fifth, we found significant positive correlations of major depression with all psychiatric disorders that we evaluated, including disorders prominent in childhood. This pattern of results indicates that the current classification scheme for major psychiatric disorders does not align well with the underlying genetic basis of these disorders. Currently, only schizophrenia has a sufficient number of genome-wide significant loci to conduct MR analysis, but the bidirectionally significant MR results are consistent with a shared biological basis for major depression and schizophrenia.

The dominant psychiatric nosological systems were principally designed for clinical utility and are based on data that emerge during human interactions (i.e., observable signs and reported symptoms) and not objective measurements of pathophysiology. MDD is frequently comorbid with other psychiatric disorders, and the phenotypic comorbidity has an underlying structure that reflects shared origins (as inferred from factor analyses and twin studies)[70–73]. Our genetic results add to this knowledge: major depression is not a discrete entity at any level of analysis. Rather, our data strongly suggest the existence of biological processes common to major depression and schizophrenia (and likely other psychiatric disorders).

Finally, as expected, we found that major depression had modest h^2_{SNP} (8.7%), as it is a complex malady with both genetic and environmental determinants. We found that major depression has a very high genetic correlation with proxy measures that can be briefly assessed. Lifetime major depressive disorder requires a constellation of signs and symptoms whose reliable scoring requires an extended interview with a trained clinician. However, the common variant genetic architecture of lifetime major depression in these seven cohorts (containing many subjects medically treated for MDD) has strong overlap with that of current depressive symptoms in general community samples. Similar relationships of clinically defined ADHD or autism with quantitative genetic variation in the population have been reported[74,75]. The 'disorder-versus-symptom' relationship has been debated extensively[76], but our data indicate that the common variant genetic overlap is very high. This finding has important implications.

One implication is for future genetic studies. In a first phase, it should be possible to elucidate the bulk of the common variant genetic architecture of MDD using a cost-effective shortcut—large studies of genotyped individuals who complete online self-report assessments of lifetime MDD (a sample size approaching 1 million MDD cases may be achievable by 2020). Use of online assessment could allow for recording of a broad range of phenotypes including comorbidities and putative environmental exposures, but with the key feature being large samples with consistently assessed measures. In a second phase, with a relatively complete understanding of the genetic basis of major depression, one could then evaluate smaller samples of carefully phenotyped individuals with MDD to understand the clinical importance of the genetic results. Subsequent empirical studies may show that it is possible to stratify MDD cases at first presentation to identify individuals at high risk for recurrence, poor outcome, poor treatment response, or who might subsequently develop a psychiatric disorder requiring alternative pharmacotherapy (for example, schizophrenia or bipolar disorder). This could form a cornerstone of precision medicine in psychiatry.

In summary, this genome-wide association meta-analysis of 135,438 MDD and major depression cases and 344,901 controls identified 44 loci. An extensive set of companion analyses provide insights into the nature of MDD as well as its neurobiology, therapeutic relevance, and genetic and biological interconnections to other psychiatric disorders. Comprehensive elucidation of these features is the primary goal of our genetic studies of MDD.

URLs. 1000 Genomes Project multi-ancestry imputation panel, https://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html; 23andMe privacy policy, https://www.23andme.com/en-eu/about/privacy; Bedtools, https://bedtools.readthedocs.io; dbGaP, https://www.ncbi.nlm.nih.gov/gap; genotype-based checksums for relatedness determination, http://www.broadinstitute.org/~sripke/share_links/checksums_download; GSMR, http://cnsgenomics.com/software/gsmr/; GTEx, http://www.gtexportal.org/home/datasets; GTMapTool, http://infochim.u-strasbg.fr/mobyle-cgi/portal.py#forms::gtmaptool; LD-Hub, http://ldsc.broadinstitute.org/; PGC, http://www.med.unc.edu/pgc; NIH NeuroBiobank, https://neurobiobank.nih.gov/; PGC ricopili genome-wide association analysis pipeline, https://github.com/Nealelab/ricopili; SMR, http://cnsgenomics.com/software/smr/#Overview; TWAS, http://gusevlab.org/projects/fusion/; UK Biobank, http://www.ukbiobank.ac.uk/.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0090-3.

Received: 28 July 2017; Accepted: 14 February 2018;
Published online: 26 April 2018

References

1. Kessler, R. C. & Bromet, E. J. The epidemiology of depression across cultures. *Annu. Rev. Public Health* **34**, 119–138 (2013).
2. Judd, L. L. The clinical course of unipolar major depressive disorders. *Arch. Gen. Psychiatry* **54**, 989–991 (1997).
3. Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T. & Murray, C. J. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* **367**, 1747–1757 (2006).
4. Wittchen, H. U. et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* **21**, 655–679 (2011).
5. Ferrari, A. J. et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med.* **10**, e1001547 (2013).
6. Angst, F., Stassen, H. H., Clayton, P. J. & Angst, J. Mortality of patients with mood disorders: follow-up over 34–38 years. *J. Affect. Disord.* **68**, 167–181 (2002).
7. Gustavsson, A. et al. Cost of disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* **21**, 718–779 (2011).
8. Murray, C. J. et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2197–2223 (2012).
9. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: review and meta-analysis. *Am. J. Psychiatry* **157**, 1552–1562 (2000).
10. Rice, F., Harold, G. & Thapar, A. The genetic aetiology of childhood depression: a review. *J. Child Psychol. Psychiatry* **43**, 65–79 (2002).
11. Viktorin, A. et al. Heritability of perinatal depression and genetic overlap with nonperinatal depression. *Am. J. Psychiatry* **73**, 158–165 (2016).
12. Levinson, D. F. et al. Genetic studies of major depressive disorder: why are there no GWAS findings, and what can we do about it. *Biol. Psychiatry* **76**, 510–512 (2014).
13. Major Depressive Disorder Working Group of the PGC. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).
14. Hek, K. et al. A genome-wide association study of depressive symptoms. *Biol. Psychiatry* **73**, 667–678 (2013).
15. CONVERGE Consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588–591 (2015).
16. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
17. Sullivan, P. F. et al. Psychiatric genomics. An update and an agenda. *Am. J. Psychiatry* **175**, 15–27 (2018).
18. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
19. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
20. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near *ODZ4*. *Nat. Genet.* **43**, 977–983 (2011).
21. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
22. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
23. Wray, N. R. et al. Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol. Psychiatry* **17**, 36–48 (2012).

24. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
25. Meier, S. M. et al. High loading of polygenic risk in cases with chronic schizophrenia. *Mol. Psychiatry* **21**, 969–974 (2016).
26. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
27. Wray, N. R. & Maier, R. Genetic basis of complex genetic disease: the contribution of disease heterogeneity to missing heritability. *Curr. Epidemiol. Rep.* **1**, 220–227 (2014).
28. Hyde, C. L. et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).
29. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
30. Berndt, S. I. et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat. Genet.* **45**, 501–512 (2013).
31. Bradfield, J. P. et al. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat. Genet.* **44**, 526–531 (2012).
32. Speliotes, E. K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
33. Willer, C. J. et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2009).
34. Thorleifsson, G. et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **41**, 18–24 (2009).
35. Fogel, B. L. et al. RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. *Hum. Mol. Genet.* **21**, 4171–4186 (2012).
36. Gehman, L. T. et al. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat. Genet.* **43**, 706–711 (2011).
37. Pariante, C. M. & Lightman, S. L. The HPA axis in major depression: classical theories and new developments. *Trends Neurosci.* **31**, 464–468 (2008).
38. Choi, Y. et al. SALM5 trans-synaptically interacts with LAR-RPTPs in a splicing-dependent manner to regulate synapse development. *Sci. Rep.* **6**, 26676 (2016).
39. Mah, W. et al. Selected SALM (synaptic adhesion-like molecule) family proteins regulate synapse formation. *J. Neurosci.* **30**, 5559–5568 (2010).
40. Zhu, Y. et al. Neuron-specific SALM5 limits inflammation in the CNS via its interaction with HVEM. *Sci. Adv.* **2**, e1500637 (2016).
41. Amiel, J. et al. Mutations in *TCF4*, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt–Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *Am. J. Hum. Genet.* **80**, 988–993 (2007).
42. Akbarian, S. et al. The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015).
43. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
44. Schmaal, L. et al. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol. Psychiatry* **22**, 900–909 (2017).
45. Cahoy, J. D. et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
46. Finucane, H. K. et al. Partitioning heritability by functional category using GWAS summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
47. Lindblad-Toh, K. et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
48. Simonti, C. N. et al. The phenotypic legacy of admixture between modern humans and Neandertals. *Science* **351**, 737–741 (2016).
49. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
50. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
51. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
52. Smemo, S. et al. Obesity-associated variants within *FTO* form long-range functional connections with *IRX3*. *Nature* **507**, 371–375 (2014).
53. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
54. Martin, J. S. et al. HUGIn. Hi-C Unifying Genomic Interrogator. *Bioinformatics* **33**, 3793–3795 (2017).
55. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18**, 199–209 (2015).
56. De Rubeis, S. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
57. Genovese, G. et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441 (2016).
58. Gaspar, H. A. & Breen, G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. *Sci. Rep.* **7**, 12460 (2017).
59. Breen, G. et al. Translating genome-wide association findings into new therapeutics for psychiatry. *Nat. Neurosci.* **19**, 1392–1396 (2016).
60. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
61. Sniekers, S. et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* **49**, 1107–1112 (2017).
62. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
63. Nikpay, M. et al. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
64. Zhu, Z. et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018).
65. Wray, N. R., Lee, S. H. & Kendler, K. S. Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. *Eur. J. Hum. Genet.* **20**, 668–674 (2012).
66. Hippocrates. *The Aphorisms of Hippocrates* (Collins & Co., New York, 1817).
67. Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* (in the press).
68. Yang, X. et al. Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* **164**, 805–817 (2016).
69. Zhang, X. et al. Cell-type-specific alternative splicing governs cell fate in the developing cerebral cortex. *Cell* **166**, 1147–1162 (2016).
70. Kessler, R. C. et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *J. Am. Med. Assoc.* **289**, 3095–3105 (2003).
71. Hasin, D. S., Goodwin, R. D., Stinson, F. S. & Grant, B. F. Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Arch. Gen. Psychiatry* **62**, 1097–1106 (2005).
72. Kendler, K. S. et al. The structure of genetic and environmental risk factors for syndromal and subsyndromal common DSM-IV axis I and all axis II disorders. *Am. J. Psychiatry* **168**, 29–39 (2011).
73. Kendler, K. S., Prescott, C. A., Myers, J. & Neale, M. C. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch. Gen. Psychiatry* **60**, 929–937 (2003).
74. Robinson, E. B. et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* **48**, 552–555 (2016).
75. Middeldorp, C. M. et al. A genome-wide association meta-analysis of attention-deficit/hyperactivity disorder symptoms in population-based pediatric cohorts. *J. Am. Acad. Child. Adolesc. Psychiatry* **55**, 896–905 (2016).
76. Kendell, R. E. The classification of depressions: a review of contemporary confusion. *Br. J. Psychiatry* **129**, 15–28 (1976).
77. Verduijn, J. et al. Using clinical characteristics to identify which patients with major depressive disorder have a higher genetic load for three psychiatric disorders. *Biol. Psychiatry* **81**, 316–324 (2017).
78. Smith, B. H. et al. Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int. J. Epidemiol.* **42**, 689–700 (2013).
79. Fernandez-Pujals, A. M. et al. Epidemiology and heritability of major depressive disorder, stratified by age of onset, sex, and illness course in Generation Scotland: Scottish Family Health Study (GS:SFHS). *PLoS One* **10**, e0142197 (2015).
80. Banda, Y. et al. Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. *Genetics* **200**, 1285–1295 (2015).
81. Pedersen, C. B. et al. The iPSYCH2012 case–cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **23**, 6–14 (2018).
82. Allen, N. E., Sudlow, C., Peakman, T., Collins, R. & UK Biobank. UK Biobank data: come and get it. *Sci. Transl. Med.* **6**, 224ed4 (2014).
83. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for ADHD. Preprint at *bioRxiv* https://www.biorxiv.org/content/early/2017/06/03/145581 (2017).

84. Boraska, V. et al. A genome-wide association study of anorexia nervosa. *Mol. Psychiatry* **19**, 1085–1094 (2014).
85. Otowa, T. et al. Meta-analysis of genome-wide association studies of anxiety disorders. *Mol. Psychiatry* **21**, 1391–1399 (2016).
86. Grove, J. et al. Common risk variants identified in autism spectrum disorder. Preprint at *bioRxiv* https://www.biorxiv.org/content/early/2017/11/27/224774 (2017).
87. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).
88. Deary, V. et al. Genetic contributions to self-reported tiredness. *Mol. Psychiatry* **23**, 609–620 (2018).
89. Rietveld, C. A. et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467–1471 (2013).
90. Lu, Y. et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat. Commun.* **7**, 10495 (2016).
91. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
92. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
93. Perry, J. R. et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
94. Barban, N. et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* **48**, 1462–1472 (2016).
95. Pilling, L. C. et al. Human longevity is influenced by many genetic variants: evidence from 75,000 UK Biobank participants. *Aging* **8**, 547–560 (2016).
96. Patel, Y. M. et al. Novel association of genetic markers affecting CYP2A6 activity and lung cancer risk. *Cancer Res.* **76**, 5768–5776 (2016).

Acknowledgements

Full acknowledgments are in the Supplementary Note. We are deeply indebted to the investigators who comprise the PGC, and to the hundreds of thousands of subjects who have shared their life experiences with PGC investigators. A full list of funding is in the Supplementary Note. Major funding for the PGC is from the US National Institutes of Health (U01 MH109528 and U01 MH109532). Statistical analyses were carried out on the NL Genetic Cluster Computer (http://www.geneticcluster.org/) hosted by SURFsara. The iPSYCH team acknowledges funding from the Lundbeck Foundation (grants R102-A9118 and R155-2014-1724), the Stanley Medical Research Institute, the European Research Council (project 294838), the Novo Nordisk Foundation for supporting the Danish National Biobank resource, and Aarhus and Copenhagen Universities and University Hospitals, including support to the iSEQ Center, the GenomeDK HPC facility, and the CIRRAU Center. This research has been conducted using the UK Biobank Resource (see URLs), including applications 4844 and 6818. Finally, we thank the members of the eQTLGen Consortium for allowing us to use their very large eQTL database ahead of publication. Its members are listed in Supplementary Table 14.

Some data used in this study were obtained from dbGaP (see URLs). dbGaP accession phs000021: funding support for the Genome-Wide Association of Schizophrenia Study was provided by the National Institute of Mental Health (R01 MH67257, R01 MH59588, R01 MH59571, R01 MH59565, R01 MH59587, R01 MH60870, R01 MH59566, R01 MH59586, R01 MH61675, R01 MH60879, R01 MH81800, U01 MH46276, U01 MH46289, U01 MH46318, U01 MH79469, and U01 MH79470), and the genotyping of samples was provided through the Genetic Association Information Network (GAIN). Samples and associated phenotype data for the Genome-Wide Association of Schizophrenia Study were provided by the Molecular Genetics of Schizophrenia Collaboration (principal investigator P. V. Gejman, Evanston Northwestern Healthcare (ENH) and Northwestern University, Evanston, IL, USA). dbGaP accession phs000196: this work used in part data from the NINDS dbGaP database from the CIDR:NGRC PARKINSON'S DISEASE STUDY. dbGaP accession phs000187: High-Density SNP Association Analysis of Melanoma: Case–Control and Outcomes Investigation. Research support to collect data and develop an application to support this project was provided by P50 CA093459, P50 CA097007, R01 ES011740, and R01 CA133996 from the NIH.

Author contributions

Writing group: G.B., A.D.B., D.F.L., C.M.L., S.R., P.F.S., N.R.W. **PGC MDD PI group:** V.A., B.T.B., K.B., D.I.B., G.B., A.D.B., S.C., U.D., J.R.D., E.D., K.D., T.E., E.J.C.d.G., H.J.G., S.P.H., C. Hayward, A.C.H., D.M.H., K.S.K., S.K., D.F.L., C.M.L., G.L., Q.S.L., S.L., P.A.F.M., P.K.M., N.G.M., A.M.M., A.M., O.M., P.B.M., B.M.-M., M. Nordentoft, M.M.N., M.C.O'D., S.A.P., N.L.P., B.W.P., R.H.P., D.J.P., J.B.P., M.P., M. Rietschel, C.S., T.G.S., J.W.S., K.S., P.F.S., H. Tiemeier, R.U., H.V., M.M.W., T.W., A.R.W., N.R.W. **Bioinformatics:** 23andMe Research Team, M.J.A., S.V.d.A., G.B., J.B., A.D.B., E.C., J.H.C., T.-K.C., J.R.I.C., L.C.-C., eQTLGen Consortium, G.E.C., C.A.C., G.D., E.M.D., T.E., A.J.F., H.A.G., P.G.-R., J.G., L.S.H., E.H., T.F.H., C. Hayward, M.H., R.J., F.J., Z.K., Q.S.L., Yihan Li, P.A.L., X.L., L.L., D.J.M., S.E.M., E.M., Y.M., J. Mill, J.N.P., B.W.P., W.J.P., G.P., P.Q., L.S., S.I.S., C.A.S., P.F.S., K.E.T., A.T., P.A.T., A.G.U., Y. Wang, S.M.W., N.R.W., H.S.X. **Clinical:** E.A., T.M.A., V.A., B.T.B., A.T.F.B., K.B., E.B.B., D.H.R.B., H.N.B., A.D.B., N. Craddock, U.D., J.R.D., N.D., K.D., M.G., F.S.G., H.J.G., A.C.H., A.M.v.H., I.B.H., M.I., S.K., J. Krogh, D.F.L., S.L., D.J.M., D.F.M., P.A.F.M., W.M., N.G.M., P. McGrath, P. McGuffin, A.M.M., A.M., C.M.M., S.S.M., F.M.M., O.M., P.B.M., D.R.N., H.O., M.J.O., C.B.P., M.G.P., J.B.P., J.A.Q., J.P.R., M. Rietschel, C.S., R. Schoevers, E.S., G.C.B.S., D.J.S., F.S., J. Strohmaier, D.U., M.M.W., J.W., T.W., G.W. **Genomic assays:** G.B., H.N.B., J.B.-G., M.B.-H., A.D.B., S.C., T.-K.C., F.D., A.J.F., S.P.H., C.S.H., A.C.H., P.H., G.H., C. Horn, J.A.K., P.A.F.M., L.M., G.W.M., M. Nauck, M.M.N., M. Rietschel, M. Rivera, E.C.S., T.G.S., S.I.S., H.S., F.S., T.E.T., J.T., A.G.U., S.H.W. **Obtained funding for primary MDD samples:** B.T.B., K.B., D.H.R.B., D.I.B., G.B., H.N.B., A.D.B., S.C., J.R.D., I.J.D., E.D., T.C.E., T.E., H.J.G., S.P.H., A.C.H., D.M.H., I.S.K., D.F.L., C.M.L., G.L., Q.S.L., S.L., P.A.F.M., W.M., N.G.M., P. McGuffin, A.M.M., A.M., G.W.M., O.M., P.B.M., M. Nordentoft, D.R.N., M.M.N., P.F.O'R., B.W.P., D.J.P., J.B.P., M.P., M. Rietschel, C.S., T.G.S., G.C.B.S., J.H.S., D.J.S., H.S., K.S., P.F.S., T.E.T., H. Tiemeier, A.G.U., H.V., M.M.W., T.W., N.R.W. **Statistical analysis:** 23andMe Research Team, A.A., M.J.A., T.F.M.A., S.V.d.A., S.-A.B., K.B., T.B.B., G.B., E.M.B., A.D.B., N. Cai, T.-K.C., J.R.I.C., B.C.-D., H.S.D., G.D., N.D., C.V.D., E.C.D., N.E., V.E.-P., T.E., H.K.F., J.F., H.A.G., S.D.G., J.G., L.S.H., C. Hayward, A.C.H., S.H., D.A.H., J.-J.H., C.L.H., M.I., E.J., F.F.H.K., J. Kraft, W.W.K., Z.K., J.M.L., C.M.L., Q.S.L., Yun Li, D.J.M., P.A.F.M., R.M.M., J. Marchini, M.M., H.M., A.M.M., S.E.M., D.M., E.M., Y.M., S.S.M., S.M., N.M., B.M.-M., B.N., M.G.N., D.R.N., P.F.O'R., R.E.P., E.P., W.J.P., G.P., D.P., S.M.P., B.P.R., S.R., M. Rivera, R. Saxena, C.S., L.S., J. Shi, S.I.S., H.S., S.S., P.F.S., K.E.T., H. Teismann, A.T., W.T., P.A.T., T.E.T., C.T., M. Traylor, V.T., M. Trzaskowski, A.V., P.M.V., Y. Wang, B.T.W., J.W., T.W., N.R.W., Y. Wu, J.Y., F.Z.

Competing interests

A.T.F.B. is on speaker's bureaus for Lundbeck and GlaxoSmithKline. G.C. is a cofounder of Element Genomics. E.D. was an employee of Hoffmann–La Roche at the time this study was conducted and a consultant to Roche and Pierre-Fabre. N.E. is employed by 23andMe, Inc., and owns stock in 23andMe, Inc. D.H. is an employee of and owns stock options in 23andMe, Inc. S.P. is an employee of Pfizer, Inc. C.L.H. is an employee of Pfizer, Inc. A.R.W. was a former employee and stockholder of Pfizer, Inc. J.A.Q. was an employee of Hoffmann–La Roche at the time this study was conducted. H.S. is an employee of deCODE Genetics/Amgen. K.S. is an employee of deCODE Genetics/Amgen. S.S. is an employee of deCODE Genetics/Amgen. P.F.S. is on the scientific advisory board for Pfizer, Inc., and the advisory committee for Lundbeck. T.E.T. is an employee of deCODE Genetics/Amgen. C.T. is an employee of and owns stock options in 23andMe, Inc.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0090-3.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to N.R.W. or P.F.S.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Naomi R. Wray[1,2,160]*, Stephan Ripke[3,4,5,160], Manuel Mattheisen[6,7,8,9,160], Maciej Trzaskowski[1,160], Enda M. Byrne[1], Abdel Abdellaoui[10], Mark J. Adams[11], Esben Agerbo[8,12,13], Tracy M. Air[14], Till M. F. Andlauer[15,16], Silviu-Alin Bacanu[17], Marie Bækvad-Hansen[8,18], Aartjan F. T. Beekman[19], Tim B. Bigdeli[17,20], Elisabeth B. Binder[15,21], Douglas R. H. Blackwood[11], Julien Bryois[22], Henriette N. Buttenschøn[7,8,23], Jonas Bybjerg-Grauholm[8,18], Na Cai[24,25], Enrique Castelao[26], Jane Hvarregaard Christensen[6,7,8], Toni-Kim Clarke[11], Jonathan I. R. Coleman[27], Lucía Colodro-Conde[28], Baptiste Couvy-Duchesne[2,29], Nick Craddock[30], Gregory E. Crawford[31,32], Cheynna A. Crowley[33], Hassan S. Dashti[3,34], Gail Davies[35], Ian J. Deary[35], Franziska Degenhardt[36,37], Eske M. Derks[28], Nese Direk[38,39], Conor V. Dolan[10], Erin C. Dunn[40,41,42], Thalia C. Eley[27], Nicholas Eriksson[43], Valentina Escott-Price[44], Farnush Hassan Farhadi Kiadeh[45], Hilary K. Finucane[46,47], Andreas J. Forstner[36,37,48,49], Josef Frank[50], Héléna A. Gaspar[27], Michael Gill[51], Paola Giusti-Rodríguez[52], Fernando S. Goes[53], Scott D. Gordon[54], Jakob Grove[6,7,8,55], Lynsey S. Hall[11,56], Eilis Hannon[57], Christine Søholm Hansen[8,18], Thomas F. Hansen[58,59,60], Stefan Herms[36,37,49], Ian B. Hickie[61], Per Hoffmann[36,37,49], Georg Homuth[62], Carsten Horn[63], Jouke-Jan Hottenga[10], David M. Hougaard[8,18], Ming Hu[64], Craig L. Hyde[65], Marcus Ising[66], Rick Jansen[19], Fulai Jin[67,68], Eric Jorgenson[69], James A. Knowles[70], Isaac S. Kohane[71,72,73], Julia Kraft[5], Warren W. Kretzschmar[74], Jesper Krogh[75], Zoltán Kutalik[76,77], Jacqueline M. Lane[3,34,78], Yihan Li[74], Yun Li[33,52], Penelope A. Lind[28], Xiaoxiao Liu[68], Leina Lu[68], Donald J. MacIntyre[79,80], Dean F. MacKinnon[53], Robert M. Maier[2], Wolfgang Maier[81], Jonathan Marchini[82], Hamdi Mbarek[10], Patrick McGrath[83], Peter McGuffin[27], Sarah E. Medland[28], Divya Mehta[2,84], Christel M. Middeldorp[10,85,86], Evelin Mihailov[87], Yuri Milaneschi[19], Lili Milani[87], Jonathan Mill[57], Francis M. Mondimore[53], Grant W. Montgomery[1], Sara Mostafavi[88,89], Niamh Mullins[27], Matthias Nauck[90,91], Bernard Ng[89], Michel G. Nivard[10], Dale R. Nyholt[92], Paul F. O'Reilly[27], Hogni Oskarsson[93], Michael J. Owen[94], Jodie N. Painter[28], Carsten Bøcker Pedersen[8,12,13], Marianne Giørtz Pedersen[8,12,13], Roseann E. Peterson[17,95], Erik Pettersson[22], Wouter J. Peyrot[19], Giorgio Pistis[26], Danielle Posthuma[96,97], Shaun M. Purcell[98], Jorge A. Quiroz[99], Per Qvist[6,7,8], John P. Rice[100], Brien P. Riley[17], Margarita Rivera[27,101], Saira Saeed Mirza[39], Richa Saxena[3,34,78], Robert Schoevers[102], Eva C. Schulte[103,104], Ling Shen[69], Jianxin Shi[105], Stanley I. Shyn[106], Engilbert Sigurdsson[107], Grant B. C. Sinnamon[108], Johannes H. Smit[19], Daniel J. Smith[109], Hreinn Stefansson[110], Stacy Steinberg[110], Craig A. Stockmeier[111], Fabian Streit[50], Jana Strohmaier[50], Katherine E. Tansey[112], Henning Teismann[113], Alexander Teumer[114], Wesley Thompson[8,59,115,116], Pippa A. Thomson[117], Thorgeir E. Thorgeirsson[110], Chao Tian[43], Matthew Traylor[118], Jens Treutlein[50], Vassily Trubetskoy[5], André G. Uitterlinden[119], Daniel Umbricht[120], Sandra Van der Auwera[121], Albert M. van Hemert[122], Alexander Viktorin[22], Peter M. Visscher[1,2], Yunpeng Wang[8,59,115], Bradley T. Webb[123], Shantel Marie Weinsheimer[8,59], Jürgen Wellmann[113], Gonneke Willemsen[10], Stephanie H. Witt[50], Yang Wu[1], Hualin S. Xi[124], Jian Yang[1,2], Futao Zhang[1], eQTLGen[125], 23andMe, Volker Arolt[126], Bernhard T. Baune[14], Klaus Berger[113], Dorret I. Boomsma[10], Sven Cichon[36,49,127,128], Udo Dannlowski[126], E. C. J. de Geus[10,129], J. Raymond DePaulo[53], Enrico Domenici[130], Katharina Domschke[131], Tõnu Esko[3,87], Hans J. Grabe[121], Steven P. Hamilton[132], Caroline Hayward[133], Andrew C. Heath[100], David A. Hinds[43], Kenneth S. Kendler[17], Stefan Kloiber[66,134,135], Glyn Lewis[136], Qingqin S. Li[137], Susanne Lucae[66], Pamela F. A. Madden[100], Patrik K. Magnusson[22], Nicholas G. Martin[54], Andrew M. McIntosh[11,35], Andres Metspalu[87,138], Ole Mors[8,139], Preben Bo Mortensen[7,8,12,13], Bertram Müller-Myhsok[15,16,140], Merete Nordentoft[8,141], Markus M. Nöthen[36,37], Michael C. O'Donovan[94], Sara A. Paciga[142], Nancy L. Pedersen[22], Brenda W. J. H. Penninx[19], Roy H. Perlis[41,143], David J. Porteous[117],

James B. Potash[144], Martin Preisig[26], Marcella Rietschel[50], Catherine Schaefer[69], Thomas G. Schulze[50,53,104,145,146], Jordan W. Smoller[40,41,42], Kari Stefansson[110,147], Henning Tiemeier[39,148,149], Rudolf Uher[150], Henry Völzke[114], Myrna M. Weissman[83,151], Thomas Werge[8,59,152], Ashley R. Winslow[153,154], Cathryn M. Lewis[27,155,160], Douglas F. Levinson[156,160], Gerome Breen[27,157,160], Anders D. Børglum[6,7,8,160], Patrick F. Sullivan[22,52,158,160]* and the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium

[1]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. [2]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [3]Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [4]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [5]Department of Psychiatry and Psychotherapy, Universitätsmedizin Berlin Campus Charité Mitte, Berlin, Germany. [6]Department of Biomedicine, Aarhus University, Aarhus, Denmark. [7]iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark. [8]iPSYCH, Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark. [9]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. [10]Department of Biological Psychology and EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [11]Division of Psychiatry, University of Edinburgh, Edinburgh, UK. [12]Centre for Integrated Register-Based Research, Aarhus University, Aarhus, Denmark. [13]National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark. [14]Discipline of Psychiatry, University of Adelaide, Adelaide, South Australia, Australia. [15]Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. [16]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [17]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA. [18]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. [19]Department of Psychiatry, Vrije Universiteit Medical Center and GGZ inGeest, Amsterdam, The Netherlands. [20]Virginia Institute for Psychiatric and Behavior Genetics, Richmond, VA, USA. [21]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA. [22]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [23]Department of Clinical Medicine, Translational Neuropsychiatry Unit, Aarhus University, Aarhus, Denmark. [24]Statistical Genomics and Systems Genetics, European Bioinformatics Institute (EMBL-EBI), Cambridge, UK. [25]Human Genetics, Wellcome Trust Sanger Institute, Cambridge, UK. [26]Department of Psychiatry, University Hospital of Lausanne, Prilly, Switzerland. [27]MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, UK. [28]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Herston, Queensland, Australia. [29]Centre for Advanced Imaging, University of Queensland, Brisbane, Queensland, Australia. [30]Psychological Medicine, Cardiff University, Cardiff, UK. [31]Center for Genomic and Computational Biology, Duke University, Durham, NC, USA. [32]Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC, USA. [33]Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [34]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [35]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [36]Institute of Human Genetics, University of Bonn, Bonn, Germany. [37]Life & Brain Center, Department of Genomics, University of Bonn, Bonn, Germany. [38]Psychiatry, Dokuz Eylul University School of Medicine, Izmir, Turkey. [39]Epidemiology, Erasmus MC, Rotterdam, The Netherlands. [40]Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA. [41]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [42]Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Massachusetts General Hospital, Boston, MA, USA. [43]Research, 23andMe, Inc., Mountain View, CA, USA. [44]Neuroscience and Mental Health, Cardiff University, Cardiff, UK. [45]Bioinformatics, University of British Columbia, Vancouver, British Columbia, Canada. [46]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [47]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA. [48]Department of Psychiatry (UPK), University of Basel, Basel, Switzerland. [49]Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, Switzerland. [50]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [51]Department of Psychiatry, Trinity College Dublin, Dublin, Ireland. [52]Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [53]Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD, USA. [54]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [55]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [56]Institute of Genetic Medicine, Newcastle University, Newcastle-upon-Tyne, UK. [57]University of Exeter Medical School, Exeter, UK. [58]Danish Headache Centre, Department of Neurology, Rigshospitalet, Glostrup, Denmark. [59]Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Capital Region of Denmark, Copenhagen, Denmark. [60]iPSYCH, Lundbeck Foundation Initiative for Psychiatric Research, Copenhagen, Denmark. [61]Brain and Mind Centre, University of Sydney, Sydney, New South Wales, Australia. [62]Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University Medicine and Ernst Moritz Arndt University Greifswald, Greifswald, Germany. [63]Roche Pharmaceutical Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, F. Hoffmann–La Roche, Ltd, Basel, Switzerland. [64]Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA. [65]Statistics, Pfizer Global Research and Development, Groton, CT, USA. [66]Max Planck Institute of Psychiatry, Munich, Germany. [67]Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH, USA. [68]Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH, USA. [69]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. [70]Psychiatry and Behavioral Sciences, University of Southern California, Los Angeles, CA, USA. [71]Informatics Program, Boston Children's Hospital, Boston, MA, USA. [72]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [73]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [74]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [75]Department of Endocrinology at Herlev University Hospital, University of Copenhagen, Copenhagen, Denmark. [76]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [77]Institute of Social and Preventive Medicine (IUMSP), University Hospital of Lausanne, Lausanne, Switzerland. [78]Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA, USA. [79]Mental Health, NHS 24 Glasgow, UK. [80]Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. [81]Department of Psychiatry and Psychotherapy, University of Bonn, Bonn, Germany. [82]Statistics, University of Oxford, Oxford, UK. [83]Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY, USA. [84]School of Psychology and Counseling, Queensland University of Technology, Brisbane, Queensland, Australia. [85]Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Service, South Brisbane, Queensland, Australia. [86]Child Health Research Centre, University of Queensland, Brisbane, Queensland, Australia. [87]Estonian Genome Center, University of Tartu, Tartu, Estonia. [88]Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada. [89]Statistics, University of British Columbia, Vancouver, British Columbia, Canada. [90]DZHK (German Centre for Cardiovascular Research), partner site Greifswald, University Medicine, University Medicine Greifswald, Greifswald, Germany. [91]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. [92]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia.

[93]Humus, Reykjavik, Iceland. [94]MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK. [95]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [96]Complex Trait Genetics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [97]Clinical Genetics, Vrije Universiteit Medical Center, Amsterdam, The Netherlands. [98]Department of Psychiatry, Brigham and Women's Hospital, Boston, MA, USA. [99]Solid Biosciences, Boston, MA, USA. [100]Department of Psychiatry, Washington University in St. Louis School of Medicine, St. Louis, MO, USA. [101]Department of Biochemistry and Molecular Biology II, Institute of Neurosciences, Center for Biomedical Research, University of Granada, Granada, Spain. [102]Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. [103]Department of Psychiatry and Psychotherapy, Medical Center of the University of Munich, Campus Innenstadt, Munich, Germany. [104]Institute of Psychiatric Phenomics and Genomics (IPPG), Medical Center of the University of Munich, Campus Innenstadt, Munich, Germany. [105]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. [106]Behavioral Health Services, Kaiser Permanente Washington, Seattle, WA, USA. [107]Faculty of Medicine, Department of Psychiatry, University of Iceland, Reykjavik, Iceland. [108]School of Medicine and Dentistry, James Cook University, Townsville, Queensland, Australia. [109]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK. [110]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [111]Psychiatry and Human Behavior, University of Mississippi Medical Center, Jackson, MS, USA. [112]College of Biomedical and Life Sciences, Cardiff University, Cardiff, UK. [113]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. [114]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [115]KG Jebsen Centre for Psychosis Research, Norway Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [116]Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA. [117]Medical Genetics Section, CGEM, IGMM, University of Edinburgh, Edinburgh, UK. [118]Clinical Neurosciences, University of Cambridge, Cambridge, UK. [119]Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. [120]Roche Pharmaceutical Research and Early Development, Neuroscience, Ophthalmology and Rare Diseases Discovery and Translational Medicine Area, Roche Innovation Center Basel, F. Hoffmann–La Roche, Ltd, Basel, Switzerland. [121]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [122]Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands. [123]Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [124]Computational Sciences Center of Emphasis, Pfizer Global Research and Development, Cambridge, MA, USA. [125]A full list of members and affiliations appears in Supplementary Table 14. [126]Department of Psychiatry, University of Münster, Munster, Germany. [127]Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, Germany. [128]Institute of Medical Genetics and Pathology, University Hospital Basel, University of Basel, Basel, Switzerland. [129]Amsterdam Public Health Institute, Vrije Universiteit Medical Center, Amsterdam, The Netherlands. [130]Centre for Integrative Biology, Università degli Studi di Trento, Trento, Italy. [131]Department of Psychiatry and Psychotherapy, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany. [132]Psychiatry, Kaiser Permanente Northern California, San Francisco, CA, USA. [133]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [134]Centre for Addiction and Mental Health, Toronto, Ontario, Canada. [135]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada. [136]Division of Psychiatry, University College London, London, UK. [137]Neuroscience Therapeutic Area, Janssen Research and Development, LLC, Titusville, NJ, USA. [138]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [139]Psychosis Research Unit, Aarhus University Hospital, Risskov, Aarhus, Denmark. [140]Institute of Translational Medicine, University of Liverpool, Liverpool, UK. [141]Mental Health Center Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark. [142]Human Genetics and Computational Biomedicine, Pfizer Global Research and Development, Groton, CT, USA. [143]Psychiatry, Harvard Medical School, Boston, MA, USA. [144]Psychiatry, University of Iowa, Iowa City, IA, USA. [145]Human Genetics Branch, NIMH Division of Intramural Research Programs, Bethesda, MD, USA. [146]Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany. [147]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [148]Child and Adolescent Psychiatry, Erasmus MC, Rotterdam, The Netherlands. [149]Psychiatry, Erasmus MC, Rotterdam, The Netherlands. [150]Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada. [151]Division of Epidemiology, New York State Psychiatric Institute, New York, NY, USA. [152]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [153]Human Genetics and Computational Biomedicine, Pfizer Global Research and Development, Cambridge, MA, USA. [154]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [155]Department of Medical and Molecular Genetics, King's College London, London, UK. [156]Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. [157]NIHR BRC for Mental Health, King's College London, London, UK. [158]Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [159]These authors contributed equally: Naomi R. Wray, Stephan Ripke, Manuel Mattheisen, Maciej Trzaskowski. [160]These authors jointly directed this work: Cathryn M. Lewis, Douglas F. Levinson, Gerome Breen, Anders D. Børglum, Patrick F. Sullivan. *e-mail: naomi.wray@uq.edu.au; pfsulliv@med.unc.edu

Methods

PGC29 cohort. Our analysis was anchored in a genome-wide association mega-analysis of 29 samples of European ancestry (16,823 MDD cases and 25,632 controls). Supplementary Table 1 summarizes the source and inclusion/exclusion criteria for cases and controls for each sample. All PGC29 samples passed a structured methodological review by MDD assessment experts (D.F.L. and K.S.K.). Cases were required to meet international consensus criteria (DSM-IV, ICD-9, or ICD-10)[97–99] for a lifetime diagnosis of MDD established using structured diagnostic instruments from assessments by trained interviewers, clinician-administered checklists, or medical record review. All cases met standard criteria for MDD, were directly interviewed (28/29 samples), or had medical record review by an expert diagnostician (1/29 samples), and most were ascertained from clinical sources (19/29 samples). Controls in most samples were screened for the absence of lifetime MDD (22/29 samples) and randomly selected from the population.

Additional cohorts. We critically evaluated six independent European-ancestry cohorts (118,635 cases and 319,269 controls). Supplementary Table 2 summarizes the source and inclusion/exclusion criteria for cases and controls for each cohort. These cohorts used a range of methods for assessing MDD or major depression. Most studies included here applied otherwise typical inclusion and exclusion criteria for both cases and controls (for example, excluding cases with lifetime bipolar disorder or schizophrenia and excluding controls with major depression).

Cohort comparability. Supplementary Table 3 summarizes the numbers of cases and controls in PGC29 and the six additional cohorts. The most direct and important way to evaluate the comparability of these cohorts for a genome-wide association meta-analysis is using SNP genotype data[22,24]. We used LD score (LDSC) regression (described below) to estimate h^2_{SNP} for each cohort (Supplementary Fig. 1 and Supplementary Table 3) and r_g for all pairwise combinations of the cohorts (Supplementary Table 3b) and to demonstrate no evidence of sample overlap. We used leave-one-sample-out GRSs, finding significant differences in case–control GRS distributions of the left-out sample for all but one PGC29 sample (Supplementary Table 4). For full details of the cohort comparability analyses including GRS analyses, see the Supplementary Note. In GRS analyses, the discovery sample was the genome-wide association sample that provided the allelic weightings for each SNP, used to generate a sum score for each individual in the independent target sample.

Genotyping and quality control. Genotyping procedures can be found in the primary reports for each cohort (summarized in Supplementary Table 3). Individual genotype data for all PGC29 samples, GERA, and iPSYCH were processed using the PGC ricopili pipeline (see URLs) for standardized quality control, imputation, and analysis[19]. The cohorts from deCODE, Generation Scotland, UK Biobank, and 23andMe were processed by the collaborating research teams using comparable procedures. SNPs and insertion–deletion polymorphisms were imputed using the 1000 Genomes Project multi-ancestry reference panel (see URLs)[100]. More detailed information on sample quality control is provided in the Supplementary Note.

LD score regression. LDSC was used to estimate h^2_{SNP} from genome-wide association summary statistics. Estimates of h^2_{SNP} on the liability scale depend on the assumed lifetime prevalence of MDD in the population (K), and we assumed $K = 0.15$ but also evaluated a range of estimates of K to explore sensitivity, including 95% confidence intervals (Supplementary Fig. 1). LDSC bivariate genetic correlations attributable to genome-wide SNPs (r_g) were estimated across all MDD and major depression cohorts and between the full cohort subjected to meta-analysis and other traits and disorders.

LDSC was also used to partition h^2_{SNP} by genomic features[24,46]. We tested for enrichment of h^2_{SNP} based on genomic annotations, partitioning h^2_{SNP} proportional to the base-pair length represented by each annotation. We used the 'baseline model', which consists of 53 functional categories. The categories are fully described elsewhere[46] and included conserved regions[47], UCSC gene models (exons, introns, promoters, UTRs), and functional genomic annotations constructed using data from ENCODE[101] and the Roadmap Epigenomics Consortium[102]. We complemented these annotations by adding introgressed regions from the Neanderthal genome in European populations[103] and open chromatin regions from the brain dorsolateral prefrontal cortex. The open chromatin regions were obtained from an ATAC–seq experiment performed in 288 samples ($n = 135$ controls, 137 schizophrenia cases, 10 bipolar cases, and 6 affective disorder cases)[104]. Peaks called with MACS[105] (1% FDR) were retained if their coordinates overlapped in at least two samples. The peaks were recentered and set to a fixed width of 300 bp using the diffbind R package[106]. To prevent upward bias in heritability enrichment estimation, we added two categories created by expanding both the Neanderthal introgressed regions and open chromatin regions by 250 bp on each side.

We used LDSC to estimate r_g between major depression and a range of other disorders, diseases, and human traits[22]. The intent of these comparisons was to evaluate the extent of shared common variant genetic architectures to suggest hypotheses about the fundamental genetic basis of major depression (given its extensive comorbidity with psychiatric and medical conditions and its association with anthropometric and other risk factors). Subject overlap of itself does not bias r_g. These r_g values were mostly based on studies of independent subjects, and the estimates should be unbiased by confounding of genetic and non-genetic effects (except if there is genotype by environment correlation). When GWAS include overlapping samples, r_g remains unbiased but the intercept of the LDSC regression is an estimate of the correlation between the association statistics attributable to sample overlap. These calculations were done using the internal PGC genome-wide association library and with LD-Hub (see URLs)[60].

Integration of genome-wide association findings to tissue and cellular gene expression. We used partitioned LDSC to evaluate which somatic tissues were enriched for major depression heritability[107]. Gene expression data generated using mRNA-seq from multiple human tissues were obtained from GTEx v6p (see URLs). Genes for which <4 samples had at least one read count per million were discarded, and samples with <100 genes with at least one read count per million were excluded. The data were normalized, and a t statistic was obtained for each tissue by comparing the expression in each tissue with the expression of all other tissues with the exception of tissues related to the tissue of interest (for example, brain cortex versus all other tissues excluding other brain samples), using sex and age as covariates. A t statistic was also obtained for each tissue among its related tissues (for example, cortex versus all other brain tissues) to test which brain region was the most associated with major depression, also using sex and age as covariates. The top 10% of the genes with the most extreme t statistic were defined as tissue specific. The coordinates for these genes were extended by a 100-kb window and tested using LD score regression. Significance was obtained from the coefficient z score, which corrects for all other categories in the baseline model.

Lists of genes specifically expressed in neurons, astrocytes, and oligodendrocytes were obtained from Cahoy et al.[45]. As these experiments were done in mice, genes were mapped to human orthologous genes using Ensembl. The coordinates for these genes were extended by a 100-kb window and tested using LD score regression as for the GTEx tissue-specific genes.

We conducted eQTL lookups of the most associated SNPs in each region and report genome-wide association SNPs in LD ($r^2 > 0.8$) with the top eQTLs in the following datasets: eQTLGen Consortium (Illumina arrays in whole blood, $n = 14{,}115$), BIOS (RNA-seq in whole blood, $n = 2{,}116$)[108], NESDA/NTR (Affymetrix arrays in whole blood, $n = 4{,}896$)[109], GEUVADIS (RNA-seq in LCLs, $n = 465$)[110], Rosmap (RNA-seq in cortex, $n = 494$)[111], GTEx (RNA-seq in 44 tissues, $n > 70$)[43], and Common Mind Consortium (CMC; prefrontal cortex, Sage Synapse accession syn5650509, $n = 467$)[51].

We used SMR[49] to identify loci with strong evidence of causality via gene expression and DNA methylation (eQTLs and meQTLs). SMR analysis is limited to significant cis-SNP expression (FDR < 0.05) and SNPs with MAF >0.01 at a Bonferroni-corrected pSMR. Owing to LD, multiple SNPs may be associated with the expression of a gene, and some SNPs are associated with the expression of more than one gene. The aim of SMR is to prioritize variants and genes for subsequent studies, and a test for heterogeneity excludes regions that may harbor multiple causal loci (pHET < 0.05; a very conservative threshold). SMR analyses were conducted using eQTLs from the eQTLGen Consortium (whole blood), GTEx (11 brain tissues), and the CMC[43,51] as well as meQTLs from whole blood[112].

We conducted a transcriptome-wide association study[50] using precomputed expression reference weights for CMC data (5,420 genes with significant cis-SNP heritability) provided with the TWAS/FUSION software. The significance threshold was 0.05/5,420.

DNA looping using Hi-C. Dorsolateral prefrontal cortex (Brodmann area 9) was dissected from post-mortem samples from three adults of European ancestry (C.S.). Cerebra from three fetal brains were obtained from the NIH NeuroBiobank (see URLs; gestation age 17–19 weeks, African ancestry). We used 'easy Hi-C' to assess DNA chromatin (looping) interactions (Supplementary Note).

Gene-wise and pathway analyses. Our approach was guided by rigorous method comparisons conducted by PGC members[55,113]. P values quantifying the degree of association of genes and gene sets with MDD were generated using MAGMA (v1.06)[114]. MAGMA uses Brown's method to combine SNP P values and account for LD. We used Ensembl gene models for 19,079 genes, giving a Bonferroni-corrected P-value threshold of 2.6×10^{-6}. Gene set P values were obtained using a competitive analysis that tests whether genes in a gene set are more strongly associated with the phenotype than other gene sets. We used European-ancestry subjects from the 1000 Genomes Project (Phase 3 v5a, MAF ≥ 0.01)[115] for the LD reference. The gene window used was 35 kb upstream and 10 kb downstream to include regulatory elements.

Gene sets were from two main sources. First, we included gene sets previously shown to be important for psychiatric disorders (71 gene sets; for example, FMRP binding partners, de novo mutations, GWAS top SNPs, ion channels)[57,116,117]. Second, we included gene sets from MSigDB (v5.2)[118], which includes canonical pathways and Gene Ontology gene sets. Canonical pathways were curated from BioCarta, KEGG, Matrisome, the Pathway Interaction Database, Reactome, Sigma-Aldrich, Signaling Gateway, Signal Transduction KE, and SuperArray. Pathways containing from 10–10,000 genes were included.

To evaluate gene sets related to antidepressants, gene sets were extracted from the Drug–Gene Interaction database (DGIdb v.2.0)[119] and the NIMH Psychoactive Drug Screening Program Ki DB[120] downloaded in June 2016. The association of 3,885 drug gene sets with major depression was estimated using MAGMA (v1.6). The drug gene sets were ordered by *P* value, and the Wilcoxon–Mann–Whitney test was used to assess whether the 42 antidepressant gene sets in the dataset (ATC code N06A in the Anatomical Therapeutic Chemical Classification System) had a higher ranking than expected by chance.

One issue is that some gene sets contain overlapping genes, and these may reflect largely overlapping results. The pathway map was constructed using the kernel generative topographic mapping algorithm (k-GTM) as described by Olier et al.[121]. GTM is a probabilistic alternative to Kohonen maps: the kernel variant is used when the input is a similarity matrix. The GTM and k-GTM algorithms are implemented in GTMapTool (see URLs). We used the Jaccard similarity matrix of FDR-significant pathways as input for the algorithm, where each pathway is encoded by a vector of binary values representing the presence (1) or absence (0) of a gene. The parameters for the k-GTM algorithm are the square root of the number of grid points (k), the square root of the number of RBF functions (m), the regularization coefficient (l), the RBF width factor (w), and the number of feature space dimensions for the kernel algorithm (b). We set k equal to the square root of the number of pathways, m equal to the square root of k, $l = 1$ (default), $w = 1$ (default), and b equal to the number of principal components explaining 99.5% of the variance in the kernel matrix. The output of the program is a set of coordinates representing the average positions of pathways on a 2D map. The x and y axes represent the dimensions of a 2D latent space. The pathway coordinates and corresponding MAGMA *P* values were used to build the pathway activity landscape using the kriging interpolation algorithm implemented in the R gstat package.

Mendelian randomization. We conducted bidirectional MR[122] analysis for four traits: years of education (EDY)[62], body mass index (BMI)[29], coronary artery disease (CAD)[63], and schizophrenia (SCZ)[19]. We denote z as a genetic variant (a SNP) that is significantly associated with x, an exposure or putative causal trait for y (the disease/trait outcome). The effect size of x on y can be estimated using a two-step least-squares (2SLS)[123] approach: $\hat{b}_{xy} = \hat{b}_{zy}/\hat{b}_{zx}$, where $\hat{b}_{zx}$ is the estimated effect size for the SNP–trait association for the exposure trait and $\hat{b}_{zy}$ is the effect size estimated for the same SNP in the GWAS of the outcome trait.

We used generalized summary statistics–based MR (GSMR)[64] to estimate $\hat{b}_{xy}$ and its standard error from multiple SNPs associated with the exposure trait at a genome-wide significance level. We conducted bidirectional GSMR analyses for each pair of traits and report results after excluding SNPs that failed the HEIDI outlier heterogeneity test (which is more conservative than excluding SNPs that have an outlying association likely driven by locus-specific pleiotropy). GSMR is more powerful than inverse-weighted MR (IVW-MR) and MR-Egger because it takes account of the sampling variation of both $\hat{b}_{zx}$ and $\hat{b}_{zy}$. GSMR also accounts for residual LD between the clumped SNPs. For comparison, we also conducted IVW-MR and MR-Egger analyses[124]. More details are provided in the Supplementary Note.

Genome build. All genomic coordinates are given in NCBI Build 37/UCSC hg19.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The PGC's policy is to make genome-wide summary results public. Summary statistics for a combined meta-analysis of PGC29 with five of the six expanded samples (deCODE, Generation Scotland, GERA, iPSYCH, and UK Biobank) are available on the PGC website (see URLs). Results for 10,000 SNPs for all seven cohorts are also available on the PGC website.

Genome-wide association summary statistics for the Hyde et al. cohort (23andMe, Inc.) must be obtained separately. These can be obtained by qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Contact David Hinds (dhinds@23andme.com) to apply for access to the data. Researchers who have the 23andMe summary statistics can readily recreate our results by performing meta-analysis of the six-cohort results file with the Hyde et al. results file from 23andMe[28].

The availability of genotype data for PGC29 is described in Supplementary Table 15. For the expanded cohorts, interested users should contact the lead principal investigators of these cohorts (which are separate from the PGC).

References

97. World Health Organization. *International Classification of Diseases* (World Health Organization, Geneva, 1978).
98. World Health Organization. *International Classification of Diseases* (World Health Organization, Geneva, 1992).
99. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, Washington, DC, 1994).
100. Abecasis, G. R. et al. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
101. ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
102. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
103. Vernot, B. et al. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352**, 235–239 (2016).
104. Bryois, J. et al. Evaluation of chromatin accessibility in prefrontal cortex of schizophrenia cases and controls. *Nat. Commun.* (in the press).
105. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
106. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
107. Finucane, H. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* https://dx.doi.org/10.1038/s41588-018-0081-4 (2018).
108. Zhernakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2017).
109. Jansen, R. et al. Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum. Mol. Genet.* **26**, 1444–1451 (2017).
110. Lappalainen, T. et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
111. Ng, B. et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017).
112. Hannon, E., Weedon, M., Bray, N., O'Donovan, M. & Mill, J. Pleiotropic effects of trait-associated genetic variation on DNA methylation: utility for refining GWAS loci. *Am. J. Hum. Genet.* **100**, 954–959 (2017).
113. de Leeuw, C. A., Neale, B. M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nat. Rev. Genet.* **17**, 353–364 (2016).
114. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
115. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
116. Turner, T. N. et al. denovo-db: a compendium of human de novo variants. *Nucleic Acids Res.* **45** (D1), D804–D811 (2017).
117. Pirooznia, M. et al. High-throughput sequencing of the synaptome in major depressive disorder. *Mol. Psychiatry* **21**, 650–655 (2016).
118. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
119. Wagner, A. H. et al. DGIdb 2.0: mining clinically relevant drug–gene interactions. *Nucleic Acids Res.* **44** (D1), D1036–D1044 (2016).
120. Roth, B. L., Kroeze, W. K., Patel, S. & Lopez, E. The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrasment of riches? *Neuroscientist* **6**, 252–262 (2000).
121. Olier, I., Vellido, A. & Giraldo, J. Kernel generative topographic mapping. in *ESANN 2010 Proc.* 28–30 (2010).
122. Smith, G. D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
123. Wooldridge, J. M. *Introductory Econometrics: A Modern Approach* (Cengage Learning, Boston, MA, 2015).
124. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).

nature genetics

ARTICLES

https://doi.org/10.1038/s41588-021-00857-4

Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology

Bipolar disorder is a heritable mental illness with complex etiology. We performed a genome-wide association study of 41,917 bipolar disorder cases and 371,549 controls of European ancestry, which identified 64 associated genomic loci. Bipolar disorder risk alleles were enriched in genes in synaptic signaling pathways and brain-expressed genes, particularly those with high specificity of expression in neurons of the prefrontal cortex and hippocampus. Significant signal enrichment was found in genes encoding targets of antipsychotics, calcium channel blockers, antiepileptics and anesthetics. Integrating expression quantitative trait locus data implicated 15 genes robustly linked to bipolar disorder via gene expression, encoding druggable targets such as HTR6, MCHR1, DCLK3 and FURIN. Analyses of bipolar disorder subtypes indicated high but imperfect genetic correlation between bipolar disorder type I and II and identified additional associated loci. Together, these results advance our understanding of the biological etiology of bipolar disorder, identify novel therapeutic leads and prioritize genes for functional follow-up studies.

Bipolar disorder (BD) is a complex mental disorder characterized by recurrent episodes of (hypo)mania and depression. It is a common condition affecting an estimated 40 to 50 million people worldwide[1]. This, combined with the typical onset in young adulthood, an often chronic course and increased risk of suicide[2], makes BD a major public health concern and a major cause of global disability[1]. Clinically, BD is classified into two main subtypes: bipolar I disorder (BD I), in which manic episodes typically alternate with depressive episodes, and bipolar II disorder (BD II), characterized by the occurrence of at least one hypomanic and one depressive episode[3]. These subtypes have a lifetime prevalence of ~1% each in the population[4,5].

Family and molecular genetic studies provide convincing evidence that BD is a multifactorial disorder, with genetic and environmental factors contributing to its development[6]. On the basis of twin and family studies, the heritability of BD is estimated at 60–85%[7,8]. Genome-wide association studies (GWASs)[9–23] have led to valuable insights into the genetic etiology of BD. The largest such study has been conducted by the Psychiatric Genomics Consortium (PGC), in which genome-wide SNP data from 29,764 patients with BD and 169,118 controls were analyzed and 30 genome-wide significant loci were identified (PGC2)[24]. SNP-based heritability $\left(h^2_{\mathrm{SNP}}\right)$ estimation using the same data suggested that common genetic variants genome-wide explain ~20% of BD's phenotypic variance[24]. Polygenic risk scores (PRSs) generated from the results of this study explained ~4% of phenotypic variance in independent samples. Across the genome, genetic associations with BD converged on specific biological pathways including regulation of insulin secretion[25,26], retrograde endocannabinoid signaling[24], glutamate receptor signaling[27] and calcium channel activity[9].

Despite this considerable progress, only a fraction of the genetic etiology of BD has been identified, and the specific biological mechanisms underlying the development of the disorder are still unknown. In the present study, we report the results of the third GWAS meta-analysis of the PGC Bipolar Disorder Working Group, comprising 41,917 individuals with BD and 371,549 controls. These results confirm and expand on many previously reported findings, identify novel therapeutic leads and prioritize genes for functional follow-up studies[28,29]. Thus, our results further illuminate the biological etiology of BD.

Results

GWAS results. We conducted a GWAS meta-analysis of 57 BD cohorts collected in Europe, North America and Australia (Supplementary Table 1), totaling 41,917 individuals with BD (cases) and 371,549 controls of European descent (effective $n=101{,}962$, see Methods). For 52 cohorts, individual-level genotype and phenotype data were shared with the PGC and cases met international consensus criteria (DSM-IV, ICD-9 or ICD-10) for lifetime BD, established using structured diagnostic interviews, clinician-administered checklists or medical record review. BD GWAS summary statistics were received for five external cohorts (iPSYCH[30], deCODE genetics[31], Estonian Biobank[32], Trøndelag Health Study (HUNT)[33] and UK Biobank[34]), in which most cases were ascertained using ICD codes. The GWAS meta-analysis identified 64 independent loci associated with BD at genome-wide significance ($P<5\times10^{-8}$; Fig. 1, Table 1 and Supplementary Table 2). Using linkage disequilibrium score regression (LDSC)[35], the h^2_{SNP} of BD was estimated to be 18.6% (s.e.=0.008, $P=5.1\times10^{-132}$) on the liability scale, assuming a BD population prevalence of 2%, and 15.6% (s.e.=0.006, $P=5.0\times10^{-132}$) assuming a population prevalence of 1% (Supplementary Table 3). The genomic inflation factor (λ_{GC}) was 1.38 and the LDSC intercept was 1.04 (s.e.=0.01, $P=2.5\times10^{-4}$; Supplementary Fig. 1). While the intercept has frequently been used as an indicator of confounding from population stratification, it can rise above 1 with increased sample size and heritability. The attenuation ratio—(LDSC intercept−1)/(mean of association chi-square statistics−1)—which is not subject to these limitations, was 0.06 (s.e.=0.02), indicating that the majority of inflation of the GWAS test statistics was due to polygenicity[35,36]. Of the 64 genome-wide significant loci, 33 are novel discoveries (that is, loci not overlapping with any locus previously reported as genome-wide significant for BD). Novel loci include the major histocompatibility complex (MHC) and loci previously reaching genome-wide significance for other psychiatric disorders, including ten for schizophrenia, four for major depression and three for childhood-onset psychiatric disorders or problematic alcohol use (Table 1).

A full list of authors and their affiliations appears at the end of the paper.

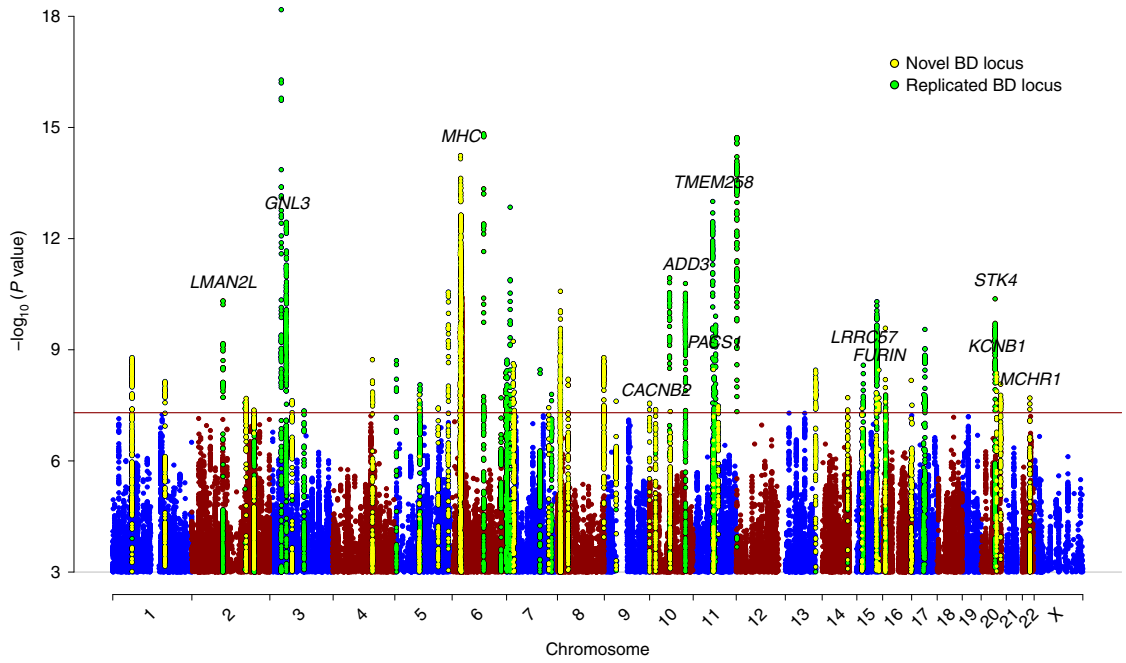


Fig. 1 | Manhattan plot of genome-wide association meta-analysis of 41,917 BD cases and 371,549 controls. The *x* axis shows genomic position (chromosomes 1–22 and X), and the *y* axis shows statistical significance as $-\log_{10}$[*P* value]. *P* values are two-sided and based on an inverse-variance-weighted fixed-effects meta-analysis. The red line shows the genome-wide significance threshold ($P < 5 \times 10^{-8}$). SNPs in genome-wide significant loci are colored green for loci previously associated with BD and yellow for novel associations from this study. The genes labeled are those prioritized by integrative eQTL analyses or notable genes in novel loci (MHC, *CACNB2* and *KCNB1*).

Enrichment analyses. Genome-wide analyses using MAGMA[37] indicated significant enrichment of BD associations in 161 genes (Supplementary Table 4) and 4 gene sets related to synaptic signaling (Supplementary Table 5). The BD association signal was enriched among genes expressed in different brain tissues (Supplementary Table 6), especially genes with high specificity of gene expression in neurons (both excitatory and inhibitory) versus other cell types, within cortical and subcortical brain regions in mice (Supplementary Fig. 2)[38]. In human brain samples, signal enrichment was also observed in hippocampal pyramidal neurons and interneurons of the prefrontal cortex and hippocampus, compared with other cell types (Supplementary Fig. 2).

In a gene-set analysis of the targets of individual drugs (from the Drug–Gene Interaction Database DGIdb v.2 (ref. [39]) and the Psychoactive Drug Screening Database Ki DB[40]), the targets of the calcium channel blockers mibefradil and nisoldipine were significantly enriched (Supplementary Table 7). Grouping drugs according to their anatomical therapeutic chemical classes[41], there was significant enrichment in the targets of four broad drug classes (Supplementary Table 8): psycholeptics (drugs with a calming effect on behavior; especially hypnotics and sedatives, antipsychotics and anxiolytics), calcium channel blockers, antiepileptics and (general) anesthetics (Supplementary Table 8).

Expression quantitative trait locus integrative analyses. We conducted a transcriptome-wide association study (TWAS) using FUSION[42] and expression quantitative trait locus (eQTL) data from the PsychENCODE Consortium (1,321 brain samples)[43]. BD-associated alleles significantly influenced expression of 77 genes in the brain (Supplementary Table 9 and Supplementary Fig. 3). These genes encompassed 40 distinct regions. We performed TWAS fine-mapping using FOCUS[44] to model the correlation among the TWAS signals and prioritize the most likely causal gene(s) in each region. Within the 90%-credible set, FOCUS prioritized 22 genes with a posterior inclusion probability (PIP) > 0.9 (encompassing 20 distinct regions) and 32 genes with a PIP > 0.7 (29 distinct regions; Supplementary Table 10).

We used summary-data-based Mendelian randomization (SMR)[45,46] to identify putative causal relationships between SNPs and BD via gene expression by integrating the BD GWAS results with brain eQTL summary statistics from the PsychENCODE[43] Consortium and blood eQTL summary statistics from the eQTLGen consortium (31,684 whole blood samples)[47]. The eQTLGen results represent the largest existing eQTL study and provide independent eQTL data. Of the 32 genes fine-mapped with PIP > 0.7, 15 were significantly associated with BD in the SMR analyses and passed the heterogeneity in dependent instruments (HEIDI) test[45,46], suggesting that their effect on BD is mediated via gene expression in the brain and/or blood (Supplementary Table 11). The genes located in genome-wide significant loci are labeled in Fig. 1. Other significant genes included *HTR6*, *DCLK3*, *HAPLN4* and *PACSIN2*.

MHC locus. Variants within and distal to the MHC locus were associated with BD at genome-wide significance. The most highly associated SNP was rs13195402, 3.2 megabases (Mb) distal to any *HLA* gene or the complement component 4 (*C4*) genes (Supplementary Fig. 4). Imputation of *C4* alleles using SNP data uncovered no association between the five most common structural forms of the *C4A*/*C4B* locus (BS, AL, AL–BS, AL–BL and AL–AL) and BD, either before or after conditioning on rs13195402 (Supplementary Fig. 5). While genetically predicted *C4A* expression initially showed a weak association with BD, this association was nonsignificant after controlling for rs13195402 (Supplementary Fig. 6).

Polygenic risk scoring. The performance of PRSs based on these GWAS results was assessed by excluding cohorts in turn from the meta-analysis to create independent test samples. PRSs explained ~4.57% of phenotypic variance in BD on the liability scale (at GWAS *P*-value threshold (GWAS P_T) < 0.1, BD population prevalence 2%), based on the weighted mean R^2 across cohorts (Fig. 2

Table 1 | Genome-wide significant loci for BD from meta-analysis of 41,917 cases and 371,549 controls

Locus	CHR	BP	SNP	*P*	OR	s.e.	A1/A2	A1 freq in controls	Previous report[a] for BD (citation)	Name for novel locus[b]	Previous report[a] for psychiatric disorders
1	1	61105668	rs2126180	1.6×10^{-9}	1.058	0.009	A/G	0.457		*LINC01748*	
2	1	163745389	rs10737496	7.2×10^{-9}	1.056	0.009	C/T	0.444		*NUF2*	CDG
3[c]	2	97416153	rs4619651	4.8×10^{-11}	1.068	0.010	G/A	0.670	*LMAN2L* (PGC2)		CDG
4	2	166152389	rs17183814	2.7×10^{-8}	1.108	0.019	G/A	0.924	*SCN2A* (PGC2)		
5	2	169481837	rs13417268	2.1×10^{-8}	1.064	0.011	C/G	0.758		*CERS6*	
6	2	193738336	rs2011302	4.3×10^{-8}	1.055	0.010	A/T	0.377		*PCGEM1*	CDG
7	2	194437889	rs2719164	4.9×10^{-8}	1.053	0.010	A/G	0.564	Intergenic (PGC2)		CDG
8[c]	3	36856030	rs9834970	6.6×10^{-19}	1.087	0.009	C/T	0.481	*TRANK1* (PGC2)		SCZ, CDG
9[c]	3	52626443	rs2336147	3.6×10^{-13}	1.070	0.009	T/C	0.498	*ITIH1* (PGC2)		SCZ, CDG
10	3	70488788	rs115694474	2.4×10^{-8}	1.068	0.012	T/A	0.799		*MDFIC2*	
11	3	107757060	rs696366	4.5×10^{-8}	1.053	0.009	C/A	0.550	*CD47* (PGC2)		
12[c]	4	123076007	rs112481526	1.9×10^{-9}	1.065	0.011	G/A	0.256		*KIAA1109*	MD
13[c]	5	7542911	rs28565152	2.0×10^{-9}	1.070	0.011	A/G	0.238	*ADCY2* (PGC2)		
14[c]	5	78849505	rs6865469	1.7×10^{-8}	1.060	0.010	T/G	0.274		*HOMER1*	
15	5	80961069	rs6887473	8.8×10^{-9}	1.062	0.011	G/A	0.739	*SSBP2* (PGC2)		
16[c]	5	137712121	rs10043984	3.7×10^{-8}	1.062	0.011	T/C	0.236		*KDM3B*	CDG
17	5	169289206	rs10866641	2.8×10^{-11}	1.065	0.009	T/C	0.575		*DOCK2*	
18[c]	6	26463575	rs13195402	5.8×10^{-15}	1.146	0.018	G/T	0.919		MHC	MD, SCZ, CDG, MOOD
19[c]	6	98565211	rs1487445	1.5×10^{-15}	1.078	0.009	T/C	0.487	*POU3F2* (PGC2)		CDG
20	6	152793572	rs4331993	2.0×10^{-8}	1.056	0.010	A/T	0.382	*SYNE1* (Green et al. 2013)		
21[c]	6	166995260	rs10455979	4.2×10^{-9}	1.057	0.010	G/C	0.500	*RPS6KA2* (PGC2)		
22[c]	7	2020995	rs12668848	1.9×10^{-9}	1.059	0.010	G/A	0.575	*MAD1L1* (Hou et al. 2016, Ikeda et al. 2018)		MD, SCZ, CDG
23[c]	7	11871787	rs113779084	1.4×10^{-13}	1.079	0.010	A/G	0.299	*THSD7A* (PGC2)		
24[c]	7	21492589	rs6954854	5.9×10^{-10}	1.060	0.009	G/A	0.425		*SP4*	
25	7	24647222	rs12672003	2.7×10^{-9}	1.096	0.016	G/A	0.113		*MPP6*	SCZ, CDG, MOOD
26	7	105043229	rs11764361	3.5×10^{-9}	1.063	0.010	A/G	0.668	*SRPK2* (PGC2)		SCZ, ASD, CDG
27	7	131870597	rs6946056	3.7×10^{-8}	1.055	0.010	C/A	0.623		*PLXNA4*	
28	7	140676153	rs10255167	1.6×10^{-8}	1.068	0.012	A/G	0.778	*MRPS33* (PGC2)		CDG
29[c]	8	9763581	rs62489493	2.6×10^{-11}	1.094	0.014	G/C	0.128		*miR124-1*	SCZ, ALC, ASD
30[c]	8	10226355	rs3088186	2.1×10^{-8}	1.058	0.010	T/C	0.287		*MSRA*	SCZ, ALC, ASD
31	8	34152492	rs2953928	6.3×10^{-9}	1.124	0.020	A/G	0.067		*RP1-84O15.2* (lincRNA)	SCZ, ADHD, CDG
32[c]	8	144993377	rs6992333	1.6×10^{-9}	1.062	0.010	G/A	0.410		*PLEC*	
33	9	37090538	rs10973201	2.5×10^{-8}	1.101	0.017	C/T	0.110		*ZCCHC7*	MD, CDG, MOOD
34[c]	9	141066490	rs62581014	2.8×10^{-8}	1.067	0.012	T/C	0.366		*TUBBP5*	
35[c]	10	18751103	rs1998820	4.1×10^{-8}	1.087	0.015	T/A	0.886		*CACNB2*	SCZ, CDG
36[c]	10	62322034	rs10994415	1.1×10^{-11}	1.125	0.017	C/T	0.082	*ANK3* (PGC2)		
37	10	64525135	rs10761661	4.7×10^{-8}	1.053	0.009	T/C	0.472		*ADO*	

Continued

Table 1 | Genome-wide significant loci for BD from meta-analysis of 41,917 cases and 371,549 controls (Continued)

Locus	CHR	BP	SNP	*P*	OR	s.e.	A1/A2	A1 freq in controls	Previous report[a] for BD (citation)	Name for novel locus[b]	Previous report[a] for psychiatric disorders
38[c]	10	111648659	rs2273738	1.6×10^{-11}	1.096	0.014	T/C	0.135	*ADD3* (Charney et al. 2017, PGC2)		
39[c]	11	61618608	rs174592	9.9×10^{-14}	1.074	0.010	G/A	0.395	*FADS2* (PGC2)		MD, CDG, MOOD
40	11	64009879	rs4672	3.4×10^{-9}	1.107	0.017	A/G	0.083		*FKBP2*	
41[c]	11	65848738	rs475805	2.0×10^{-9}	1.070	0.011	A/G	0.767	*PACS1* (PGC2)		
42[c]	11	66324583	rs678397	5.5×10^{-9}	1.056	0.009	T/C	0.457	*PC* (PGC1, PGC2)		
43[c]	11	70517927	rs12575685	1.2×10^{-10}	1.067	0.010	A/G	0.327	*SHANK2* (PGC2)		MD
44	11	79092527	rs12289486	3.3×10^{-8}	1.086	0.015	T/C	0.115	*ODZ4* (PGC1)		
45[c]	12	2348844	rs11062170	1.9×10^{-15}	1.081	0.010	C/G	0.333	*CACNA1C* (PGC2)		SCZ, CDG, MOOD
46	13	113869045	rs35306827	3.6×10^{-9}	1.068	0.011	G/A	0.775		*CUL4A*	
47	14	99719219	rs2693698	2.0×10^{-8}	1.055	0.009	G/A	0.551		*BCL11B*	SCZ, CDG
48[c]	15	38973793	rs35958438	3.8×10^{-8}	1.066	0.012	G/A	0.772		*C15orf53*	CDG
49[c]	15	42904904	rs4447398	2.6×10^{-9}	1.086	0.014	A/C	0.131	*STARD9* (PGC2)		
50	15	83531774	rs62011709	1.4×10^{-8}	1.064	0.011	T/A	0.747		*HOMER2*	SCZ
51[c]	15	85149575	rs748455	5.0×10^{-11}	1.070	0.010	T/C	0.719	*ZNF592* (PGC2)		SCZ, CDG
52	15	91426560	rs4702	3.5×10^{-9}	1.059	0.010	G/A	0.446		*FURIN*	SCZ, CDG
53	16	9230816	rs28455634	2.6×10^{-10}	1.065	0.010	G/A	0.620		*C16orf72*	
54	16	9926348	rs7199910	1.7×10^{-8}	1.057	0.010	G/T	0.312	*GRIN2A* (PGC2)		SCZ, CDG
55	16	89632725	rs12932628	6.7×10^{-9}	1.058	0.010	T/G	0.487		*RPL13*	
56	17	1835482	rs4790841	3.1×10^{-8}	1.075	0.013	T/C	0.151		*RTN4RL1*	
57	17	38129841	rs11870683	2.8×10^{-8}	1.059	0.010	T/A	0.650	*ERBB2* (Hou et al. 2016)		
58	17	38220432	rs61554907	1.6×10^{-8}	1.091	0.015	T/G	0.124	*ERBB2* (Hou et al. 2016)		
59[c]	17	42191893	rs228768	2.8×10^{-10}	1.067	0.010	G/T	0.294	*HDAC5* (PGC2)		
60[c]	20	43682551	rs67712855	4.2×10^{-11}	1.070	0.010	T/G	0.687	*STK4* (PGC2)		
61[c]	20	43944323	rs6032110	1.0×10^{-9}	1.059	0.009	A/G	0.512	*WFDC12* (PGC2)		
62[c]	20	48033127	rs237460	4.3×10^{-9}	1.057	0.009	T/C	0.412		*KCNB1*	CDG
63	20	60865815	rs13044225	8.5×10^{-9}	1.056	0.010	G/A	0.440		*OSBPL2*	
64	22	41153879	rs5758064	2.0×10^{-8}	1.054	0.009	T/C	0.523		*SLC25A17*	MD, SCZ, CDG, MOOD

[a]Previous report refers to previous association of a SNP in the locus with the psychiatric disorder at genome-wide significance. PGC1, ref. [9]; PGC2, ref. [24]; Hou et al. 2016, ref. [17]; Ikeda et al. 2018, ref. [59]; Green et al. 2013, ref. [15]; Charney et al. 2017, ref. [11]. [b]Novel loci are named using the nearest gene to the index SNP. [c]Locus overlaps with a genome-wide significant locus for BD I. CHR, chromosome; BP, GRCh37 base-pair position; OR, odds ratio; A1, tested allele; A2, other allele; freq, frequency; CDG, cross-disorder GWAS of the PGC; MD, major depression; SCZ, schizophrenia; MOOD, mood disorders; ASD, autism spectrum disorder; ALC, alcohol use disorder or problematic alcohol use; ADHD, attention deficit/hyperactivity disorder. *P* values are two-sided and based on an inverse-variance-weighted fixed-effects meta-analysis.

and Supplementary Table 12). This corresponds to a weighted mean area under the curve of 65%. Results per cohort and per wave of recruitment to the PGC are in Supplementary Tables 12 and 13 and Supplementary Fig. 7. At GWAS $P_T < 0.1$, individuals in the top 10% of BD PRSs had an odds ratio of 3.5 (95% confidence interval (CI) 1.7–7.3) of being affected with the disorder compared with individuals in the middle decile (based on the weighted mean OR across PGC cohorts), and an odds ratio of 9.3 (95% CI 1.7–49.3) compared with individuals in the lowest decile. The generalizability of PRSs from this meta-analysis was examined in several non-European cohorts. PRSs explained up to 2.3% and 1.9% of variance in BD in two East Asian samples, and 1.2% and 0.4% in two admixed African American samples (Fig. 2 and Supplementary Table 14). The variance explained by the PRSs increased in every cohort with increasing sample size of the PGC BD European discovery sample (Supplementary Fig. 8 and Supplementary Table 14).

Genetic architecture of BD and other traits. The genome-wide genetic correlation (r_g) of BD with a range of diseases and traits was assessed on LD Hub[48]. After correction for multiple testing, BD showed significant r_g with 16 traits among 255 tested from published GWASs (Supplementary Table 15). Genetic correlation was positive with all psychiatric disorders assessed, particularly schizophrenia ($r_g = 0.68$) and major depression ($r_g = 0.44$), and to a lesser degree anorexia, attention deficit/hyperactivity disorder and autism spectrum disorder ($r_g \approx 0.2$). We found evidence of positive r_g between BD

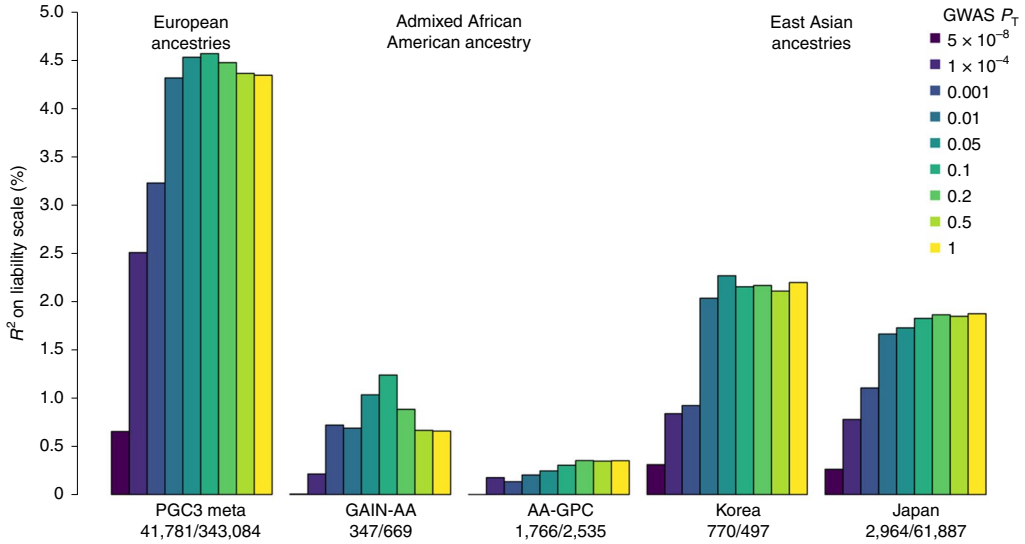


Fig. 2 | Phenotypic variance in BD explained by PRSs. Variance explained is presented on the liability scale, assuming a 2% population prevalence of BD. For European ancestries, the results shown are the weighted mean R^2 values across all 57 cohorts in the PGC3 meta-analysis, weighted by the effective *n* per cohort. The numbers of cases and controls are shown from left to right under the bar plot for each study. GWAS P_T, the color of the bars represents the *P*-value threshold used to select SNPs from the discovery GWAS; GAIN-AA, Genetic Association Information Network African American cohort; AA-GPC, African American Genomic Psychiatry Cohort.

and smoking initiation, cigarettes per day, problematic alcohol use and drinks per week (Fig. 3). BD was also positively genetically correlated with measures of sleep quality (daytime sleepiness, insomnia and sleep duration; Fig. 3). Among 514 traits measured in the general population of the UK Biobank, there was significant r_g between BD and many psychiatric-relevant traits or symptoms, dissatisfaction with interpersonal relationships, poorer overall health rating and feelings of loneliness or isolation (Supplementary Table 16).

Bivariate gaussian mixture models were applied to the GWAS summary statistics for BD and other complex traits using the MiXeR tool[49,50] to estimate the number of variants influencing each trait that explain 90% of h^2_{SNP} and their overlap between traits. MiXeR estimated that approximately 8,600 (s.e. = 200) variants influence BD, which is similar to the estimate for schizophrenia (9,700, s.e. = 200) and lower than that for major depression (12,300, s.e. = 600; Supplementary Table 17 and Supplementary Fig. 9). When considering the number of shared loci as a proportion of the total polygenicity of each trait, the vast majority of loci influencing BD were also estimated to influence major depression (97%) and schizophrenia (96%; Supplementary Table 17 and Supplementary Fig. 9). Interestingly, within these shared components, the variants that influenced both BD and schizophrenia had high concordance in direction of effect (80%, s.e. = 2%), while the portion of concordant variants between BD and major depressive disorder was only 69% (s.e. = 1%; Supplementary Table 17).

Genetic and causal relationships between BD and modifiable risk factors. Ten traits associated with BD from clinical and epidemiological studies were investigated in detail for genetic and potentially causal relationships with BD via LDSC[35], generalized summary statistics–based Mendelian randomization (GSMR)[51] and bivariate gaussian mixture modeling[49]. BD has been strongly linked with sleep disturbances[52], alcohol use[53], smoking[54], higher educational attainment[55,56] and mood instability[57]. Most of these traits had modest but significant genetic correlations with BD ($r_g = -0.05$ to 0.35; Fig. 3). Examining the effects of these traits on BD via GSMR, smoking initiation was associated with BD, corresponding to an OR of 1.49 (95% CI 1.38–1.61) for developing the disorder ($P = 1.74 \times 10^{-22}$; Fig. 3). Testing the effect of BD on the traits, we found that BD was significantly associated with reduced likelihood of being a morning person and increased number of drinks per week ($P < 1.47 \times 10^{-3}$; Fig. 3). Positive bidirectional relationships were identified between BD and longer sleep duration, problematic alcohol use, educational attainment and mood instability (Fig. 3). Notably, the instrumental variables for mood instability were selected from a GWAS conducted in the general population, excluding individuals with psychiatric disorders[58]. For all of the aforementioned BD–trait relationships, the effect size estimates from GSMR were consistent with those calculated using the inverse-variance-weighted regression method, and there was no evidence of bias from horizontal pleiotropy. Full MR results are in Supplementary Tables 18 and 19. Bivariate gaussian mixture modeling using MiXeR indicated large proportions of variants influencing both BD and all other traits tested, particularly educational attainment, where approximately 98% of variants influencing BD were estimated to also influence educational attainment. While cigarettes per day was a trait of interest, MiXeR could not model these data due to low polygenicity and heritability, and the effect of cigarettes per day on BD was inconsistent between MR methods, suggesting a violation of MR assumptions (Supplementary Tables 18–20).

BD subtypes. We conducted GWAS meta-analyses of BD I (25,060 cases, 449,978 controls) and BD II (6,781 cases, 364,075 controls). The BD I analysis identified 44 genome-wide significant loci, 31 of which overlapped with genome-wide significant loci from the main BD GWAS (Table 1 and Supplementary Table 21). The remaining 13 genome-wide significant loci for BD I all had $P < 4.0 \times 10^{-5}$ in the main BD GWAS. One genome-wide significant locus was identified in the GWAS meta-analysis of BD II and had a $P < 1.1 \times 10^{-4}$ in the main GWAS of BD (Supplementary Table 21). The h^2_{SNP} estimates on the liability scale for BD I and BD II were 20.9% (s.e. = 0.009, $P = 1.0 \times 10^{-111}$) and 11.6% (s.e. = 0.01, $P = 3.9 \times 10^{-15}$), respectively, assuming a 1% population prevalence of each subtype. These heritability values are significantly different from each other ($P = 2.4 \times 10^{-25}$, block jackknife). The genetic correlation between BD I and BD II was 0.85

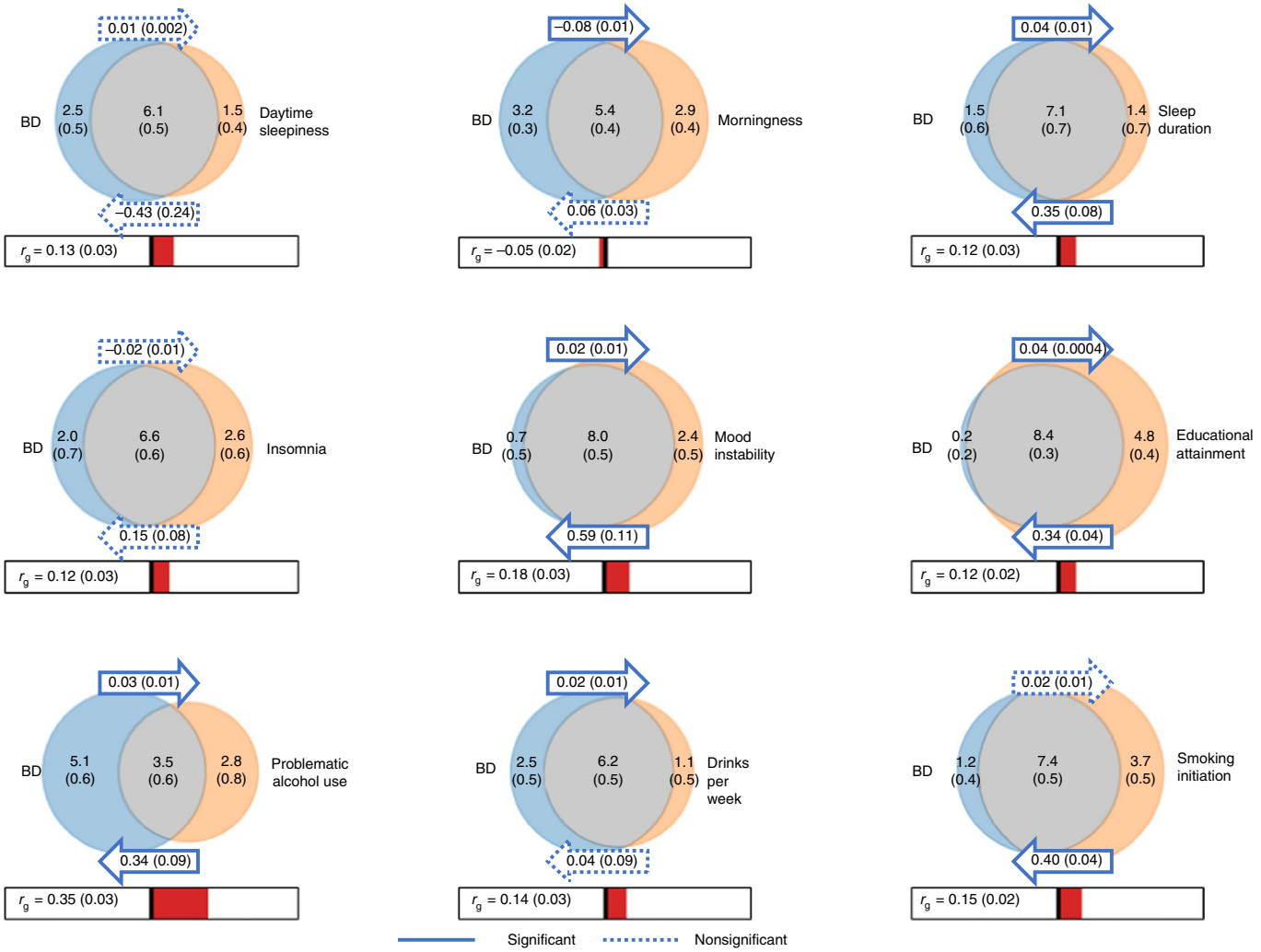


Fig. 3 | Relationships between BD and modifiable risk factors based on genetic correlations, GSMR and bivariate gaussian mixture modeling. The Venn diagrams depict MiXeR results of the estimated number of influencing variants shared between BD and each trait of interest (gray), unique to BD (blue) and unique to the trait of interest (orange). The number of influencing variants and standard error are shown in thousands. The size of the circles reflects the polygenicity of each trait, with larger circles corresponding to greater polygenicity. The estimated genetic correlation (r_g) between BD and each trait of interest and standard error from LDSC is shown below the corresponding Venn diagram, with an accompanying scale (−1 to +1). The arrows above and below the Venn diagrams indicate the results of GSMR of BD on the trait of interest, and the trait of interest on BD, respectively. The GSMR effect size and standard error is shown inside the corresponding arrow. Solid arrows indicate a significant relationship between the exposure and the outcome, after correction for multiple comparisons ($P < 1.47 \times 10^{-3}$), and dotted arrows indicate a nonsignificant relationship.

(s.e. = 0.05, $P = 2.88 \times 10^{-54}$), which is significantly different from 1 ($P = 1.6 \times 10^{-3}$). The genetic correlation of BD I with schizophrenia ($r_g = 0.66$, s.e. = 0.02) was higher than that of BD II ($r_g = 0.54$, s.e. = 0.05), whereas major depression was more strongly genetically correlated with BD II ($r_g = 0.66$, s.e. = 0.05) than with BD I ($r_g = 0.34$, s.e. = 0.03; Supplementary Table 22).

Discussion

In a GWAS of 41,917 BD cases, we identified 64 associated genomic loci, 33 of which are novel discoveries. With a 1.5-fold increase in effective sample size compared with the PGC2 BD GWAS, this study more than doubled the number of associated loci, representing an inflection point in the rate of risk variant discovery. We observed consistent replication of known BD loci, including 28/30 loci from the PGC2 GWAS[24] and several implicated by other BD GWAS[15–17], including a study of East Asian cases[59].

The 33 novel loci discovered here encompass genes of expected biological relevance to BD, such as the ion-channel-encoding genes *CACNB2* and *KCNB1*. Among the 64 BD loci, 17 have previously been implicated in GWAS of schizophrenia[60], and 7 in GWAS of major depression[61], representing the first overlap of genome-wide significant loci between the mood disorders. For these genome-wide significant loci shared across disorders, 17/17 and 5/7 of the BD index SNPs had the same direction of effect on schizophrenia and major depression, respectively (Supplementary Table 23). More generally, 50/64 and 62/64 BD loci had a consistent direction of effect on major depression and schizophrenia, respectively, considerably greater than chance ($P < 1 \times 10^{-5}$, binomial test). Bivariate gaussian mixture modeling estimated that across the entire genome, almost all variants influencing BD also influence schizophrenia and major depression, albeit with variable effects[62]. SNPs in and around the MHC locus reached genome-wide significance for BD for the first time. However, unlike in schizophrenia, we found no influence of *C4* structural alleles or gene expression[63]. Rather, the association was driven by variation outside the classical MHC locus, with the index SNP (rs13195402) being a missense

variant in *BTN2A1*, a brain-expressed gene[64] encoding a plasma membrane protein.

The genetic correlation of BD with other psychiatric disorders was consistent with previous reports[65,66]. Our results also corroborate previous genetic and clinical evidence of associations between BD and sleep disturbances[67], problematic alcohol use[68] and smoking[69]. While the genome-wide genetic correlations with these traits were modest ($r_g = -0.05$ to 0.35), MiXeR estimated that, for all traits, more than 55% of trait-influencing variants also influence BD (Fig. 3). Taken together, these results point to shared biology as one possible explanation for the high prevalence of substance use in BD. However, excluding genetic variants associated with both traits, MR analyses suggested that smoking is also a putatively 'causal' risk factor for BD, while BD has no effect on smoking, consistent with a previous report[70]. (We use the word 'causal' with caution here as we consider MR an exploratory analysis to identify potentially modifiable risk factors that warrant more detailed investigations to understand their complex relationship with BD.) In contrast, MR indicated that BD had bidirectional 'causal' relationships with problematic alcohol use, longer sleep duration and mood instability. Insights into the relationship of such behavioral correlates with BD may have future impact on clinical decision-making in the prophylaxis or management of the disorder. Higher educational attainment has previously been associated with BD in epidemiological studies[55,56], while lower educational attainment has been associated with schizophrenia and major depression[71,72]. Here, educational attainment had a significant positive effect on risk of BD and vice versa. Interestingly, MiXeR estimated that almost all variants that influence BD also influence educational attainment. The substantial genetic overlap observed between BD and the other phenotypes suggests that many variants likely influence multiple phenotypes, which may be differentiated by phenotype-specific effect size distributions among the shared influencing variants.

The integration of eQTL data with our GWAS results yielded 15 high-confidence genes for which there was converging evidence that their association with BD is mediated via gene expression. Among these were *HTR6*, encoding a serotonin receptor targeted by antipsychotics and antidepressants[73], and *MCHR1* (melanin-concentrating hormone receptor 1), encoding a target of the antipsychotic haloperidol[73]. We note that, for both of these genes, their top eQTLs have opposite directions of effect on gene expression in the brain and blood, possibly playing a role in the tissue-specific gene regulation influencing BD[74]. BD was associated with decreased expression of *FURIN*, a gene with a neurodevelopmental role that has already been the subject of functional genomics experiments in neuronal cells following its association with schizophrenia in GWAS[75]. The top association in our GWAS was in the *TRANK1* locus on chromosome 3, which has previously been implicated in BD[12,18,59]. Although BD-associated SNPs in this locus are known to regulate *TRANK1* expression[76], our eQTL analyses support a stronger but correlated regulation of *DCLK3*, located 87 kilobases (kb) upstream of *TRANK1* (refs. [43,77]). Both *FURIN* and *DCLK3* also encode druggable proteins (although they are not targets for any current psychiatric medications)[73,78]. These eQTL results provide promising BD candidate genes for functional follow-up experiments[29]. While several of these are in genome-wide significant loci, many are not the closest gene to the index SNP, highlighting the value of probing underlying molecular mechanisms to prioritize the most likely causal genes in the loci.

GWAS signals were enriched in the gene targets of existing BD pharmacological agents, such as antipsychotics, mood stabilizers and antiepileptics. However, enrichment was also found in the targets of calcium channel blockers used to treat hypertension and GABA-receptor-targeting anesthetics (Supplementary Table 8). Calcium channel antagonists have long been investigated for the treatment of BD, without becoming an established therapeutic approach, and there is evidence that some antiepileptics have calcium-channel-inhibiting effects[79,80]. These results underscore the opportunity for repurposing some classes of drugs, particularly calcium channel antagonists, as potential BD treatments[81].

BD associations were enriched in gene sets involving neuronal parts and synaptic signaling. Neuronal and synaptic pathways have been described in cross-disorder GWAS of multiple psychiatric disorders including BD[82–84]. Dysregulation of such pathways has also been suggested by previous functional and animal studies[85]. Analysis of single-cell gene expression data revealed enrichment in genes with high specificity of gene expression in neurons (both excitatory and inhibitory) of many brain regions, in particular the cortex and hippocampus. These findings are similar to those reported in GWAS data of schizophrenia[86] and major depressive disorder[38].

PRSs for BD explained on average 4.57% of phenotypic variance (liability scale) across European cohorts, although this varied in different waves of the BD GWAS, ranging from 6.6% in the PGC1 cohorts to 2.9% in the external biobank studies (Supplementary Fig. 7 and Supplementary Table 12). These results are in line with the h^2_{SNP} of BD per wave, which ranged from 24.6% (s.e. = 0.01) in PGC1 to 11.9% (s.e. = 0.01) in external studies (Supplementary Table 3). Some variability in h^2_{SNP} estimates may arise from the inclusion of cases from population biobanks, who may have more heterogeneous clinical presentations or less severe illness than patients with BD ascertained via inpatient or outpatient psychiatric clinics. Across the waves of clinically ascertained samples within the PGC, h^2_{SNP} and the R^2 of PRSs also varied, likely reflecting clinical and genetic heterogeneity in the type of BD cases ascertained; the PGC1 cohorts consisted mostly of BD I cases[9], known to be the most heritable of the BD subtypes[11,24], while later waves included more individuals with BD II[24]. Overall, the h^2_{SNP} of BD calculated from the meta-analysis summary statistics was 18% on the liability scale, a decrease of ~2% compared with the PGC2 GWAS[24], which may be due to the addition of cohorts with lower h^2_{SNP} estimates and heterogeneity between cohorts (Supplementary Table 3). However, despite differences in h^2_{SNP} and R^2 of PRSs per wave, the genetic correlation of BD between all waves was high (weighted mean $r_g = 0.94$, s.e. = 0.03), supporting our rationale for combining cases with different BD subtypes or ascertainment to increase power for discovery of risk variants. In Europeans, individuals in the top 10% of PRSs had an OR of 3.5 for BD, compared with individuals with average PRSs (middle decile), which translates into a modest absolute lifetime risk of the disorder (7% based on PRSs alone). While PRSs are invaluable tools in research settings, the current BD PRSs lack sufficient power to separate individuals into clinically meaningful risk categories, and therefore have no clinical utility at present[87,88]. PRSs from this European BD meta-analysis yield higher R^2 values in diverse ancestry samples than PRSs based on any currently available BD GWAS within the same ancestry[59]. However, performance still greatly lags behind that in Europeans, with ~2% variance explained in East Asian samples and substantially less in admixed African American samples, likely due to differences in allele frequencies and LD structures, consistent with previous studies[89,90]. There is a pressing need for more and larger studies in other ancestry groups to ensure that any future clinical utility is broadly applicable. Exploiting the differences in LD structure between diverse ancestry samples will also assist in the fine-mapping of risk loci for BD.

Our analyses confirmed that BD is a highly polygenic disorder, with an estimated 8,600 variants explaining 90% of its h^2_{SNP}. Hence, many more SNPs than those identified here are expected to account for the common variant architecture underlying BD. This GWAS marks an inflection point in risk variant discovery, and we expect that, from this point forward, the addition of more samples will lead to a dramatic increase in genetic findings. Nevertheless, fewer genome-wide significant loci have been identified in BD than in a schizophrenia GWASs of comparable sample size[60]. This may be due to the clinical and genetic heterogeneity that exists in BD.

Our GWAS of subtypes BD I and BD II identified additional associated loci. Consistent with previous findings[24], our analysis showed that the two subtypes were highly but imperfectly genetically correlated ($r_g = 0.85$), and that BD I is more genetically correlated with schizophrenia, while BD II has stronger genetic correlation with major depression. The subtypes are sufficiently similar to justify joint analysis as BD, but are not identical in their genetic composition, and as such contribute to the genetic heterogeneity of BD[91]. We identified 13 loci passing genome-wide significance for BD I, and one for BD II, which did not reach significance in the main BD GWAS, further illustrating the partially differing genetic composition of the two subtypes. Understanding the shared and distinct genetic components of BD subtypes and symptoms requires detailed phenotyping efforts in large cohorts and is an important area for future psychiatric genetics research.

In summary, these new data advance our understanding of the biological etiology of BD and prioritize a set of candidate genes for functional follow-up experiments. Several lines of evidence converge on the involvement of calcium channel signaling, providing a promising avenue for future therapeutic development.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00857-4.

Received: 2 September 2020; Accepted: 25 March 2021;
Published online: 17 May 2021

References

1. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1211–1259 (2017).
2. Plans, L. et al. Association between completed suicide and bipolar disorder: a systematic review of the literature. *J. Affect. Disord.* **242**, 111–122 (2019).
3. American Psychiatric Association *Diagnostic and Statistical Manual of Mental Disorders* 5th edn (American Psychiatric Association Publishing, 2013).
4. Merikangas, K. R. et al. Lifetime and 12-month prevalence of bipolar spectrum disorder in the national comorbidity survey replication. *Arch. Gen. Psychiatry* **64**, 543–552 (2007).
5. Merikangas, K. R. et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Arch. Gen. Psychiatry* **68**, 241–251 (2011).
6. Craddock, N. & Sklar, P. Genetics of bipolar disorder. *Lancet* **381**, 1654–1662 (2013).
7. Song, J. et al. Bipolar disorder and its relation to major psychiatric disorders: a family-based study in the Swedish population. *Bipolar Disord.* **17**, 184–193 (2015).
8. Bienvenu, O. J., Davydow, D. S. & Kendler, K. S. Psychiatric 'diseases' versus behavioral disorders and degree of genetic influence. *Psychol. Med.* **41**, 33–40 (2011).
9. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near *ODZ4*. *Nat. Genet.* **43**, 977–983 (2011).
10. Baum, A. E. et al. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol. Psychiatry* **13**, 197–207 (2008).
11. Charney, A. W. et al. Evidence for genetic heterogeneity between clinical subtypes of bipolar disorder. *Transl. Psychiatry* **7**, e993 (2017).
12. Chen, D. T. et al. Genome-wide association study meta-analysis of European and Asian-ancestry samples identifies three novel loci associated with bipolar disorder. *Mol. Psychiatry* **18**, 195–205 (2013).
13. Cichon, S. et al. Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am. J. Hum. Genet.* **88**, 372–381 (2011).
14. Ferreira, M. A. R. et al. Collaborative genome-wide association analysis supports a role for *ANK3* and *CACNA1C* in bipolar disorder. *Nat. Genet.* **40**, 1056–1058 (2008).
15. Green, E. K. et al. Association at *SYNE1* in both bipolar disorder and recurrent major depression. *Mol. Psychiatry* **18**, 614–617 (2013).
16. Green, E. K. et al. Replication of bipolar disorder susceptibility alleles and identification of two novel genome-wide significant associations in a new bipolar disorder case–control sample. *Mol. Psychiatry* **18**, 1302–1307 (2013).
17. Hou, L. et al. Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. *Hum. Mol. Genet.* **25**, 3383–3394 (2016).
18. Mühleisen, T. W. et al. Genome-wide association study reveals two new risk loci for bipolar disorder. *Nat. Commun.* **5**, 3339 (2014).
19. Schulze, T. G. et al. Two variants in Ankyrin 3 (*ANK3*) are independent genetic risk factors for bipolar disorder. *Mol. Psychiatry* **14**, 487–491 (2009).
20. Scott, L. J. et al. Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proc. Natl Acad. Sci. USA* **106**, 7501–7506 (2009).
21. Sklar, P. et al. Whole-genome association study of bipolar disorder. *Mol. Psychiatry* **13**, 558–569 (2008).
22. Smith, E. N. et al. Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol. Psychiatry* **14**, 755–763 (2009).
23. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
24. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
25. Lee, S.-H., Zabolotny, J. M., Huang, H., Lee, H. & Kim, Y.-B. Insulin in the nervous system and the mind: functions in metabolism, memory, and mood. *Mol. Metab.* **5**, 589–601 (2016).
26. McIntyre, R. S. et al. A randomized, double-blind, controlled trial evaluating the effect of intranasal insulin on neurocognitive function in euthymic patients with bipolar disorder. *Bipolar Disord.* **14**, 697–706 (2012).
27. Nurnberger, J. I. Jr et al. Identification of pathways for bipolar disorder: a meta-analysis. *JAMA Psychiatry* **71**, 657–664 (2014).
28. Gordovez, F. J. A. & McMahon, F. J. The genetics of bipolar disorder. *Mol. Psychiatry* **25**, 544–559 (2020).
29. Zhang, C., Xiao, X., Li, T. & Li, M. Translational genomics and beyond in bipolar disorder. *Mol. Psychiatry* **26**, 186–202 (2021).
30. Pedersen, C. B. et al. The iPSYCH2012 case–cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **1**, 6–14 (2018).
31. Gudbjartsson, D. F. et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
32. Leitsalu, L. et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int. J. Epidemiol.* **44**, 1137–1147 (2015).
33. Krokstad, S. et al. Cohort Profile: the HUNT Study, Norway. *Int. J. Epidemiol.* **42**, 968–977 (2013).
34. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
35. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
36. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
37. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
38. Bryois, J. et al. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nat. Genet.* **52**, 482–493 (2020).
39. Wagner, A. H. et al. DGIdb 2.0: mining clinically relevant drug–gene interactions. *Nucleic Acids Res.* **44**, D1036–D1044 (2016).
40. Roth, B. L., Lopez, E., Patel, S. & Kroeze, W. K. The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrassment of riches? *Neuroscientist* **6**, 252–262 (2000).
41. Gaspar, H. A. & Breen, G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. *Sci. Rep.* **7**, 12460 (2017).
42. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
43. Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).
44. Mancuso, N. et al. Probabilistic fine-mapping of transcriptome-wide association studies. *Nat. Genet.* **51**, 675–682 (2019).
45. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
46. Wu, Y. et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat. Commun.* **9**, 918 (2018).

47. Võsa, U. et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. Preprint at *bioRxiv* https://doi.org/10.1101/447367 (2018).
48. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
49. Frei, O. et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nat. Commun.* **10**, 2417 (2019).
50. Holland, D. et al. Beyond SNP heritability: polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. *PLoS Genet.* **16**, e1008612 (2020).
51. Zhu, Z. et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018).
52. Steardo, L. Jr et al. Sleep disturbance in bipolar disorder: neuroglia and circadian rhythms. *Front. Psychiatry* **10**, 501 (2019).
53. Hunt, G. E., Malhi, G. S., Cleary, M., Lai, H. M. X. & Sitharthan, T. Prevalence of comorbid bipolar and substance use disorders in clinical settings, 1990–2015: systematic review and meta-analysis. *J. Affect. Disord.* **206**, 331–349 (2016).
54. Heffner, J. L., Strawn, J. R., DelBello, M. P., Strakowski, S. M. & Anthenelli, R. M. The co-occurrence of cigarette smoking and bipolar disorder: phenomenology and treatment considerations. *Bipolar Disord.* **13**, 439–453 (2011).
55. Vreeker, A. et al. High educational performance is a distinctive feature of bipolar disorder: a study on cognition in bipolar disorder, schizophrenia patients, relatives and controls. *Psychol. Med.* **46**, 807–818 (2016).
56. MacCabe, J. H. et al. Excellent school performance at age 16 and risk of adult bipolar disorder: national cohort study. *Br. J. Psychiatry* **196**, 109–115 (2010).
57. Broome, M. R., Saunders, K. E. A., Harrison, P. J. & Marwaha, S. Mood instability: significance, definition and measurement. *Br. J. Psychiatry* **207**, 283–285 (2015).
58. Ward, J. et al. The genomic basis of mood instability: identification of 46 loci in 363,705 UK Biobank participants, genetic correlation with psychiatric disorders, and association with gene expression and function. *Mol. Psychiatry* **25**, 3091–3099 (2020).
59. Ikeda, M. et al. A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. *Mol. Psychiatry* **23**, 639–647 (2018).
60. Pardiñas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
61. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
62. Smeland, O. B., Frei, O., Dale, A. M. & Andreassen, O. A. The polygenic architecture of schizophrenia – rethinking pathogenesis and nosology. *Nat. Rev. Neurol.* **16**, 366–379 (2020).
63. Sekar, A. et al. Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016).
64. GTEx Consortium et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
65. Brainstorm Consortium et al. Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757 (2018).
66. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell* **179**, 1469–1482 (2019).
67. Lewis, K. J. S. et al. Comparison of genetic liability for sleep traits among individuals with bipolar disorder I or II and control participants. *JAMA Psychiatry* **77**, 303–310 (2020).
68. Zhou, H. et al. Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. *Nat. Neurosci.* **23**, 809–818 (2020).
69. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
70. Vermeulen, J. M. et al. Smoking and the risk for bipolar disorder: evidence from a bidirectional Mendelian randomisation study. *Br. J. Psychiatry* **218**, 88–94 (2021).
71. Peyrot, W. J. et al. The association between lower educational attainment and depression owing to shared genetic effects? Results in ~25 000 subjects. *Mol. Psychiatry* **20**, 735–743 (2015).
72. Swanson, C. L. Jr, Gur, R. C., Bilker, W., Petty, R. G. & Gur, R. E. Premorbid educational attainment in schizophrenia: association with symptoms, functioning, and neurobehavioral measures. *Biol. Psychiatry* **44**, 739–747 (1998).
73. Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
74. Mizuno, A. & Okada, Y. Biological characterization of expression quantitative trait loci (eQTLs) showing tissue-specific opposite directional effects. *Eur. J. Hum. Genet.* **27**, 1745–1756 (2019).
75. Schrode, N. et al. Synergistic effects of common schizophrenia risk variants. *Nat. Genet.* **51**, 1475–1485 (2019).
76. Jiang, X. et al. Sodium valproate rescues expression of *TRANK1* in iPSC-derived neural cells that carry a genetic variant associated with serious mental illness. *Mol. Psychiatry* **24**, 613–624 (2019).
77. Huckins, L. M. et al. Transcriptomic imputation of bipolar disorder and bipolar subtypes reveals 29 novel associated genes. Preprint at *bioRxiv* https://doi.org/10.1101/222786 (2017).
78. Finan, C. et al. The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* **9**, eaag1166 (2017).
79. von Wegerer, J., Hesslinger, B., Berger, M. & Walden, J. A calcium antagonistic effect of the new antiepileptic drug lamotrigine. *Eur. Neuropsychopharmacol.* **7**, 77–81 (1997).
80. Cipriani, A. et al. A systematic review of calcium channel antagonists in bipolar disorder and some considerations for their future development. *Mol. Psychiatry* **21**, 1324–1332 (2016).
81. Harrison, P. J., Tunbridge, E. M., Dolphin, A. C. & Hall, J. Voltage-gated calcium channel blockers for psychiatric disorders: genomic reappraisal. *Br. J. Psychiatry* **216**, 250–253 (2020).
82. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18**, 199–209 (2015).
83. Forstner, A. J. et al. Identification of shared risk loci and pathways for bipolar disorder and schizophrenia. *PLoS ONE* **12**, e0171595 (2017).
84. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* **173**, 1705–1715 (2018).
85. Lee, Y., Zhang, Y., Kim, S. & Han, K. Excitatory and inhibitory synaptic dysfunction in mania: an emerging hypothesis from animal model studies. *Exp. Mol. Med.* **50**, 1–11 (2018).
86. Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).
87. Lewis, C. M. & Vassos, E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* **12**, 44 (2020).
88. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19**, 581–590 (2018).
89. Duncan, L. et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10**, 3328 (2019).
90. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
91. Coleman, J. R. I. et al. The genetics of the mood disorder spectrum: genome-wide association analyses of more than 185,000 cases and 439,000 controls. *Biol. Psychiatry* **88**, 169–184 (2020).

Niamh Mullins[1,2,235] ✉, Andreas J. Forstner[3,4,5,235], Kevin S. O'Connell[6,7], Brandon Coombes[8], Jonathan R. I. Coleman[9,10], Zhen Qiao[11], Thomas D. Als[12,13,14], Tim B. Bigdeli[15,16], Sigrid Børte[17,18,19], Julien Bryois[20], Alexander W. Charney[2], Ole Kristian Drange[21,22], Michael J. Gandal[23], Saskia P. Hagenaars[9,10], Masashi Ikeda[24], Nolan Kamitaki[25,26], Minsoo Kim[23], Kristi Krebs[27], Georgia Panagiotaropoulou[28], Brian M. Schilder[1,29,30,31], Laura G. Sloofman[1], Stacy Steinberg[32], Vassily Trubetskoy[28], Bendik S. Winsvold[19,33], Hong-Hee Won[34], Liliya Abramova[35], Kristina Adorjan[36,37], Esben Agerbo[14,38,39], Mariam Al Eissa[40], Diego Albani[41], Ney Alliey-Rodriguez[42,43],

Adebayo Anjorin[44], Verneri Antilla[45], Anastasia Antoniou[46], Swapnil Awasthi[28], Ji Hyun Baek[47], Marie Bækvad-Hansen[14,48], Nicholas Bass[40], Michael Bauer[49], Eva C. Beins[3], Sarah E. Bergen[20], Armin Birner[50], Carsten Bøcker Pedersen[14,38,39], Erlend Bøen[51], Marco P. Boks[52], Rosa Bosch[53,54,55,56], Murielle Brum[57], Ben M. Brumpton[19], Nathalie Brunkhorst-Kanaan[57], Monika Budde[36], Jonas Bybjerg-Grauholm[14,48], William Byerley[58], Murray Cairns[59], Miquel Casas[53,54,55,56], Pablo Cervantes[60], Toni-Kim Clarke[61], Cristiana Cruceanu[60,62], Alfredo Cuellar-Barboza[63,64], Julie Cunningham[65], David Curtis[66,67], Piotr M. Czerski[68], Anders M. Dale[69], Nina Dalkner[50], Friederike S. David[3], Franziska Degenhardt[3,70], Srdjan Djurovic[71,72], Amanda L. Dobbyn[1,2], Athanassios Douzenis[46], Torbjørn Elvsåshagen[18,73,74], Valentina Escott-Price[75], I. Nicol Ferrier[76], Alessia Fiorentino[40], Tatiana M. Foroud[77], Liz Forty[75], Josef Frank[78], Oleksandr Frei[6,18], Nelson B. Freimer[23,79], Louise Frisén[80], Katrin Gade[36,81], Julie Garnham[82], Joel Gelernter[83,84,85], Marianne Giørtz Pedersen[14,38,39], Ian R. Gizer[86], Scott D. Gordon[87], Katherine Gordon-Smith[88], Tiffany A. Greenwood[89], Jakob Grove[12,13,14,90], José Guzman-Parra[91], Kyooseob Ha[92], Magnus Haraldsson[93], Martin Hautzinger[94], Urs Heilbronner[36], Dennis Hellgren[20], Stefan Herms[3,95,96], Per Hoffmann[3,95,96], Peter A. Holmans[75], Laura Huckins[1,2], Stéphane Jamain[97,98], Jessica S. Johnson[1,2], Janos L. Kalman[36,37,99], Yoichiro Kamatani[100,101], James L. Kennedy[102,103,104,105], Sarah Kittel-Schneider[57,106], James A. Knowles[107,108], Manolis Kogevinas[109], Maria Koromina[110], Thorsten M. Kranz[57], Henry R. Kranzler[111,112], Michiaki Kubo[113], Ralph Kupka[114,115,116], Steven A. Kushner[117], Catharina Lavebratt[118,119], Jacob Lawrence[120], Markus Leber[121], Heon-Jeong Lee[122], Phil H. Lee[123], Shawn E. Levy[124], Catrin Lewis[75], Calwing Liao[125,126], Susanne Lucae[62], Martin Lundberg[118,119], Donald J. MacIntyre[127], Sigurdur H. Magnusson[32], Wolfgang Maier[128], Adam Maihofer[89], Dolores Malaspina[1,2], Eirini Maratou[129], Lina Martinsson[80], Manuel Mattheisen[12,13,14,106,130], Steven A. McCarroll[25,26], Nathaniel W. McGregor[131], Peter McGuffin[9], James D. McKay[132], Helena Medeiros[108], Sarah E. Medland[87], Vincent Millischer[118,119], Grant W. Montgomery[11], Jennifer L. Moran[25,133], Derek W. Morris[134], Thomas W. Mühleisen[4,95], Niamh O'Brien[40], Claire O'Donovan[82], Loes M. Olde Loohuis[23,79], Lilijana Oruc[135], Sergi Papiol[36,37], Antonio F. Pardiñas[75], Amy Perry[88], Andrea Pfennig[49], Evgenia Porichi[46], James B. Potash[136], Digby Quested[137,138], Towfique Raj[1,29,30,31], Mark H. Rapaport[139], J. Raymond DePaulo[136], Eline J. Regeer[140], John P. Rice[141], Fabio Rivas[91], Margarita Rivera[142,143], Julian Roth[106], Panos Roussos[1,2,29], Douglas M. Ruderfer[144], Cristina Sánchez-Mora[53,54,56,145], Eva C. Schulte[36,37], Fanny Senner[36,37], Sally Sharp[40], Paul D. Shilling[89], Engilbert Sigurdsson[93,146], Lea Sirignano[78], Claire Slaney[82], Olav B. Smeland[6,7], Daniel J. Smith[147], Janet L. Sobell[148], Christine Søholm Hansen[14,48], Maria Soler Artigas[53,54,56,145], Anne T. Spijker[149], Dan J. Stein[150], John S. Strauss[102], Beata Świątkowska[151], Chikashi Terao[101], Thorgeir E. Thorgeirsson[32], Claudio Toma[152,153,154], Paul Tooney[59], Evangelia-Eirini Tsermpini[110], Marquis P. Vawter[155], Helmut Vedder[156], James T. R. Walters[75], Stephanie H. Witt[78], Simon Xi[157], Wei Xu[158], Jessica Mei Kay Yang[75], Allan H. Young[159,160], Hannah Young[1], Peter P. Zandi[136], Hang Zhou[83,84], Lea Zillich[78], HUNT All-In Psychiatry*, Rolf Adolfsson[161], Ingrid Agartz[51,130,162], Martin Alda[82,163], Lars Alfredsson[164], Gulja Babadjanova[165], Lena Backlund[118,119], Bernhard T. Baune[166,167,168], Frank Bellivier[169,170], Susanne Bengesser[50], Wade H. Berrettini[171], Douglas H. R. Blackwood[61], Michael Boehnke[172], Anders D. Børglum[14,173,174], Gerome Breen[9,10], Vaughan J. Carr[175], Stanley Catts[176], Aiden Corvin[177], Nicholas Craddock[75], Udo Dannlowski[166], Dimitris Dikeos[178], Tõnu Esko[26,27,179,180], Bruno Etain[169,170], Panagiotis Ferentinos[9,46], Mark Frye[64], Janice M. Fullerton[152,153], Micha Gawlik[106], Elliot S. Gershon[42,181],

Fernando S. Goes[136], Melissa J. Green[152,175], Maria Grigoroiu-Serbanescu[182], Joanna Hauser[68], Frans Henskens[59], Jan Hillert[80], Kyung Sue Hong[47], David M. Hougaard[14,48], Christina M. Hultman[20], Kristian Hveem[19,183], Nakao Iwata[24], Assen V. Jablensky[184], Ian Jones[75], Lisa A. Jones[88], René S. Kahn[2,52], John R. Kelsoe[89], George Kirov[75], Mikael Landén[20,185], Marion Leboyer[97,98,186], Cathryn M. Lewis[9,10,187], Qingqin S. Li[188], Jolanta Lissowska[189], Christine Lochner[190], Carmel Loughland[59], Nicholas G. Martin[87,191], Carol A. Mathews[192], Fermin Mayoral[91], Susan L. McElroy[193], Andrew M. McIntosh[127,194], Francis J. McMahon[195], Ingrid Melle[6,196], Patricia Michie[59], Lili Milani[27], Philip B. Mitchell[175], Gunnar Morken[21,197], Ole Mors[14,198], Preben Bo Mortensen[12,14,38,39], Bryan Mowry[176], Bertram Müller-Myhsok[62,199,200], Richard M. Myers[124], Benjamin M. Neale[25,45,179], Caroline M. Nievergelt[89,201], Merete Nordentoft[14,202], Markus M. Nöthen[3], Michael C. O'Donovan[75], Ketil J. Oedegaard[203,204], Tomas Olsson[205], Michael J. Owen[75], Sara A. Paciga[206], Chris Pantelis[207], Carlos Pato[108], Michele T. Pato[108], George P. Patrinos[110,208,209], Roy H. Perlis[210,211], Danielle Posthuma[212,213], Josep Antoni Ramos-Quiroga[53,54,55,56], Andreas Reif[57], Eva Z. Reininghaus[50], Marta Ribasés[53,54,56,145], Marcella Rietschel[78], Stephan Ripke[25,28,45], Guy A. Rouleau[126,214], Takeo Saito[24], Ulrich Schall[59], Martin Schalling[118,119], Peter R. Schofield[152,153], Thomas G. Schulze[36,78,81,136,215], Laura J. Scott[172], Rodney J. Scott[59], Alessandro Serretti[216], Cynthia Shannon Weickert[152,175,217], Jordan W. Smoller[25,133,218], Hreinn Stefansson[32], Kari Stefansson[32,219], Eystein Stordal[220,221], Fabian Streit[78], Patrick F. Sullivan[20,222,223], Gustavo Turecki[224], Arne E. Vaaler[225], Eduard Vieta[226], John B. Vincent[102], Irwin D. Waldman[227], Thomas W. Weickert[152,175,217], Thomas Werge[14,228,229,230], Naomi R. Wray[11,231], John-Anker Zwart[18,19,33], Joanna M. Biernacka[8,64], John I. Nurnberger[232], Sven Cichon[3,4,95,96], Howard J. Edenberg[77,233], Eli A. Stahl[1,2,179,236], Andrew McQuillin[40,236], Arianna Di Florio[75,223,236], Roel A. Ophoff[23,79,117,234,236] and Ole A. Andreassen[6,7,236] ✉

[1]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [2]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [3]Institute of Human Genetics, University of Bonn, School of Medicine and University Hospital Bonn, Bonn, Germany. [4]Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Jülich, Germany. [5]Centre for Human Genetics, University of Marburg, Marburg, Germany. [6]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [7]NORMENT, University of Oslo, Oslo, Norway. [8]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA. [9]Social, Genetic and Developmental Psychiatry Centre, King's College London, London, UK. [10]NIHR Maudsley BRC, King's College London, London, UK. [11]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia. [12]iSEQ, Center for Integrative Sequencing, Aarhus University, Aarhus, Denmark. [13]Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark. [14]iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark. [15]Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, NY, USA. [16]VA NY Harbor Healthcare System, Brooklyn, NY, USA. [17]Research and Communication Unit for Musculoskeletal Health, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway. [18]Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [19]K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway. [20]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [21]Department of Mental Health, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. [22]Department of Østmarka, Division of Mental Health Care, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. [23]Department of Psychiatry and Biobehavioral Science, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [24]Department of Psychiatry, School of Medicine, Fujita Health University, Toyoake, Japan. [25]Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA. [26]Department of Genetics, Harvard Medical School, Boston, MA, USA. [27]Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia. [28]Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin, Berlin, Germany. [29]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [30]Ronald M. Loeb Center for Alzheimer's Disease, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [31]Estelle and Daniel Maggin Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [32]deCODE Genetics/Amgen, Reykjavik, Iceland. [33]Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway. [34]Samsung Advanced Institute for Health Sciences and Technology (SAIHST), Samsung Medical Center, Sungkyunkwan University, Seoul, South Korea. [35]Russian Academy of Medical Sciences, Mental Health Research Center, Moscow, Russian Federation. [36]Institute of Psychiatric Phenomics and Genomics (IPPG), University Hospital, LMU Munich, Munich, Germany. [37]Department of Psychiatry and Psychotherapy, University Hospital, LMU Munich, Munich, Germany. [38]National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark. [39]Centre for Integrated Register-Based Research, Aarhus University, Aarhus, Denmark. [40]Division of Psychiatry, University College London, London, UK. [41]Department of Neuroscience, Istituto Di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy. [42]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, USA. [43]Northwestern University, Chicago, IL, USA. [44]Psychiatry, Berkshire Healthcare NHS Foundation Trust, Bracknell, UK. [45]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [46]2nd Department of Psychiatry, Attikon General Hospital, National and Kapodistrian University of Athens, Athens, Greece. [47]Department of Psychiatry, Samsung Medical Center, School of Medicine, Sungkyunkwan University, Seoul, South Korea. [48]Center for Neonatal Screening, Department for Congenital

Disorders, Statens Serum Institut, Copenhagen, Denmark. [49]Department of Psychiatry and Psychotherapy, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. [50]Department of Psychiatry and Psychotherapeutic Medicine, Medical University of Graz, Graz, Austria. [51]Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, Norway. [52]Psychiatry, Brain Center UMC Utrecht, Utrecht, the Netherlands. [53]Instituto de Salud Carlos III, Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain. [54]Department of Psychiatry, Hospital Universitari Vall d´Hebron, Barcelona, Spain. [55]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain. [56]Psychiatric Genetics Unit, Group of Psychiatry Mental Health and Addictions, Vall d´Hebron Research Institut (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. [57]Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Frankfurt am Main, Germany. [58]Psychiatry, University of California San Francisco, San Francisco, CA, USA. [59]University of Newcastle, Newcastle, New South Wales, Australia. [60]Mood Disorders Program, Department of Psychiatry, McGill University Health Center, Montreal, Quebec, Canada. [61]Division of Psychiatry, University of Edinburgh, Edinburgh, UK. [62]Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. [63]Department of Psychiatry, Universidad Autonoma de Nuevo Leon, Monterrey, Mexico. [64]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, USA. [65]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. [66]Centre for Psychiatry, Queen Mary University of London, London, UK. [67]UCL Genetics Institute, University College London, London, UK. [68]Department of Psychiatry, Laboratory of Psychiatric Genetics, Poznan University of Medical Sciences, Poznan, Poland. [69]Center for Multimodal Imaging and Genetics, Departments of Neurosciences, Radiology, and Psychiatry, University of California, San Diego, CA, USA. [70]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Essen, University of Duisburg-Essen, Duisburg, Germany. [71]Department of Medical Genetics, Oslo University Hospital, Oslo, Norway. [72]NORMENT, Department of Clinical Science, University of Bergen, Bergen, Norway. [73]Department of Neurology, Oslo University Hospital, Oslo, Norway. [74]NORMENT, KG Jebsen Centre for Psychosis Research, Oslo University Hospital, Oslo, Norway. [75]Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK. [76]Academic Psychiatry, Newcastle University, Newcastle upon Tyne, UK. [77]Department of Medical and Molecular Genetics, Indiana University, Indianapolis, IN, USA. [78]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [79]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, Los Angeles, CA, USA. [80]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. [81]Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany. [82]Department of Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada. [83]Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA. [84]Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA. [85]Departments of Genetics and Neuroscience, Yale University School of Medicine, New Haven, CT, USA. [86]Department of Psychological Sciences, University of Missouri, Columbia, MO, USA. [87]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [88]Psychological Medicine, University of Worcester, Worcester, UK. [89]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. [90]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [91]Mental Health Department, University Regional Hospital, Biomedicine Institute (IBIMA), Málaga, Spain. [92]Department of Psychiatry, Seoul National University College of Medicine, Seoul, South Korea. [93]Landspitali University Hospital, Reykjavik, Iceland. [94]Department of Psychology, Eberhard Karls Universität Tübingen, Tübingen, Germany. [95]Department of Biomedicine, University of Basel, Basel, Switzerland. [96]Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland. [97]Neuropsychiatrie Translationnelle, Inserm U955, Créteil, France. [98]Faculté de Santé, Université Paris Est, Créteil, France. [99]International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, Germany. [100]Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. [101]Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [102]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada. [103]Neurogenetics Section, Centre for Addiction and Mental Health, Toronto, Ontario, Canada. [104]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada. [105]Institute of Medical Sciences, University of Toronto, Toronto, Ontario, Canada. [106]Department of Psychiatry, Psychosomatics and Psychotherapy, Center of Mental Health, University Hospital Würzburg, Würzburg, Germany. [107]Cell Biology, SUNY Downstate Medical Center College of Medicine, Brooklyn, NY, USA. [108]Institute for Genomic Health, SUNY Downstate Medical Center College of Medicine, Brooklyn, NY, USA. [109]ISGlobal, Barcelona, Spain. [110]Laboratory of Pharmacogenomics and Individualized Therapy, Department of Pharmacy, School of Health Sciences, University of Patras, Patras, Greece. [111]Mental Illness Research, Education and Clinical Center, Crescenz VAMC, Philadelphia, PA, USA. [112]Center for Studies of Addiction, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [113]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [114]Psychiatry, Altrecht, Utrecht, the Netherlands. [115]Psychiatry, GGZ inGeest, Amsterdam, the Netherlands. [116]Psychiatry, VU Medisch Centrum, Amsterdam, the Netherlands. [117]Department of Psychiatry, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. [118]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. [119]Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden. [120]Psychiatry, North East London NHS Foundation Trust, Ilford, UK. [121]Clinic for Psychiatry and Psychotherapy, University Hospital Cologne, Cologne, Germany. [122]Department of Psychiatry, Korea University College of Medicine, Seoul, South Korea. [123]Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [124]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA. [125]Department of Human Genetics, McGill University, Montréal, Quebec, Canada. [126]Montreal Neurological Institute and Hospital, McGill University, Montréal, Quebec, Canada. [127]Division of Psychiatry, Centre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, UK. [128]Department of Psychiatry and Psychotherapy, University of Bonn, Bonn, Germany. [129]Clinical Biochemistry Laboratory, Attikon General Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece. [130]Department of Clinical Neuroscience, Centre for Psychiatry Research, Karolinska Institutet, Stockholm, Sweden. [131]Systems Genetics Working Group, Department of Genetics, Stellenbosch University, Stellenbosch, South Africa. [132]Genetic Cancer Susceptibility Group, International Agency for Research on Cancer, Lyon, France. [133]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [134]Centre for Neuroimaging and Cognitive Genomics (NICOG), National University of Ireland Galway, Galway, Ireland. [135]Medical Faculty, School of Science and Technology, University Sarajevo, Sarajevo, Bosnia and Herzegovina. [136]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [137]Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK. [138]Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK. [139]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA. [140]Outpatient Clinic for Bipolar Disorder, Altrecht, Utrecht, the Netherlands. [141]Department of Psychiatry, Washington University in Saint Louis, Saint Louis, MO, USA. [142]Department of Biochemistry and Molecular Biology II, Faculty of Pharmacy, University of Granada, Granada, Spain. [143]Institute of Neurosciences, Biomedical Research Center (CIBM), University of Granada, Granada, Spain. [144]Medicine, Psychiatry, Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. [145]Department of Genetics, Microbiology and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain. [146]Faculty of Medicine, Department of Psychiatry, School of Health Sciences, University of Iceland, Reykjavik, Iceland. [147]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK. [148]Psychiatry and the Behavioral Sciences, University of Southern California, Los Angeles, CA, USA. [149]Mood Disorders, PsyQ, Rotterdam, the Netherlands. [150]SAMRC Unit on Risk and Resilience in Mental Disorders, Department of Psychiatry and Neuroscience Institute, University of Cape Town, Cape Town, South Africa. [151]Department of Environmental Epidemiology, Nofer Institute of Occupational Medicine, Lodz, Poland. [152]Neuroscience Research Australia, Sydney, New South Wales, Australia. [153]School of Medical Sciences, University of New South Wales, Sydney, New South Wales, Australia. [154]Centro de Biología Molecular Severo Ochoa, Universidad Autónoma de Madrid and CSIC, Madrid, Spain.

[155]Department of Psychiatry and Human Behavior, School of Medicine, University of California, Irvine, Irvine, CA, USA. [156]Psychiatry, Psychiatrisches Zentrum Nordbaden, Wiesloch, Germany. [157]Computational Sciences Center of Emphasis, Pfizer Global Research and Development, Cambridge, MA, USA. [158]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. [159]Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. [160]South London and Maudsley NHS Foundation Trust, Bethlem Royal Hospital, Beckenham, UK. [161]Department of Clinical Sciences, Psychiatry, Umeå University Medical Faculty, Umeå, Sweden. [162]NORMENT, KG Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, Institute of Clinical Medicine and Diakonhjemmet Hospital, University of Oslo, Oslo, Norway. [163]National Institute of Mental Health, Klecany, Czech Republic. [164]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [165]Institute of Pulmonology, Russian State Medical University, Moscow, Russian Federation. [166]Department of Psychiatry, University of Münster, Münster, Germany. [167]Department of Psychiatry, Melbourne Medical School, The University of Melbourne, Melbourne, Victoria, Australia. [168]The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, Victoria, Australia. [169]Université de Paris, INSERM, Optimisation Thérapeutique en Neuropsychopharmacologie, UMRS 1144, Paris, France. [170]APHP Nord, DMU Neurosciences, Département de Psychiatrie et de Médecine Addictologique, GHU Saint Louis-Lariboisière-Fernand Widal, Paris, France. [171]Psychiatry, University of Pennsylvania, Philadelphia, PA, USA. [172]Center for Statistical Genetics and Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. [173]Department of Biomedicine and the iSEQ Center, Aarhus University, Aarhus, Denmark. [174]Center for Genomics and Personalized Medicine, CGPM, Aarhus, Denmark. [175]School of Psychiatry, University of New South Wales, Sydney, New South Wales, Australia. [176]University of Queensland, Brisbane, Queensland, Australia. [177]Neuropsychiatric Genetics Research Group, Department of Psychiatry and Trinity Translational Medicine Institute, Trinity College Dublin, Dublin, Ireland. [178]1st Department of Psychiatry, Eginition Hospital, National and Kapodistrian University of Athens, Athens, Greece. [179]Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [180]Division of Endocrinology, Children's Hospital Boston, Boston, MA, USA. [181]Department of Human Genetics, University of Chicago, Chicago, IL, USA. [182]Biometric Psychiatric Genetics Research Unit, Alexandru Obregia Clinical Psychiatric Hospital, Bucharest, Romania. [183]HUNT Research Center, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway. [184]University of Western Australia, Nedlands, Western Australia, Australia. [185]Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden. [186]Department of Psychiatry and Addiction Medicine, Assistance Publique - Hôpitaux de Paris, Paris, France. [187]Department of Medical and Molecular Genetics, King's College London, London, UK. [188]Neuroscience Therapeutic Area, Janssen Research and Development, LLC, Titusville, NJ, USA. [189]Cancer Epidemiology and Prevention, M. Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland. [190]SA MRC Unit on Risk and Resilience in Mental Disorders, Department of Psychiatry, Stellenbosch University, Stellenbosch, South Africa. [191]School of Psychology, The University of Queensland, Brisbane, Queensland, Australia. [192]Department of Psychiatry and Genetics Institute, University of Florida, Gainesville, FL, USA. [193]Research Institute, Lindner Center of HOPE, Mason, OH, USA. [194]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [195]Human Genetics Branch, Intramural Research Program, National Institute of Mental Health, Bethesda, MD, USA. [196]Division of Mental Health and Addiction, University of Oslo, Institute of Clinical Medicine, Oslo, Norway. [197]Psychiatry, St Olavs University Hospital, Trondheim, Norway. [198]Psychosis Research Unit, Aarhus University Hospital - Psychiatry, Risskov, Denmark. [199]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [200]University of Liverpool, Liverpool, UK. [201]Research/Psychiatry, Veterans Affairs San Diego Healthcare System, San Diego, CA, USA. [202]Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark. [203]Division of Psychiatry, Haukeland Universitetssjukehus, Bergen, Norway. [204]Faculty of Medicine and Dentistry, University of Bergen, Bergen, Norway. [205]Department of Clinical Neuroscience and Center for Molecular Medicine, Karolinska Institutet at Karolinska University Hospital, Solna, Sweden. [206]Human Genetics and Computational Biomedicine, Pfizer Global Research and Development, Groton, CT, USA. [207]University of Melbourne, Melbourne, Victoria, Australia. [208]Department of Pathology, College of Medicine and Health Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates. [209]Zayed Center of Health Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates. [210]Psychiatry, Harvard Medical School, Boston, MA, USA. [211]Division of Clinical Research, Massachusetts General Hospital, Boston, MA, USA. [212]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [213]Department of Clinical Genetics, Amsterdam Neuroscience, Vrije Universiteit Medical Center, Amsterdam, the Netherlands. [214]Department of Neurology and Neurosurgery, Faculty of Medicine, McGill University, Montreal, Quebec, Canada. [215]Department of Psychiatry and Behavioral Sciences, SUNY Upstate Medical University, Syracuse, NY, USA. [216]Department of Biomedical and NeuroMotor Sciences, University of Bologna, Bologna, Italy. [217]Department of Neuroscience, SUNY Upstate Medical University, Syracuse, NY, USA. [218]Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Massachusetts General Hospital, Boston, MA, USA. [219]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [220]Department of Psychiatry, Hospital Namsos, Namsos, Norway. [221]Department of Neuroscience, Norges Teknisk Naturvitenskapelige Universitet Fakultet for Naturvitenskap og Teknologi, Trondheim, Norway. [222]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [223]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [224]Department of Psychiatry, McGill University, Montreal, Quebec, Canada. [225]Department of Psychiatry, Sankt Olavs Hospital Universitetssykehuset i Trondheim, Trondheim, Norway. [226]Clinical Institute of Neuroscience, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERSAM, Barcelona, Spain. [227]Department of Psychology, Emory University, Atlanta, GA, USA. [228]Institute of Biological Psychiatry, Mental Health Services, Copenhagen University Hospital, Copenhagen, Denmark. [229]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [230]Center for GeoGenetics, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark. [231]Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia. [232]Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA. [233]Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA. [234]Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. [235]These authors contributed equally: Niamh Mullins, Andreas J. Forstner. [236]These authors jointly supervised this work: Eli A. Stahl, Andrew McQuillin, Arianna Di Florio, Roel A. Ophoff, Ole A. Andreassen. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: niamh.mullins@mssm.edu; ole.andreassen@medisin.uio.no

HUNT All-In Psychiatry

Bendik S. Winsvold[19,33], Eystein Stordal[220,221], Gunnar Morken[21,197], John-Anker Zwart[18,19,33], Ole Kristian Drange[21,22] and Sigrid Børte[17,18,19]

A list of members and their affiliations appears in the Supplementary Information.

Methods

Sample description. The meta-analysis sample comprises 57 cohorts collected in Europe, North America and Australia, totaling 41,917 BD cases and 371,549 controls of European descent (Supplementary Table 1). The total effective *n*, equivalent to an equal number of cases and controls in each cohort ($4\times n_{cases}\times n_{controls}/(n_{cases}+n_{controls})$), is 101,962. For 52 cohorts, individual-level genotype and phenotype data were shared with the PGC. Cohorts have been added to the PGC in five waves (PGC1 (ref. [9]), PGC2 (ref. [24]), PGC PsychChip, PGC3 and External studies); all cohorts from previous PGC BD GWASs were included. The source and inclusion/exclusion criteria for cases and controls for each cohort are described in the Supplementary Note. Cases were required to meet international consensus criteria (DSM-IV, ICD-9 or ICD-10) for a lifetime diagnosis of BD, established using structured diagnostic instruments from assessments by trained interviewers, clinician-administered checklists or medical record review. In most cohorts, controls were screened for the absence of lifetime psychiatric disorders and randomly selected from the population. For five cohorts (iPSYCH[30], deCODE genetics[31], Estonian Biobank[32], HUNT[33] and UK Biobank[34]), GWAS summary statistics for BD were shared with the PGC. In these cohorts, BD cases were ascertained using ICD codes or self-report during a nurse interview, and the majority of controls were screened for the absence of psychiatric disorders via ICD codes. Follow-up analyses included four non-European BD case–control cohorts, two from East Asia (Japan[59] and Korea[92]) and two admixed African American cohorts[22,93], providing a total of 5,847 cases and 65,588 controls. These BD cases were ascertained using international consensus criteria (DSM-IV)[22,93] through psychiatric interviews (Supplementary Note).

Genotyping, quality control and imputation. For 52 cohorts internal to the PGC, genotyping was performed following local protocols and genotypes were called using standard genotype calling software from commercial sources (Affymetrix and Illumina). Subsequently, standardized quality control, imputation and statistical analyses were performed centrally using Rapid Imputation for Consortias Pipeline (RICOPILI; version 2018_Nov_23.001)[94], separately for each cohort. Briefly, the quality control parameters for retaining SNPs and subjects were: SNP missingness < 0.05 (before sample removal), subject missingness < 0.02, autosomal heterozygosity deviation (F_{het} < 0.2), SNP missingness < 0.02 (after sample removal), difference in SNP missingness between cases and controls < 0.02, SNP Hardy–Weinberg equilibrium ($P > 10\times10^{-10}$ in psychiatric cases and $P > 10\times10^{-6}$ in controls). Relatedness was calculated across cohorts using identity by descent and one of each pair of related individuals (pi_hat > 0.2) was excluded. Principal components (PCs) were generated using genotyped SNPs in each cohort separately using EIGENSTRAT v6.1.4 (ref. [95]). On the basis of visual inspection of plots of PCs for each dataset (which were all of European descent according to self-report/clinical data), we excluded samples to obtain more clearly homogeneous datasets. Genotype imputation was performed using the prephasing/imputation stepwise approach implemented in Eagle v2.3.5 (ref. [96]) and Minimac3 (ref. [97]) to the Haplotype Reference Consortium (HRC) reference panel v1.0 (ref. [98]). Data on the X chromosome were available for 50 cohorts internal to the PGC and one external cohort (HUNT), and the X chromosome was imputed to the HRC reference panel in males and females separately within each cohort. The five external cohorts were processed by the collaborating research teams using comparable procedures and imputed to the HRC or a custom reference panel as appropriate. Full details of the genotyping, quality control and imputation for each of these cohorts are available in the Supplementary Note. Identical individuals between PGC cohorts and the Estonian Biobank and UK Biobank cohorts were detected using genotype-based checksums (https://personal.broadinstitute.org/sripke/share_links/zpXkV8INxUg9bayDpLToG4g58TMtjN_PGC_SCZ_w3.0718d.76) and removed from PGC cohorts.

GWAS. For PGC cohorts, GWASs were conducted within each cohort using an additive logistic regression model in PLINK v1.90 (ref. [99]), covarying for PCs 1–5 and any others as required. Association analyses of the X chromosome were conducted in males and females separately using the same procedures, with males coded as 0 or 2 for 0 or 1 copies of the reference allele. Results from males and females were then meta-analyzed within each cohort. For external cohorts, GWASs were conducted by the collaborating research teams using comparable procedures (Supplementary Note). To control test statistic inflation at SNPs with low minor allele frequency (MAF) in small cohorts, SNPs were retained only if cohort MAF was >1% and minor allele count was >10 in either cases or controls (whichever had smaller *n*). There was no evidence of stratification artifacts or uncontrolled inflation of test statistics in the results from any cohort (λ_{GC} = 0.97–1.05; Supplementary Table 1). Meta-analysis of GWAS summary statistics was conducted using an inverse-variance-weighted fixed-effects model in METAL (version 2011-03-25)[100] across 57 cohorts for the autosomes (41,917 BD cases and 371,549 controls) and 51 cohorts for the X chromosome (35,691 BD cases and 96,731 controls). A genome-wide significant locus was defined as the region around a SNP with $P < 5\times10^{-8}$, with LD $r^2 > 0.1$, within a 3,000-kb window. Regional association plots and forest plots of the index SNPs for all genome-wide significant loci are presented in Supplementary Data 1 and 2, respectively.

Overlap of loci with other psychiatric disorders. Genome-wide significant loci for BD were assessed for overlap with genome-wide significant loci for other psychiatric disorders, using the largest available GWAS results for major depression[61], schizophrenia[60], attention deficit/hyperactivity disorder[101], post-traumatic stress disorder[102], lifetime anxiety disorder[103], Tourette's syndrome[104], anorexia nervosa[105], alcohol use disorder or problematic alcohol use[68], autism spectrum disorder[106], mood disorders[91] and the cross-disorder GWAS of the PGC[66]. The boundaries of the genome-wide significant loci were calculated in the original publications. Overlap of loci was calculated using bedtools v2.29.2 (ref. [107]).

Enrichment analyses. *P* values quantifying the degree of association of genes and gene sets with BD were calculated using MAGMA v1.08 (ref. [37]), implemented in FUMA v1.3.6a (refs. [64,108]). Gene-based tests were performed for 19,576 genes (Bonferroni-corrected *P*-value threshold = 2.55×10^{-6}). A total of 11,858 curated gene sets including at least 10 genes from MSigDB V7.0 were tested for association with BD (Bonferroni-corrected *P*-value threshold = 4.22×10^{-6}). Competitive gene-set tests were conducted correcting for gene size, variant density and LD within and between genes. Tissue-set enrichment analyses were also performed using MAGMA implemented in FUMA, to test for enrichment of association signal in genes expressed in 54 tissue types from GTEx V8 (Bonferroni-corrected *P*-value threshold = 9.26×10^{-4})[64,108].

For single-cell enrichment analyses, publicly available single-cell RNA-sequencing data were compiled from five studies of the adult human and mouse brain[86,109–112]. The mean expression for each gene in each cell type was computed from the single-cell expression data (if not provided). For the Zeisel dataset[109], we used the mean expression at level 4 (39 cell types from 19 regions for the mouse nervous system). For the Saunders dataset[110], we computed the mean expression of the different classes in each of the 9 different brain regions sampled (88 cell types in total). We filtered out any genes with nonunique names, genes not expressed in any cell types, non-protein-coding genes and, for mouse datasets, genes that had no expert-curated 1:1 orthologs between mouse and human (Mouse Genome Informatics, The Jackson Laboratory, version 11/22/2016, http://www.informatics.jax.org/downloads/reports/index.html#homology), resulting in 16,472 genes. Gene expression was then scaled to a total of 1 million unique molecular identifiers (or transcripts per million) for each cell type/tissue. Using a previously described method[38], a metric of gene expression specificity was calculated by dividing the expression of each gene in each cell type by the total expression of that gene in all cell types, leading to values ranging from 0 to 1 for each gene (0 meaning that the gene is not expressed in that cell type and 1 meaning that all of the expression of the gene is in that cell type). We then selected the top 10% most specific genes for each cell type/tissue for enrichment analysis. MAGMA v1.08 (ref. [37]) was used to test gene-set enrichment using GWAS summary statistics, covarying for gene size, gene density, mean sample size for tested SNPs per gene, the inverse of the minor allele counts per gene and the log of these metrics. We excluded any SNPs with INFO score < 0.6, with MAF < 1% or with estimated odds ratio > 25 or smaller than 1/25, as well as SNPs located in the MHC region (chr6:25–34 Mb). We set a window of 35 kb upstream to 10 kb downstream of the gene coordinates to compute gene-level association statistics and used the European reference panel from phase 3 of the 1000 Genomes Project as the reference population[113]. We then used MAGMA to test whether the 10% most specific genes (with an expression of at least 1 transcript per million or 1 unique molecular identifier per million) for each cell type/tissue were associated with BD. The *P*-value threshold for significance was $P < 9.1\times10^{-3}$, representing a 5% false discovery rate across datasets.

Further gene-set analyses were performed restricted to genes targeted by drugs, assessing individual drugs and grouping drugs with similar actions. This approach has been described previously[41]. Gene-level and gene-set analyses were performed in MAGMA v1.08 (ref. [37]). Gene boundaries were defined using build 37 reference data from the National Center for Biotechnology Information, available on the MAGMA website (https://ctg.cncr.nl/software/magma), extended 35 kb upstream and 10 kb downstream to include regulatory regions outside the transcribed region. Gene-level association statistics were defined as the aggregate of the mean and the lowest variant-level *P* value within the gene boundary, converted to a *Z* value. Gene sets were defined comprising the targets of each drug in the Drug–Gene Interaction database DGIdb v.2 (ref. [39]) and in the Psychoactive Drug Screening Database Ki DB[40], both downloaded in June 2016[41]. Analyses were performed using competitive gene-set analyses in MAGMA. Results from the drug-set analysis were then grouped according to the Anatomical Therapeutic Chemical class of the drug[41]. Only drug classes with at least ten valid drug gene sets within them were analyzed. Drug-class analysis was performed using enrichment curves. All drug gene sets were ranked by their association in the drug-set analysis, and then for a given drug class an enrichment curve was drawn scoring a 'hit' if the drug gene set was within the class, or a 'miss' if it was outside the class. The area under the curve was calculated, and a *P* value for this was calculated using the Wilcoxon Mann–Whitney test comparing drug gene sets within the class to drug gene sets outside the class[41]. Multiple testing was controlled using a Bonferroni-corrected significance threshold of $P < 5.60\times10^{-5}$ for drug-set analysis and $P < 7.93\times10^{-4}$ for drug-class analysis, accounting for 893 drug sets and 63 drug classes tested.

eQTL integrative analysis. A TWAS was conducted using the precomputed gene expression weights from PsychENCODE data (1,321 brain samples)[43], available online with the FUSION software[42]. For genes with significant *cis*-SNP heritability (13,435 genes), FUSION software (vOct 1, 2019) was used to test whether SNPs influencing gene expression are also associated with BD (Bonferroni-corrected *P*-value threshold $< 3.72\times10^{-6}$). For regions including a TWAS-significant gene, TWAS fine-mapping of the region was conducted using FOCUS (fine-mapping of causal gene sets, v0.6.10)[44]. Regions were defined using the correlation matrix of predicted effects on gene expression around TWAS-significant genes[44]. A PIP was assigned to each gene for being causal for the observed TWAS association signal. Based on the PIP of each gene and a null model, whereby no gene in the region is causal for the TWAS signal, the 90%-credible gene set for each region was computed[44].

SMR (v1.03)[45,46] was applied to further investigate putative causal relationships between SNPs and BD via gene expression. SMR was performed using eQTL summary statistics from the eQTLGen (31,684 blood samples)[47] and PsychENCODE[43] consortia. SMR analysis is limited to transcripts with at least one significant *cis*-eQTL ($P < 5\times10^{-8}$) in each dataset (15,610 in eQTLGen; 10,871 in PsychENCODE). The Bonferroni-corrected significance threshold was $P < 3.20\times10^{-6}$ and $P < 4.60\times10^{-6}$ for eQTLGen and PsychENCODE, respectively. The significance threshold for the HEIDI test was $P_{HEIDI} \geq 0.01$ (ref. [46]). While the results of TWAS and SMR indicate an association between BD and gene expression, a nonsignificant HEIDI test additionally indicates either a direct causal role or a pleiotropic effect of the BD-associated SNPs on gene expression.

***C4* imputation.** To investigate the MHC (chr6:24–34 Mb on hg19), the alleles of *C4* genes (*C4A* and *C4B*) were imputed in 47 PGC cohorts for which individual-level genotype data were accessible, totaling 32,749 BD cases and 53,370 controls. The imputation reference panel comprised 2,530 reference haplotypes of MHC SNPs and *C4* alleles, generated using a sample of 1,265 individuals with whole-genome sequence data, from the Genomic Psychiatry cohort[114]. Briefly, imputation of *C4* as a multiallelic variant was performed using Beagle v4.1 (refs. [115,116]), using SNPs from the MHC region that were also in the haplotype reference panel. Within the Beagle pipeline, the reference panel was first converted to bref format. We used the conform-gt tool to perform strand-flipping and filtering of specific SNPs for which the strand remained ambiguous. Beagle was run using default parameters with two key exceptions: we used the GRCh37 PLINK recombination map, and we set the output to include genotype probability (that is, GP field in VCF) for correct downstream probabilistic estimation of *C4A* and *C4B* joint dosages. The output consisted of dosage estimates for each of the common *C4* structural haplotypes for each individual. The five most common structural forms of the *C4A*/*C4B* locus (BS, AL, AL–BS, AL–BL and AL–AL) could be inferred with reasonably high accuracy (generally $0.70 < r^2 < 1.00$). The imputed *C4* alleles were tested for association with BD in a joint logistic regression that included terms for dosages of the five most common *C4* structural haplotypes (AL–BS, AL–BL, AL–AL, BS and AL), rs13195402 genotype (top lead SNP in the MHC) and PCs as per the GWAS. The genetically regulated expression of *C4A* was predicted from the imputed *C4* alleles using a model previously described[63]. Predicted *C4A* expression was tested for association with BD in a joint logistic regression that included predicted *C4A* expression, rs13195402 genotype (top lead SNP in the MHC) and PCs as per the GWAS.

Polygenic risk scoring. PRSs from our GWAS meta-analysis were tested for association with BD in individual cohorts, using a discovery GWAS where the target cohort was left out of the meta-analysis. Briefly, the GWAS results from each discovery GWAS were pruned for LD using the *P*-value-informed clumping method in PLINK v1.90 (ref. [99]; r^2 0.1 within a 500-kb window) based on the LD structure of the HRC reference panel[98]. Subsets of SNPs were selected from the results below nine increasingly liberal *P*-value thresholds (GWAS P_T: 5×10^{-8}, 1×10^{-4}, 1×10^{-3}, 0.01, 0.05, 0.1, 0.2, 0.5, 1). Sets of alleles, weighted by their log odds ratios from the discovery GWAS, were summed into PRSs for each individual in the target datasets, using PLINK v1.90 implemented via RICOPILI[94,99]. PRSs were tested for association with BD in the target dataset using logistic regression, covarying for PCs as per the GWAS in each cohort. PRSs were tested in the external cohorts by the collaborating research teams using comparable procedures. The variance explained by the PRSs (R^2) was converted to the liability scale to account for the proportion of cases in each target dataset, using a BD population prevalence of 2% and 1%[117]. The weighted average R^2 values were calculated using the effective *n* for each cohort. The odds ratios for BD for individuals in the top decile of PRSs compared with those in the lowest decile and middle decile were calculated in the 52 datasets internal to the PGC. To assess cross-ancestry performance, PRSs generated from the meta-analysis results were tested for association with BD using similar methods in a Japanese sample[59], a Korean sample[92] and two admixed African American samples. Full details of the QC, imputation and analysis of these samples are in the Supplementary Note.

LDSC. LDSC[35] was used to estimate the h^2_{SNP} of BD from GWAS summary statistics. h^2_{SNP} was converted to the liability scale, using a lifetime BD prevalence of 2% and 1%. LDSC bivariate genetic correlations attributable to genome-wide SNPs (r_g) were estimated with 255 human diseases and traits from published GWASs and 514 GWASs of phenotypes in the UK Biobank from LD Hub[48]. Adjusting for the number of traits tested, the Bonferroni-corrected *P*-value thresholds were $P < 1.96\times10^{-4}$ and $P < 9.73\times10^{-5}$, respectively.

MiXeR. We applied causal mixture models[49,50] to the GWAS summary statistics, using MiXeR v1.3. MiXeR provides univariate estimates of the proportion of non-null SNPs ('polygenicity') and the variance of effect sizes of non-null SNPs ('discoverability') in each phenotype. For each SNP, *i*, univariate MiXeR models its additive genetic effect of allele substitution, β_i, as a point-normal mixture, $\beta_i = (1-\pi_1)\,N(0,0) + \pi_1 N(0, \sigma^2_\beta)$, where π_1 represents the proportion of non-null SNPs ('polygenicity') and σ^2_β represents variance of effect sizes of non-null SNPs ('discoverability'). Then, for each SNP, *j*, MiXeR incorporates LD information and allele frequencies for $M = 9{,}997{,}231$ SNPs extracted from 1000 Genomes phase 3 data to estimate the expected probability distribution of the signed test statistic, $z_j = \delta_j + \varepsilon_j = N\sum_i \sqrt{H_i} r_{ij}\beta_i + \varepsilon_j$, where *N* is sample size, H_i indicates heterozygosity of the *i*th SNP, r_{ij} indicates allelic correlation between the *i*th and *j*th SNPs and $\varepsilon_j \sim N(0, \sigma^2_0)$ is the residual variance. Further, the three parameters, π_1, σ^2_β and σ^2_0, are fitted by direct maximization of the likelihood function. The optimization is based on a set of approximately 600,000 SNPs, obtained by selecting a random set of 2,000,000 SNPs with MAF of 5% or higher, followed by LD pruning at LD $r^2 = 0.8$ threshold. The random SNP selection and full optimization procedure are repeated 20 times to obtain the means and standard errors of model parameters. The log-likelihood figures show individual curves for each of the 20 runs, each shifted vertically so that the best log-likelihood point is shown at the zero ordinate.

The total number of trait-influencing variants is estimated as $M\pi_1$, where $M = 9{,}997{,}231$ gives the number of SNPs in the reference panel. MiXeR Venn diagrams report the effective number of influencing variants, $\eta M\pi_1$, where η is a fixed number, $\eta = 0.319$, which gives the fraction of influencing variants contributing to 90% of the trait's heritability (with rationale for this adjustment being that the remaining 68.1% of influencing variants are small and cumulatively explain only 10% of the trait's heritability). Phenotypic variance explained on average by an influencing genetic variant is calculated as $\underline{H}\sigma^2_\beta$, where $\underline{H} = \frac{1}{M}\sum_i H_i = 0.2075$ is the average heterozygosity across SNPs in the reference panel. Under the assumptions of the MiXeR model, SNP heritability is then calculated as $h^2_{SNP} = M\pi_1 \times \underline{H}\sigma^2_\beta$.

In the cross-trait analysis, MiXeR models additive genetic effects as a mixture of four components, representing null SNPs in both traits (π_0); SNPs with a specific effect on the first and on the second trait (π_1 and π_2, respectively); and SNPs with nonzero effect on both traits (π_{12}). In the last component, MiXeR models the variance–covariance matrix as $\Sigma_{12} = \begin{bmatrix} \sigma^2_1 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma^2_2 \end{bmatrix}$, where ρ_{12} indicates correlation of effect sizes within the shared component, and σ^2_1 and σ^2_2 correspond to the discoverability parameter estimated in the univariate analysis of the two traits. These components are then plotted in Venn diagrams. After fitting parameters of the model, the Dice coefficient of polygenic overlap is then calculated as $\frac{2\pi_{12}}{\pi_1+2\pi_{12}+\pi_2}$, and genetic correlation is calculated as $r_g = \frac{\rho_{12}\pi_{12}}{\sqrt{(\pi_1+\pi_{12})(\pi_2+\pi_{12})}}$. The fraction of influencing variants with concordant effect direction is calculated as twice the multivariate normal cumulative distribution function at point (0, 0) for the bivariate normal distribution with zero mean and variance–covariance matrix Σ_{12}. All code is available online (https://github.com/precimed/mixer).

MR. We selected 17 traits associated with BD in clinical or epidemiological studies for MR to dissect their relationship with BD (Supplementary Note). Bidirectional GSMR[51] analyses were performed between BD and the traits of interest using GWAS summary statistics, implemented in GCTA software (v1.93.1f beta). The instrumental variables were selected by a clumping procedure internal to the GSMR software with parameters: --gwas-thresh 5×10^{-8} --clump-r2 0.01. Traits with fewer than 10 instrumental variables available were excluded from the GSMR analyses to avoid conducting underpowered tests[51], resulting in 10 traits tested (Bonferroni-corrected *P*-value threshold $< 2.5\times10^{-3}$). The HEIDI-outlier test was applied to test for horizontal pleiotropy ($P_{HEIDI} < 0.01$)[51]. For comparison, the MR analyses were also performed using the inverse-variance-weighted regression method, implemented via the TwoSampleMR R package, using the instrumental variables selected by GSMR[118,119]. To further investigate horizontal pleiotropy, the MR Egger intercept test was conducted using the TwoSampleMR package[118,119] and MR-PRESSO software was used to perform the global test and the distortion test[120].

BD subtypes. GWAS meta-analyses were conducted for BD I (25,060 cases, 449,978 controls from 55 cohorts, effective $n = 64{,}802$) and BD II (6,781 cases, 364,075 controls from 31 cohorts, effective $n = 22{,}560$; Supplementary Table 1) using the same procedures described for the main GWAS. BD subtypes were defined based on international consensus criteria (DSM-IV, ICD-9 or ICD-10), established using structured diagnostic instruments from assessments by

trained interviewers, clinician-administered checklists or medical record review. In the external biobank cohorts, BD subtypes were defined using ICD codes (Supplementary Note). LDSC[35] was used to estimate the h^2_{SNP} of each subtype, and the genetic correlation between the subtypes. The difference between the LDSC h^2_{SNP} estimates for BD I and BD II was tested for deviation from 0 using the block jackknife[121]. The LDSC genetic correlation (r_g) was tested for difference from 1 by calculating a chi-square statistic corresponding to the estimated r_g as $((r_g-1)/\text{s.e.})^2$.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

GWAS summary statistics are publicly available on the PGC website (https://www.med.unc.edu/pgc/results-and-downloads). Individual-level data are accessible through collaborative analysis proposals to the Bipolar Disorder Working Group of the PGC (https://www.med.unc.edu/pgc/shared-methods/how-to/). This study included some publicly available datasets accessed through dbGaP (PGC bundle phs001254) and the HRC reference panel v1.0 (http://www.haplotype-reference-consortium.org/home). Databases used: Drug–Gene Interaction Database DGIdb v.2 (https://www.dgidb.org); Psychoactive Drug Screening Database Ki DB (https://pdsp.unc.edu/databases/kidb.php); DrugBank 5.0 (https://www.drugbank.ca); LD Hub (http://ldsc.broadinstitute.org); FUMA (https://fuma.ctglab.nl).

Code availability

All software used is publicly available at the URLs or references cited.

References

92. Moon, S. et al. The Korea Biobank Array: design and identification of coding variants associated with blood biochemical traits. *Sci. Rep.* **9**, 1382 (2019).
93. Bigdeli, T. B. et al. Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. *Mol. Psychiatry* **25**, 2455–2467 (2020).
94. Lam, M. et al. RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics* **36**, 930–933 (2020).
95. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
96. Loh, P.-R. et al. Reference-based phasing using the haplotype reference consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
97. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
98. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
99. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
100. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
101. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
102. Nievergelt, C. M. et al. International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nat. Commun.* **10**, 4558 (2019).
103. Purves, K. L. et al. A major role for common genetic variation in anxiety disorders. *Mol. Psychiatry* **25**, 3292–3303 (2020).
104. Yu, D. et al. Interrogating the genetic determinants of Tourette's syndrome and other tic disorders through genome-wide association studies. *Am. J. Psychiatry* **176**, 217–227 (2019).
105. Watson, H. J. et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214 (2019).
106. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
107. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
108. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
109. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014 (2018).
110. Saunders, A. et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030 (2018).
111. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
112. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
113. 1000 Genomes Project Consortium. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
114. Kamitaki, N. et al. Complement genes contribute sex-biased vulnerability in diverse disorders. *Nature* **582**, 577–581 (2020).
115. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
116. Browning, B. L. & Browning, S. R. Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **98**, 116–126 (2016).
117. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.* **36**, 214–224 (2012).
118. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).
119. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408 (2018).
120. Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
121. Hübel, C. et al. Genomics of body fat percentage may contribute to sex bias in anorexia nervosa. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **180**, 428–438 (2019).

Acknowledgements

We thank the participants who donated their time, life experiences and DNA to this research and the clinical and scientific teams that worked with them. We are deeply indebted to the investigators who make up the PGC. The PGC has received major funding from the US National Institute of Mental Health (PGC3: U01 MH109528; PGC2: U01 MH094421; PGC1: U01 MH085520). Statistical analyses were carried out on the NL Genetic Cluster Computer (http://www.geneticcluster.org) hosted by SURFsara and the Mount Sinai high-performance computing cluster (http://hpc.mssm.edu), which is supported by the Office of Research Infrastructure of the National Institutes of Health under award numbers S10OD018522 and S10OD026880. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Full acknowledgements are included in the Supplementary Note.

Author contributions

Writing group: N.M., A.J.F., K.S.O'C., B.C., J.R.I.C., J.M.B., J.I.N., S. Cichon, H.J.E., E.A.S., A. McQuillin, A.D.F., R.A.O., O.A.A. PGC BD PI group: A.J.F., M.I., H.-H.W., D.C., R.A., I.A., M.A., L. Alfredsson, G. Babadjanova, L.B., B.T.B., F.B., S. Bengesser, W.H.B., D.H.R.B., M. Boehnke, A.D.B., G. Breen, V.J.C., S. Catts, A.C., N.C., U.D., D.D., T. Esko, B.E., P.F., M.F., J.M.F., M.G., E.S.G., F.S.G., M. J. Green, M.G.-S., J. Hauser, F.H., J. Hillert, K.S.H., D.M.H., C.M.H., K. Hveem, N.I., A.V.J., I.J., L.A.J., R.S.K., J.R.K., G.K., M. Landén, M. Leboyer, C.M.L., Q.S.L., J. Lissowska, C. Lochner, C. Loughland, N.G.M., C.A.M., F.M., S.L.M., A.M.M., F.J.M., I.M., P. Michie, L. Milani, P. B. Mitchell, G.M., O.M., P. B. Mortensen, B.M., B.M.-M., R.M.M., B.M.N., C.M.N., M.N., M.M.N., M.C.O'D., K.J.O., T.O., M.J.O., S.A.P., C. Pantelis, C. Pato, M.T.P., G.P.P., R.H.P., D.P., J.A.R.-Q., A.R., E.Z.R., M. Ribasés, M. Rietschel, S.R., G.A.R., T.S., U.S., M.S., P.R.S., T.G.S., L.J.S., R.J.S., A.S., C.S.W., J.W.S., H.S., K.S., E. Stordal, F. Streit, P.F.S., G.T., A.E.V., E.V., J.B.V., I.D.W., T.W.W., T.W., N.R.W., J.-A.Z., J.M.B., J.I.N., S. Cichon, H.J.E., E.A.S., A. McQuillin, A.D.F., R.A.O., O.A.A. Bioinformatics: N.M., A.J.F., J.R.I.C., S. Børte, M.J. Gandal, M. Kim, B.M.S., L.G.S., B.S.W., H.-H.W., N.A.-R., S.E.B., B.M.B., V.E.-P., S.H., P.A.H., Y.K., M. Koromina, M. Kubo, M. Leber, P.H.L., C. Liao, L.M.O.L., T.R., P.R., P.D.S., M.S.A., C. Terao, T.E.T., S.X., H.Y., P.P.Z., S. Bengesser, G. Breen, P.F., E.S.G., Q.S.L., G.A.R., H.S., T.W., E.A.S. Clinical: O.K.D., M.I., L. Abramova, K.A., E.A., N.A.-R., A. Anjorin, A. Antoniou, J.H.B., N.B., M. Bauer, A.B., C.B.P., E.B., M.P.B., R.B., M. Brum, N.B.-K., M. Budde, W.B., M. Cairns, M. Casas, P.C., A.C.-B., D.C., P.M.C., N.D., A.D., T. Elvsåshagen, L. Forty, L. Frisén, K.G., J. Garnham, M.G.P., I.R.G., K.G.-S., J. Grove, J.G.-P., K. Ha, M. Haraldsson, M. Hautzinger, U.H., D.H., J. L. Kalman, J. L. Kennedy, S.K.-S., M. Kogevinas, T.M.K., R.K., S.A.K., J. Lawrence, H.-J.L., C. Lewis, S.L., M. Lundberg, D.J.M., W.M., D.M., L. Martinsson, M.M., P. McGuffin, H.M., V.M., C.O'D., L.O., S.P., A. Perry, A. Pfennig, E.P., J.B.P., D.Q., M.H.R., J.R.D., E.J.R., J.P.R., F.R., J.R., E.C.S., F. Senner, E. Sigurdsson, L.S., C.S., O.B.S., D. J. Smith, J.L.S., A.T.S., J.S.S., B.Ś., P.T., M.P.V., H.V., A.H.Y., L.Z., HUNT All-In Psychiatry, R.A., I.A., M.A., G. Babadjanova, L.B., B.T.B., F.B., S. Bengesser, D.H.R.B., A.D.B., A.C., N.C., U.D., D.D., B.E., P.F., M.F., M.G., E.S.G., F.S.G., M. J. Green, M.G.-S., J. Hauser, K.S.H., N.I., I.J., L.A.J., R.S.K., G.K., M. Landén, C.M.L., J. Lissowska, N.G.M., C.A.M., F.M., S.L.M., A.M.M., I.M., P. B. Mitchell, G.M., O.M., P. B. Mortensen, M.C.O'D., K.J.O., M.J.O., C. Pato, M.T.P., R.H.P., J.A.R.-Q., A.R., E.Z.R., M. Rietschel, T.S., T.G.S., A.S., C.S.W., J.W.S., E. Stordal, F. Streit, A.E.V., E.V., J.B.V., I.D.W., T.W.W., T.W., J.I.N., A. McQuillin, A.D.F. Genomic assays/data generation: A.J.F., M.I., E.A., M.A.E., D.A., M.B.-H., E.C.B., C.B.P., J.B.-G., M. Cairns, T.-K.C., C.C., J.C., F.S.D., F.D., S.D., A.F., J.F., N.B.F., J. Gelernter, M.G.P., P.H., S.J., Y.K., H.R.K., M. Kubo, S.E.L., C. Liao, E.M., N.W.M., J.D.M., G.W.M., J.L.M., D.W.M., T.W.M., N.O'B., M. Rivera, C.S.-M., S. Sharp, C.S.H., C. Terao, C. Toma, E.-E.T., S.H.W., HUNT All-In Psychiatry, G. Breen, A.C., T. Esko, J.M.F., E.S.G., D.M.H., N.I., F.J.M., L. Milani,

R.M.M., M.M.N., M. Ribasés, G.A.R., T.S., G.T., S. Cichon. Obtained funding for BD samples: M.I., M. Cairns, I.N.F., L. Frisén, S.J., Y.K., J.A.K., M. Kubo, C. Lavebratt, S.L., D.M., P. McGuffin, G.W.M., J.B.P., M.H.R., J.R.D., D. J. Stein, J.S.S., C. Terao, A.H.Y., P.P.Z., M.A., L. Alfredsson, L.B., B.T.B., F.B., W.H.B., M. Boehnke, A.D.B., G. Breen, A.C., N.C., B.E., M.F., J.M.F., E.S.G., M. J. Green, M.G.-S., K.S.H., K. Hveem, N.I., I.J., L.A.J., M. Landén, M. Leboyer, N.G.M., F.J.M., P. B. Mitchell, O.M., P. B. Mortensen, B.M.N., M.N., M.M.N., M.C.O'D., T.O., M.J.O., C. Pato, M.T.P., G.P.P., M. Rietschel, G.A.R., T.S., M.S., P.R.S., T.G.S., C.S.W., J.W.S., G.T., J.B.V., T.W.W., T.W., J.M.B., J.I.N., H.J.E., R.A.O., O.A.A. Statistical analysis: N.M., K.S.O'C., B.C., J.R.I.C., Z.Q., T.D.A., T.B.B., S. Børte, J.B., A.W.C., O.K.D., M. J. Gandal, S.P.H., N.K., M. Kim, K.K., G.P., B.M.S., L.G.S., S. Steinberg, V.T., B.S.W., H.-H.W., V.A., S.A., S.E.B., B.M.B., A.M.D., A.L.D., V.E.-P., T.M.F., O.F., S.D.G., T.A.G., J. Grove, P.A.H., L.H., J.S.J., Y.K., M. Kubo, C. Lavebratt, M. Leber, P.H.L., S.H.M., A. Maihofer, M.M., S.A.M., S.E.M., L.M.O.L., A.F.P., T.R., P.R., D.M.R., O.B.S., C. Terao, T.E.T., J.T.R.W., W.X., J.M.K.Y., H.Y., P.P.Z., H.Z., A.D.B., G. Breen, E.S.G., F.S.G., Q.S.L., B.M.-M., C.M.N., D.P., S.R., H.S., P.F.S., T.W., N.R.W., J.M.B., E.A.S. K.S.O'C., B.C., J.R.I.C. and Z.Q. contributed equally to this work and should be regarded as joint second authors.

Competing interests

T.E.T., S. Steinberg, H.S. and K.S. are employed by deCODE Genetics/Amgen. Multiple additional authors work for pharmaceutical or biotechnology companies in a manner directly analogous to academic coauthors and collaborators. A.H.Y. has given paid lectures and served on advisory boards relating to drugs used in affective and related disorders for several companies (AstraZeneca, Eli Lilly, Lundbeck, Sunovion, Servier, Livanova, Janssen, Allergan, Bionomics and Sumitomo Dainippon Pharma), was Lead Investigator for Embolden Study (AstraZeneca), BCI Neuroplasticity study and Aripiprazole Mania Study, and is an investigator for Janssen, Lundbeck, Livanova and Compass. J.I.N. is an investigator for Janssen. P.F.S. reports the following potentially competing financial interests: Lundbeck (advisory committee), Pfizer (Scientific Advisory Board member) and Roche (grant recipient, speaker reimbursement). G. Breen reports consultancy and speaker fees from Eli Lilly and Illumina and grant funding from Eli Lilly. M. Landén has received speaker fees from Lundbeck. O.A.A. has received speaker fees from Lundbeck and Sunovion, and is a consultant to HealthLytix. J.A.R.-Q. was on the speakers bureau and/or acted as consultant for Eli Lilly, Janssen-Cilag, Novartis, Shire, Lundbeck, Almirall, Braingaze, Sincrolab and Rubió in the last 5 years. He also received travel awards (air tickets and hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire and Eli Lilly. The Department of Psychiatry chaired by him received unrestricted educational and research support from the following companies in the last 5 years: Eli Lilly, Lundbeck, Janssen-Cilag, Actelion, Shire, Ferrer, Oryzon, Roche, Psious and Rubió. E.V. has received grants and served as a consultant, advisor or CME speaker for the following entities: AB-Biotics, Abbott, Allergan, Angelini, AstraZeneca, Bristol Myers Squibb, Dainippon Sumitomo Pharma, Farmindustria, Ferrer, Forest Research Institute, Gedeon Richter, GlaxoSmithKline, Janssen, Lundbeck, Otsuka, Pfizer, Roche, SAGE, Sanofi-Aventis, Servier, Shire, Sunovion, Takeda, the Brain and Behaviour Foundation, the Catalan Government (AGAUR and PERIS), the Spanish Ministry of Science, Innovation, and Universities (AES and CIBERSAM), the Seventh European Framework Programme and Horizon 2020 and the Stanley Medical Research Institute. T. Elvsåshagen has received speaker fees from Lundbeck. S.K.-S. received author's and consultant honoraria from Medice Arzneimittel Pütter GmbH and Shire/Takeda. A.S. is or has been a consultant/speaker for: Abbott, Abbvie, Angelini, AstraZeneca, Clinical Data, Boheringer, Bristol Myers Squibb, Eli Lilly, GlaxoSmithKline, Innovapharma, Italfarmaco, Janssen, Lundbeck, Naurex, Pfizer, Polifarma, Sanofi, Servier. J.R.D. has served as an unpaid consultant to Myriad – Neuroscience (formerly Assurex Health) in 2017 and 2019 and owns stock in CVS Health. H.R.K. serves as an advisory board member for Dicerna Pharmaceuticals, and is a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was sponsored in the past 3 years by AbbVie, Alkermes, Amygdala Neurosciences, Arbor Pharmaceuticals, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka and Pfizer. H.R.K. is named as an inventor on PCT patent application no. 15/878,640 entitled: Genotype-guided dosing of opioid agonists, filed January 24, 2018. B.M.N. is a member of the scientific advisory board at Deep Genomics and consultant for Camp4 Therapeutics, Takeda Pharmaceutical and Biogen. All other authors declare no financial interests or potential conflicts of interest.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00857-4.

Correspondence and requests for materials should be addressed to N.M. or O.A.A.

Peer review information *Nature Genetics* thanks Na Cai, Qiang Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

4 DISCUSSION

The projects presented herein have an important unifying factor in that they have all contributed to a better understanding the complete genetic architecture of a range of different common, genetically complex neuropsychiatric traits. They are all puzzle pieces affirming that the current concept of the genetic framework underlying these disorders is one of a spectrum of genetic variation of all frequencies and effect sizes that jointly make-up the genetic puzzle of these disorders, independent of the diagnostic categories used clinically. Although this work includes a number of studies that were able to confidently and robustly link specific genes or variants to a given phenotype across the frequency spectrum, they also illustrate the conundrums commonly encountered in the study of the genetic underpinnings of common, complex disorders. For one, extremely large, ideally very deeply phenotyped cohorts are needed to robustly associate variants of different frequencies with common complex disorders. This is illustrated by the fact that some of the largest GWAS are currently performed on > 1.5 million individuals (Initiative, 2021) and that equally large or larger numbers are likely to be needed for successful genome-wide rare variant associations as well. While massive efforts are already underway in the field of neuropsychiatric genetics to perform such RVAS(e.g.(Fu, 2021b; Palmer, 2020; Satterstrom et al., 2019; Singh, 2020)), currently these "only" assess up to 200,000 individuals and focus only on coding variation. Accordingly, even nearly two decades into the genomic era, relevant samples sizes and financial resources are important limiting factors to a fuller understanding of the genetics of neuropsychiatric traits.

When smaller sample sizes are used in more focused analyses, results tend to be less reliable. One reason for this is the so-called "winner's curse" or regression towards the mean, which in genetics refers to the statistical tendency of an extreme event (such as the occurrence of a number of rare variants in a given gene or the identification of a locus of genome-wide significance in a GWAS) to be followed by a less extreme event. This has led to the report of false positive association findings and the failure of studies to replicate in the field of common, complex genetics(Grinde et al., 2017; Kesselmeier and Lorenzo Bermejo, 2017; Marigorta et al., 2018). The projects depicted herein have contributed to uncovering such—likely—false positives(Schulte et al., 2012), but they have also generated genetic candidates that await—and may never

see—replication(Schulte et al., 2014a). It is important to remain cognizant of this, especially in the study of rare variants.

Establishing statistical evidence for association is further complicated by the fact that many genes are likely to harbor both protective and predisposing variants for a given phenotype, a phenomenon which has prominently been shown for the Alzheimer's disease gene *APP*, for example(Jonsson et al., 2012). Here, more advanced statistical approaches—like the SKAT/SKAT-O test used in the study of rare variation in RLS-linked genes(Schulte et al., 2014b)–that are currently being developed may help to alleviate this problem. Also, the inclusion of synonymous variants in burden analyses yielded, in many instances, a more significant signal for an enrichment of rare coding variants. While this deserves attention, the true reason for this could not be established and could range from statistical to fine-scale population substructure to LD-dependent signal amplification to a true causal contribution as has been illustrated for other neuropsychiatric disorders(Cruchaga et al., 2014).

Next to the "winner's curse", the phenomenon that rare variants are more likely to be functionally relevant further hinders translation of genetic findings. This means that rare, functionally relevant variants are more likely to bridge the existing gap in our understanding between genetic variation and pathophysiology while it is, at the same time, very difficult to statistically link rare variants to a given phenotype simply because of the fact that they are so rare. This is a phenomenon that has been widely reported for, for example, drug target genes(Nelson et al., 2012). Even if variants are identified in an unbiased, comprehensive approach like WES and are located in functionally plausible genes like *PLXNA4*(Schulte et al., 2013b), *LRRK1*(Schulte et al., 2014a), or *APP*(Schulte et al., 2015a), firmly establishing causality of the given rare and very rare variants is very difficult. Yet, sometimes important novel genetic factors can robustly be identified by the approaches used. Rare variants in *VPS35*, for instance, have, since we first linked the gene to PD(Zimprich et al., 2011), been established as the cause of 0.3 to 1.3% of late-onset PD cases (Guella et al., 2012; Kumar et al., 2012) and more than 200 articles have been published that investigate *VPS35* in the context of PD.

The works presented herein investigated neuropsychiatric disorders belonging to the neurologic as well as the psychiatric spectrum. While the vast majority of investigated disorders is deemed to be common and genetically complex, it is interesting to note that the underlying genetic architecture still seems to differ. Overall, complex, common psychiatric disorders seem to have a much higher polygenicity(Smeland et al., 2020; Sullivan and Geschwind, 2019) than neurological disorders and, at the same time, a much higher genetic overlap than neurologic conditions (Brainstorm, 2018; PGC, 2019). As already mentioned above, this is reflected in the size and timing dynamics of GWAS, where we see that large differences exist between the neurologic and psychiatric conditions addressed in this work. However, there are also important differences between the three common severe psychiatric disorders themselves when it comes to the number of study participants needed to identify loci of genome-wide significance (Kendall et al., 2021)(Figure 3).

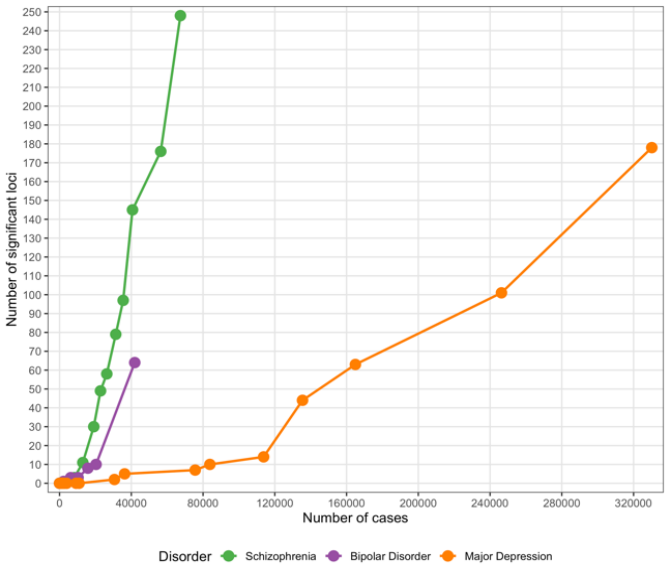


Figure 3: Relationship between GWAS case sample size and number of identified loci of genome-wide significance for different neuropsychiatric disorders. SCZ (green), BPD (purple), and MDD (orange)(taken from (Kendall et al., 2021)).

While some of these differences can likely be ascribed to true differences in genetic architecture, one also needs to be aware of the fact that many other variables can influence the perceived polygenicity and genetic architecture depicted in these studies. The list of potential confounds is long and includes biases in the selection of cases, heterogeneous case definitions, phenotypic heterogeneity—especially in psychiatric disorders where not all disorders receiving the same diagnostic code are also the same underlying disorder—, environmental confounders, sex differences, and epistatic effects, among others(Cai et al., 2020; Converge, 2015; Iwaki et al., 2020).

High polygenicity with contributions of hundreds, thousands, or tens of thousands of common variants of very small effect on the phenotype has been firmly established for many common, complex psychiatric disorders. In addition, the works depicted herein but also those of many others (e.g.(Fu, 2021b; Palmer, 2020; Satterstrom et al.,

2019; Singh, 2020)) have confidently placed rare variants on the genomic map of common complex neuropsychiatric disorders. For many of these disorders, it has been shown that rare and common variants converge on the same pathways and, accordingly, the same pathophysiologic mechanisms (e.g. (Schulte et al., 2014b; Singh, 2020)). Some even go so far as to postulate that all genetic variants in all genes expressed in a disease-relevant cell type contribute to bringing about a given phenotype under what is referred to as the "omnigenic" model (Boyle et al., 2017). The ability to now whole genome sequence tens and eventually hundreds of thousands and millions of individuals will ultimately bridge many—though certainly not all—gaps still present in our understanding of the genetic architecture of neuropsychiatric disorders.

5 PERSPECTIVES

While the identification of genomic loci associated with neuropsychiatric disorders represents an important component of the journey towards the full comprehension of the genetic framework of these disorders, the translation of the acquired knowledge into changes in the clinical management of individuals with these disorders is of equal if not greater importance. The past 10 or 15 years have seen genetic discoveries at an unprecedented scale. By 2020, for example, GWAS had reported more than 55,000 unique loci for almost 5,000 traits(Loos, 2020; MacArthur et al., 2017). Yet, clinical translation of genetic findings has not been able to keep up with the enormous speed at which new risk loci are discovered(Visscher et al., 2017; Zeggini et al., 2019).

To date, most studies in the field have investigated variants of different frequencies separately. Yet, a more accurate and complete model of the genetic architecture underlying common complex genetic disorders would require an integration of variants of all different frequencies and—ideally—also structural variation like CNVs. In the context of COVID-19, machine learning approaches have been used to generate first integrated scores which use coding genetic variants of all different frequencies as features to model the contribution of common, low-frequency, rare, and ultra-rare variants to disease severity(Fallerini et al., 2022). Similarly, in a preprint, one of the largest WES studies conducted to date utilized a Bayesian model to jointly analyze rare coding variants and CNVs for a given gene in the exomes of 63,237 individuals with autism spectrum disorders(Fu, 2021b). This led to the identification of 71 genes

associated with autism spectrum disorders at a false-discovery-rate (FDR)-corrected p-value < 0.001(Fu, 2021b). First frameworks for such integrated analyses are beginning to be explored but will surely become more sophisticated and powerful in coming years.

Despite the undisputedly great successes of genetic studies since the turn of the millennium, the road to a full biological understanding of the effects of the genetic variants in most neuropsychiatric disorders is still long. The integration of large, often publicly available omics datasets, such as from the Genotype-Tissue Expression (GTEx)(GTEx, 2013) or the ENCODE(ENCODE, 2012) projects, has led to a much better understanding of the pathophysiologic changes linked to genetic risk variants(Zeggini et al., 2019). RNA expression data, for example, was used to show that common risk factors involved in BPD show enriched expression in hippocampal pyramidal neurons and interneurons of the hippocampus and prefrontal cortex, implicating these cell types in BPD pathogenesis(Mullins et al., 2021). Prospectively, single-cell and single-nucleus RNA sequencing data in conjunction with, for instance, unsupervised clustering approaches can add even more detail and, thus, biological insights to such "*in silico*" functional follow-ups(Fu, 2021b; Peng et al., 2021) and large-scale efforts like the Human Cell Atlas(Rozenblatt-Rosen et al., 2017) will provide the highest-resolution tissue reference map yet using single-cell omics techniques.

Still, these approaches cannot fully replace functional, molecular biology follow-ups to truly understand the biological consequences of specific genetic variants in most disease contexts. Especially rare coding genetic variants of large effect harbor great potential to further our knowledge of the pathophysiology underlying neuropsychiatric disorders because larger effects are much more readily and more robustly detected in model systems. The individual analysis of all potentially relevant genetic variants in a given phenotype has become near impossible due to the vast number of genetic variants that have been linked to different traits. One way out of this dilemma are large-scale functional screens. Currently, this is an area of highly dynamic methodological development. One such methodology are pooled screens using CRISPR-base editors that introduce transition mutations at specific locations in the genome(Gaudelli et al., 2017; Hanna et al., 2021; Komor et al., 2016). Using this approach, Hanna *et al.*, in a proof-of-principle study, generated a library of single guide RNAs—RNAs used to target specific genomic base pairs in CRISPR-Cas9 genome editing—for 52,034 genetic variants of known clinical importance in 3,584

genes followed by cellular screens to identify loss-of-function phenotypes in several DNA damage repair genes(Hanna et al., 2021). This and several other very recent studies provide a glimpse into future opportunities in translating genetic risk variation into meaningful biological insights at a massively parallel scale(Klein et al., 2020; Zeggini et al., 2019).

In order for the genetic findings of the past decade to become more clinically translatable, it is also of vital importance to increase the ancestral representativeness of participants in these studies. Most genetic studies performed over the last decade have focused solely on individuals of European ancestry. Currently, only approximately 20% of all participants in GWAS are of non-European ancestry, while individuals of non-European descent represent 84% of the world's population(Loos, 2020; Martin et al., 2019; Mills and Rahal, 2020). This precludes in-depth insights to be gained from more genetically diverse and more ancient human populations such as many of those found on the African continent whose study could be vital in gaining a truly complete picture of the genetic architecture underlying a trait(Gurdasani et al., 2015; Zeggini et al., 2019). More importantly, though, it also generates biases and decreased transferability of genetic results that may also result in inequalities in access to clinical care informed by genetic findings(Loos, 2020; Martin et al., 2019). One example for this are polygenic risk scores (PRS), which represent a cumulative estimate of the total genetic risk derived from the common genetic variants an individual carries and which may be used in the future as part of risk prediction and personalized clinical management(Khera et al., 2018; Zeggini et al., 2019). PRS perform much worse if applied to individuals of ancestries different from those used to establish the scores (i.e. mostly European-only)(Martin et al., 2019). For BPD, the PRS derived from the most recent GWAS explain about 4.6% of the phenotypic variance in cohorts of European ancestry, but only a maximum of 2.3% in those of East Asian or 1.2% in those of admixed American ancestry(Mullins et al., 2021). The only remedy is to increase the participation of non-Europeans in genetic studies with several important efforts to do so already underway(Gurdasani et al., 2015; Stevenson et al., 2019) but many more needed.

6 REFERENCES

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. Nat Methods 7(4), 248-249.

All of Us Research Program, I., Denny, J.C., Rutter, J.L., Goldstein, D.B., Philippakis, A., Smoller, J.W., Jenkins, G., Dishman, E., 2019. The "All of Us" Research Program. N Engl J Med 381(7), 668-676.

Antonarakis, S.E., Chakravarti, A., Cohen, J.C., Hardy, J., 2010. Mendelian disorders and multifactorial traits: the big divide or one for all? Nat Rev Genet 11(5), 380-384.
Bonnefond, A., Clement, N., Fawcett, K., Yengo, L., Vaillant, E., Guillaume, J.L., Dechaume, A., Payne, F., Roussel, R., Czernichow, S., Hercberg, S., Hadjadj, S., Balkau, B., Marre, M., Lantieri, O., Langenberg, C., Bouatia-Naji, N., Meta-Analysis of, G., Insulin-Related Traits, C., Charpentier, G., Vaxillaire, M., Rocheleau, G., Wareham, N.J., Sladek, R., McCarthy, M.I., Dina, C., Barroso, I., Jockers, R., Froguel, P., 2012. Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat Genet 44(3), 297-301.

Boyle, E.A., Li, Y.I., Pritchard, J.K., 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169(7), 1177-1186.

Brainstorm, 2018. Analysis of shared heritability in common disorders of the brain. Science 360(6395).

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., Marchini, J., 2018. The UK Biobank resource with deep phenotyping and genomic data. Nature 562(7726), 203-209.

Cai, N., Revez, J.A., Adams, M.J., Andlauer, T.F.M., Breen, G., Byrne, E.M., Clarke, T.K., Forstner, A.J., Grabe, H.J., Hamilton, S.P., Levinson, D.F., Lewis, C.M., Lewis, G., Martin, N.G., Milaneschi, Y., Mors, O., Muller-Myhsok, B., Penninx, B., Perlis, R.H., Pistis, G., Potash, J.B., Preisig, M., Shi, J., Smoller, J.W., Streit, F., Tiemeier, H., Uher, R., Van der Auwera, S., Viktorin, A., Weissman, M.M., Consortium, M.D.D.W.G.o.t.P.G., Kendler, K.S., Flint, J., 2020. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nat Genet 52(4), 437-447.

Chartier-Harlin, M.C., Dachsel, J.C., Vilarino-Guell, C., Lincoln, S.J., Lepretre, F., Hulihan, M.M., Kachergus, J., Milnerwood, A.J., Tapia, L., Song, M.S., Le Rhun, E., Mutez, E., Larvor, L., Duflot, A., Vanbesien-Mailliot, C., Kreisler, A., Ross, O.A., Nishioka, K., Soto-Ortolaza, A.I., Cobb, S.A., Melrose, H.L., Behrouz, B., Keeling, B.H., Bacon, J.A., Hentati, E., Williams, L., Yanagiya, A., Sonenberg, N., Lockhart, P.J., Zubair, A.C., Uitti, R.J., Aasly, J.O., Krygowska-Wajs, A., Opala, G., Wszolek, Z.K., Frigerio, R., Maraganore, D.M., Gosal, D., Lynch, T., Hutchinson, M., Bentivoglio, A.R., Valente, E.M., Nichols, W.C., Pankratz, N., Foroud, T., Gibson, R.A., Hentati, F., Dickson, D.W., Destee, A., Farrer, M.J., 2011. Translation initiator EIF4G1 mutations in familial Parkinson disease. Am J Hum Genet 89(3), 398-406.

Chen, X., Kordich, J.K., Williams, E.T., Levine, N., Cole-Strauss, A., Marshall, L., Labrie, V., Ma, J., Lipton, J.W., Moore, D.J., 2019. Parkinson's disease-linked D620N VPS35 knockin mice manifest tau neuropathology and dopaminergic neurodegeneration. Proc Natl Acad Sci U S A 116(12), 5765-5774.

Chiu, C.C., Weng, Y.H., Huang, Y.Z., Chen, R.S., Liu, Y.C., Yeh, T.H., Lu, C.S., Lin, Y.W., Chen, Y.J., Hsu, C.C., Chiu, C.H., Wang, Y.T., Chen, W.S., Liu, S.Y., Wang, H.L., 2020. (D620N) VPS35 causes the impairment of Wnt/beta-catenin signaling cascade and mitochondrial dysfunction in a PARK17 knockin mouse model. Cell Death Dis 11(11), 1018.

Cirulli, E.T., Goldstein, D.B., 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11(6), 415-425.
Converge, 2015. Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature 523(7562), 588-591.

Cruchaga, C., Karch, C.M., Jin, S.C., Benitez, B.A., Cai, Y., Guerreiro, R., Harari, O., Norton, J., Budde, J., Bertelsen, S., Jeng, A.T., Cooper, B., Skorupa, T., Carrell, D., Levitch, D., Hsu, S., Choi, J., Ryten, M., Sassi, C., Bras, J., Gibbs, R.J., Hernandez, D.G., Lupton, M.K., Powell, J., Forabosco, P., Ridge, P.G., Corcoran, C.D., Tschanz, J.T., Norton, M.C., Munger, R.G., Schmutz, C., Leary, M., Demirci, F.Y., Bamne, M.N., Wang, X., Lopez, O.L., Ganguli, M., Medway, C., Turton, J., Lord, J., Braae, A., Barber, I., Brown, K., Alzheimer's Research, U.K.C., Pastor, P., Lorenzo-Betancor, O., Brkanac, Z., Scott, E., Topol, E., Morgan, K., Rogaeva, E., Singleton, A., Hardy, J., Kamboh, M.I., George-Hyslop, P.S., Cairns, N., Morris, J.C., Kauwe, J.S.K., Goate, A.M., 2014. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature 505(7484), 550-554.

Deng, H., Wu, Y., Jankovic, J., 2015. The EIF4G1 gene and Parkinson's disease. Acta Neurol Scand 132(2), 73-78.

Dhungel, N., Eleuteri, S., Li, L.B., Kramer, N.J., Chartron, J.W., Spencer, B., Kosberg, K., Fields, J.A., Stafa, K., Adame, A., Lashuel, H., Frydman, J., Shen, K., Masliah, E., Gitler, A.D., 2015. Parkinson's disease genes VPS35 and EIF4G1 interact genetically and converge on alpha-synuclein. Neuron 85(1), 76-87.

ENCODE, 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489(7414), 57-74.

Erickson, T., French, C.R., Waskiewicz, A.J., 2010. Meis1 specifies positional information in the retina and tectum to organize the zebrafish visual system. Neural Dev 5, 22.
Fallerini, C., Picchiotti, N., Baldassarri, M., Zguro, K., Daga, S., Fava, F., Benetti, E., Amitrano, S., Bruttini, M., Palmieri, M., Croci, S., Lista, M., Beligni, G., Valentino, F., Meloni, I., Tanfoni, M., Minnai, F., Colombo, F., Cabri, E., Fratelli, M., Gabbi, C., Mantovani, S., Frullanti, E., Gori, M., Crawley, F.P., Butler-Laporte, G., Richards, B., Zeberg, H., Lipcsey, M., Hultstrom, M., Ludwig, K.U., Schulte, E.C., Pairo-Castineira, E., Baillie, J.K., Schmidt, A., Frithiof, R., HGI, W.W.W.G.W.t., Gen, O.C., Study, G.-C.M., Mari, F., Renieri, A., Furini, S., 2022. Common, low-frequency, rare, and ultra-rare coding variants contribute to COVID-19 severity. Hum Genet 141(1), 147-173.

Fu, J.M.S., F.K.; Peng, M.; Brand, H.; Collins, R.L.; Dong, S.; Klei, L.; Stevens, C.R.; Cusick, C.; Babadi, M.; et al. , 2021a. Rare coding variation illuminates the allelic architecture, risk genes, cellular expression patterns, and phenotypic context of autism. medRxiv.

Fu, J.M.S., F.K.; Peng, M.; Brand, H.; Collins, R.L.; Dong, S.; Klei, L.; Stevens, C.R.; et al., 2021b. Rare coding variation illuminates the allelic architecture, risk genes, cellular expression patterns, and phenotypic context of autism.

Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., Nickerson, D.A., Bamshad, M.J., Project, N.E.S., Akey, J.M., 2013. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 493(7431), 216-220.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., Liu, D.R., 2017. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature 551(7681), 464-471.

Gaziano, J.M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J., Whitbourne, S., Deen, J., Shannon, C., Humphries, D., Guarino, P., Aslan, M., Anderson, D., LaFleur, R., Hammond, T., Schaa, K., Moser, J., Huang, G., Muralidhar, S., Przygodzki, R., O'Leary, T.J., 2016. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 70, 214-223.

Gibson, G., 2012. Rare and common variants: twenty arguments. Nat Rev Genet 13(2), 135-145.

Grinde, K.E., Arbet, J., Green, A., O'Connell, M., Valcarcel, A., Westra, J., Tintle, N., 2017. Illustrating, Quantifying, and Correcting for Bias in Post-hoc Analysis of Gene-Based Rare Variant Tests of Association. Front Genet 8, 117.

GTEx, 2013. The Genotype-Tissue Expression (GTEx) project. Nat Genet 45(6), 580-585. Guella, I., Solda, G., Cilia, R., Pezzoli, G., Asselta, R., Duga, S., Goldwurm, S., 2012. The Asp620asn mutation in VPS35 is not a common cause of familial Parkinson's disease. Mov Disord 27(6), 800-801.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A., Ritchie, G.R., Xue, Y., Asimit, J., Nsubuga, R.N., Young, E.H., Pomilla, C., Kivinen, K., Rockett, K., Kamali, A., Doumatey, A.P., Asiki, G., Seeley, J., Sisay-Joof, F., Jallow, M., Tollman, S., Mekonnen, E., Ekong, R., Oljira, T., Bradman, N., Bojang, K., Ramsay, M., Adeyemo, A., Bekele, E., Motala, A., Norris, S.A., Pirie, F., Kaleebu, P., Kwiatkowski, D., Tyler-Smith, C., Rotimi, C., Zeggini, E., Sandhu, M.S., 2015. The African Genome Variation Project shapes medical genetics in Africa. Nature 517(7534), 327-332.

Hanna, R.E., Hegde, M., Fagre, C.R., DeWeirdt, P.C., Sangree, A.K., Szegletes, Z., Griffith, A., Feeley, M.N., Sanson, K.R., Baidi, Y., Koblan, L.W., Liu, D.R., Neal, J.T., Doench, J.G., 2021. Massively parallel assessment of human variants with base editor screens. Cell 184(4), 1064-1080 e1020.

Hanss, Z., Larsen, S.B., Antony, P., Mencke, P., Massart, F., Jarazo, J., Schwamborn, J.C., Barbuti, P.A., Mellick, G.D., Kruger, R., 2021. Mitochondrial and Clearance Impairment in p.D620N VPS35 Patient-Derived Neurons. Mov Disord 36(3), 704-715.

Hopfner, F., Schulte, E.C., Mollenhauer, B., Bereznai, B., Knauf, F., Lichtner, P., Zimprich, A., Haubenberger, D., Pirker, W., Brucke, T., Peters, A., Gieger, C., Kuhlenbaumer, G., Trenkwalder, C., Winkelmann, J., 2013. The role of SCARB2 as susceptibility factor in Parkinson's disease. Mov Disord 28(4), 538-540.

Huttenlocher, J., Kruger, R., Capetian, P., Lohmann, K., Brockmann, K., Csoti, I., Klein, C., Berg, D., Gasser, T., Bonin, M., Riess, O., Bauer, P., 2015. EIF4G1 is neither a strong nor a common risk factor for Parkinson's disease: evidence from large European cohorts. J Med Genet 52(1), 37-41.

Initiative, C.-H.G., 2021. Mapping the human genetic architecture of COVID-19. Nature 600(7889), 472-477.

International Parkinson Disease Genomics, C., 2020. Ten Years of the International Parkinson Disease Genomics Consortium: Progress and Next Steps. J Parkinsons Dis 10(1), 19-30.

Iwaki, H., Blauwendraat, C., Makarious, M.B., Bandres-Ciga, S., Leonard, H.L., Gibbs, J.R., Hernandez, D.G., Scholz, S.W., Faghri, F., International Parkinson's Disease Genomics, C., Nalls, M.A., Singleton, A.B., 2020. Penetrance of Parkinson's Disease in LRRK2 p.G2019S Carriers Is Modified by a Polygenic Risk Score. Mov Disord 35(5), 774-780.

Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J., de Bakker, P.I., 2008. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24(24), 2938-2939.

Jonsson, T., Atwal, J.K., Steinberg, S., Snaedal, J., Jonsson, P.V., Bjornsson, S., Stefansson, H., Sulem, P., Gudbjartsson, D., Maloney, J., Hoyte, K., Gustafson, A., Liu, Y., Lu, Y., Bhangale, T., Graham, R.R., Huttenlocher, J., Bjornsdottir, G., Andreassen, O.A., Jonsson, E.G., Palotie, A., Behrens, T.W., Magnusson, O.T., Kong, A., Thorsteinsdottir, U., Watts, R.J., Stefansson, K., 2012. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature 488(7409), 96-99.

Kazazian, H.H., Jr., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G., Antonarakis, S.E., 1988. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 332(6160), 164-166.

Kendall, K.M., Van Assche, E., Andlauer, T.F.M., Choi, K.W., Luykx, J.J., Schulte, E.C., Lu, Y., 2021. The genetic basis of major depression. Psychol Med, 1-14.

Kesselmeier, M., Lorenzo Bermejo, J., 2017. Robust logistic regression to narrow down the winner's curse for rare and recessive susceptibility variants. Brief Bioinform 18(6), 962-972.

Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T., Kathiresan, S., 2018. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 50(9), 1219-1224.

Klein, J.C., Agarwal, V., Inoue, F., Keith, A., Martin, B., Kircher, M., Ahituv, N., Shendure, J., 2020. A systematic evaluation of the design and context dependencies of massively parallel reporter assays. Nat Methods 17(11), 1083-1091.

Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T., Bracken, M.B., Ferris, F.L., Ott, J., Barnstable, C., Hoh, J., 2005. Complement factor H polymorphism in age-related macular degeneration. Science 308(5720), 385-389.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., Liu, D.R., 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533(7603), 420-424.

Kondrashov, A.S., 2003. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. Hum Mutat 21(1), 12-27.

Kumar, K.R., Weissbach, A., Heldmann, M., Kasten, M., Tunc, S., Sue, C.M., Svetel, M., Kostic, V.S., Segura-Aguilar, J., Ramirez, A., Simon, D.K., Vieregge, P., Munte, T.F., Hagenah, J., Klein, C., Lohmann, K., 2012. Frequency of the D620N mutation in VPS35 in Parkinson disease. Arch Neurol 69(10), 1360-1364.

Lesage, S., Condroyer, C., Klebe, S., Lohmann, E., Durif, F., Damier, P., Tison, F., Anheim, M., Honore, A., Viallet, F., Bonnet, A.M., Ouvrard-Hernandez, A.M., Vidailhet, M., Durr, A., Brice, A., French Parkinson's Disease Genetics Study, G., 2012. EIF4G1 in familial Parkinson's disease: pathogenic mutations or rare benign variants? Neurobiol Aging 33(9), 2233 e2231-2233 e2235.

Loos, R.J.F., 2020. 15 years of genome-wide association studies and no signs of slowing down. Nat Commun 11(1), 5900.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., Pendlington, Z.M., Welter, D., Burdett, T., Hindorff, L., Flicek, P., Cunningham, F., Parkinson, H., 2017. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res 45(D1), D896-D901.

Madsen, B.E., Browning, S.R., 2009. A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet 5(2), e1000384.

Maher, B., 2008. Personal genomes: The case of the missing heritability. Nature 456(7218), 18-21.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.,

McCarroll, S.A., Visscher, P.M., 2009. Finding the missing heritability of complex diseases. Nature 461(7265), 747-753.

Mardis, E.R., 2011. A decade's perspective on DNA sequencing technology. Nature 470(7333), 198-203.

Marigorta, U.M., Rodriguez, J.A., Gibson, G., Navarro, A., 2018. Replicability and Prediction: Lessons and Challenges from GWAS. Trends Genet 34(7), 504-517.
Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., Daly, M.J., 2019. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 51(4), 584-591.

McGuire, A.L., Gabriel, S., Tishkoff, S.A., Wonkam, A., Chakravarti, A., Furlong, E.E.M., Treutlein, B., Meissner, A., Chang, H.Y., Lopez-Bigas, N., Segal, E., Kim, J.S., 2020. The road ahead in genetics and genomics. Nat Rev Genet 21(10), 581-596.

Mills, M.C., Rahal, C., 2020. The GWAS Diversity Monitor tracks diversity by disease in real time. Nat Genet 52(3), 242-243.

Mullins, N., Forstner, A.J., O'Connell, K.S., Coombes, B., Coleman, J.R.I., Qiao, Z., Als, T.D., Bigdeli, T.B., Borte, S., Bryois, J., Charney, A.W., Drange, O.K., Gandal, M.J., Hagenaars, S.P., Ikeda, M., Kamitaki, N., Kim, M., Krebs, K., Panagiotaropoulou, G., Schilder, B.M., Sloofman, L.G., Steinberg, S., Trubetskoy, V., Winsvold, B.S., Won, H.H., Abramova, L., Adorjan, K., Agerbo, E., Al Eissa, M., Albani, D., Alliey-Rodriguez, N., Anjorin, A., Antilla, V., Antoniou, A., Awasthi, S., Baek, J.H., Baekvad-Hansen, M., Bass, N., Bauer, M., Beins, E.C., Bergen, S.E., Birner, A., Bocker Pedersen, C., Boen, E., Boks, M.P., Bosch, R., Brum, M., Brumpton, B.M., Brunkhorst-Kanaan, N., Budde, M., Bybjerg-Grauholm, J., Byerley, W., Cairns, M., Casas, M., Cervantes, P., Clarke, T.K., Cruceanu, C., Cuellar-Barboza, A., Cunningham, J., Curtis, D., Czerski, P.M., Dale, A.M., Dalkner, N., David, F.S., Degenhardt, F., Djurovic, S., Dobbyn, A.L., Douzenis, A., Elvsashagen, T., Escott-Price, V., Ferrier, I.N., Fiorentino, A., Foroud, T.M., Forty, L., Frank, J., Frei, O., Freimer, N.B., Frisen, L., Gade, K., Garnham, J., Gelernter, J., Giortz Pedersen, M., Gizer, I.R., Gordon, S.D., Gordon-Smith, K., Greenwood, T.A., Grove, J., Guzman-Parra, J., Ha, K., Haraldsson, M., Hautzinger, M., Heilbronner, U., Hellgren, D., Herms, S., Hoffmann, P., Holmans, P.A., Huckins, L., Jamain, S., Johnson, J.S., Kalman, J.L., Kamatani, Y., Kennedy, J.L., Kittel-Schneider, S., Knowles, J.A., Kogevinas, M., Koromina, M., Kranz, T.M., Kranzler, H.R., Kubo, M., Kupka, R., Kushner, S.A., Lavebratt, C., Lawrence, J., Leber, M., Lee, H.J., Lee, P.H., Levy, S.E., Lewis, C., Liao, C., Lucae, S., Lundberg, M., MacIntyre, D.J., Magnusson, S.H., Maier, W., Maihofer, A., Malaspina, D., Maratou, E., Martinsson, L., Mattheisen, M., McCarroll, S.A., McGregor, N.W., McGuffin, P., McKay, J.D., Medeiros, H., Medland, S.E., Millischer, V., Montgomery, G.W., Moran, J.L., Morris, D.W., Muhleisen, T.W., O'Brien, N., O'Donovan, C., Olde Loohuis, L.M., Oruc, L., Papiol, S., Pardinas, A.F., Perry, A., Pfennig, A., Porichi, E., Potash, J.B., Quested, D., Raj, T., Rapaport, M.H., DePaulo, J.R., Regeer, E.J., Rice, J.P., Rivas, F., Rivera, M., Roth, J., Roussos, P., Ruderfer, D.M., Sanchez-Mora, C., Schulte, E.C., Senner, F., Sharp, S., Shilling, P.D., Sigurdsson, E., Sirignano, L., Slaney, C., Smeland, O.B., Smith, D.J., Sobell, J.L., Soholm Hansen, C., Soler Artigas, M., Spijker, A.T., Stein, D.J., Strauss, J.S., Swiatkowska, B., Terao, C., Thorgeirsson, T.E., Toma, C., Tooney, P., Tsermpini, E.E., Vawter, M.P., Vedder, H., Walters, J.T.R., Witt, S.H., Xi, S., Xu, W., Yang, J.M.K., Young, A.H., Young, H., Zandi, P.P., Zhou, H., Zillich, L., Psychiatry, H.A.-I., Adolfsson, R., Agartz, I., Alda, M., Alfredsson, L., Babadjanova, G., Backlund,

L., Baune, B.T., Bellivier, F., Bengesser, S., Berrettini, W.H., Blackwood, D.H.R., Boehnke, M., Borglum, A.D., Breen, G., Carr, V.J., Catts, S., Corvin, A., Craddock, N., Dannlowski, U., Dikeos, D., Esko, T., Etain, B., Ferentinos, P., Frye, M., Fullerton, J.M., Gawlik, M., Gershon, E.S., Goes, F.S., Green, M.J., Grigoroiu-Serbanescu, M., Hauser, J., Henskens, F., Hillert, J., Hong, K.S., Hougaard, D.M., Hultman, C.M., Hveem, K., Iwata, N., Jablensky, A.V., Jones, I., Jones, L.A., Kahn, R.S., Kelsoe, J.R., Kirov, G., Landen, M., Leboyer, M., Lewis, C.M., Li, Q.S., Lissowska, J., Lochner, C., Loughland, C., Martin, N.G., Mathews, C.A., Mayoral, F., McElroy, S.L., McIntosh, A.M., McMahon, F.J., Melle, I., Michie, P., Milani, L., Mitchell, P.B., Morken, G., Mors, O., Mortensen, P.B., Mowry, B., Muller-Myhsok, B., Myers, R.M., Neale, B.M., Nievergelt, C.M., Nordentoft, M., Nothen, M.M., O'Donovan, M.C., Oedegaard, K.J., Olsson, T., Owen, M.J., Paciga, S.A., Pantelis, C., Pato, C., Pato, M.T., Patrinos, G.P., Perlis, R.H., Posthuma, D., Ramos-Quiroga, J.A., Reif, A., Reininghaus, E.Z., Ribases, M., Rietschel, M., Ripke, S., Rouleau, G.A., Saito, T., Schall, U., Schalling, M., Schofield, P.R., Schulze, T.G., Scott, L.J., Scott, R.J., Serretti, A., Shannon Weickert, C., Smoller, J.W., Stefansson, H., Stefansson, K., Stordal, E., Streit, F., Sullivan, P.F., Turecki, G., Vaaler, A.E., Vieta, E., Vincent, J.B., Waldman, I.D., Weickert, T.W., Werge, T., Wray, N.R., Zwart, J.A., Biernacka, J.M., Nurnberger, J.I., Cichon, S., Edenberg, H.J., Stahl, E.A., McQuillin, A., Di Florio, A., Ophoff, R.A., Andreassen, O.A., 2021. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet 53(6), 817-829.

Nachman, M.W., Crowell, S.L., 2000. Estimate of the mutation rate per nucleotide in humans. Genetics 156(1), 297-304.

Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., St Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S.A., Fraser, D., Warren, L., Aponte, J., Zawistowski, M., Liu, X., Zhang, H., Zhang, Y., Li, J., Li, Y., Li, L., Woollard, P., Topp, S., Hall, M.D., Nangle, K., Wang, J., Abecasis, G., Cardon, L.R., Zollner, S., Whittaker, J.C., Chissoe, S.L., Novembre, J., Mooser, V., 2012. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science 337(6090), 100-104.
Ng, P.C., Henikoff, S., 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31(13), 3812-3814.

Nichols, N., Bras, J.M., Hernandez, D.G., Jansen, I.E., Lesage, S., Lubbe, S., Singleton, A.B., International Parkinson's Disease Genomics, C., 2015. EIF4G1 mutations do not cause Parkinson's disease. Neurobiol Aging 36(8), 2444 e2441-2444.

Niu, M., Zhao, F., Bondelid, K., Siedlak, S.L., Torres, S., Fujioka, H., Wang, W., Liu, J., Zhu, X., 2021. VPS35 D620N knockin mice recapitulate cardinal features of Parkinson's disease. Aging Cell 20(5), e13347.

Palmer, D.S.H., D.P.; Chapman, S.B.; Adolfsson, R.; Bass, N.; Blackwood, D.; Boks, M.P.M.; Chen, C.Y.; et al., 2020. Exome sequencing in bipolar disorder reveals shared risk gene AKAP11 with schizophrenia.

Pardinas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., Han, J., Hubbard, L., Lynham, A., Mantripragada, K., Rees, E., MacCabe, J.H., McCarroll, S.A., Baune, B.T., Breen, G., Byrne, E.M., Dannlowski, U., Eley, T.C., Hayward, C., Martin, N.G., McIntosh, A.M., Plomin, R., Porteous, D.J., Wray, N.R., Caballero, A., Geschwind, D.H., Huckins, L.M.,

Ruderfer, D.M., Santiago, E., Sklar, P., Stahl, E.A., Won, H., Agerbo, E., Als, T.D., Andreassen, O.A., Baekvad-Hansen, M., Mortensen, P.B., Pedersen, C.B., Borglum, A.D., Bybjerg-Grauholm, J., Djurovic, S., Durmishi, N., Pedersen, M.G., Golimbet, V., Grove, J., Hougaard, D.M., Mattheisen, M., Molden, E., Mors, O., Nordentoft, M., Pejovic-Milovancevic, M., Sigurdsson, E., Silagadze, T., Hansen, C.S., Stefansson, K., Stefansson, H., Steinberg, S., Tosato, S., Werge, T., Consortium, G., Consortium, C., Collier, D.A., Rujescu, D., Kirov, G., Owen, M.J., O'Donovan, M.C., Walters, J.T.R., 2018. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet 50(3), 381-389.

Peng, M., Wamsley, B., Elkins, A.G., Geschwind, D.H., Wei, Y., Roeder, K., 2021. Cell type hierarchy reconstruction via reconciliation of multi-resolution cluster tree. Nucleic Acids Res 49(16), e91.

PGC, 2019. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell 179(7), 1469-1482 e1411.

Pritchard, J.K., Cox, N.J., 2002. The allelic architecture of human disease genes: common disease-common variant...or not? Hum Mol Genet 11(20), 2417-2423.

Ramensky, V., Bork, P., Sunyaev, S., 2002. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30(17), 3894-3900.

Rivas, M.A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C.K., Boucher, G., Ripke, S., Ellinghaus, D., Burtt, N., Fennell, T., Kirby, A., Latiano, A., Goyette, P., Green, T., Halfvarson, J., Haritunians, T., Korn, J.M., Kuruvilla, F., Lagace, C., Neale, B., Lo, K.S., Schumm, P., Torkvist, L., National Institute of, D., Digestive Kidney Diseases Inflammatory Bowel Disease Genetics, C., United Kingdom Inflammatory Bowel Disease Genetics, C., International Inflammatory Bowel Disease Genetics, C., Dubinsky, M.C., Brant, S.R., Silverberg, M.S., Duerr, R.H., Altshuler, D., Gabriel, S., Lettre, G., Franke, A., D'Amato, M., McGovern, D.P., Cho, J.H., Rioux, J.D., Xavier, R.J., Daly, M.J., 2011. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 43(11), 1066-1073.

Roach, J.C., Glusman, G., Smit, A.F., Huff, C.D., Hubley, R., Shannon, P.T., Rowen, L., Pant, K.P., Goodman, N., Bamshad, M., Shendure, J., Drmanac, R., Jorde, L.B., Hood, L., Galas, D.J., 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328(5978), 636-639.

Rozenblatt-Rosen, O., Stubbington, M.J.T., Regev, A., Teichmann, S.A., 2017. The Human Cell Atlas: from vision to reality. Nature 550(7677), 451-453.

Satterstrom, F.K., Walters, R.K., Singh, T., Wigdor, E.M., Lescai, F., Demontis, D., Kosmicki, J.A., Grove, J., Stevens, C., Bybjerg-Grauholm, J., Baekvad-Hansen, M., Palmer, D.S., Maller, J.B., i, P.-B.C., Nordentoft, M., Mors, O., Robinson, E.B., Hougaard, D.M., Werge, T.M., Bo Mortensen, P., Neale, B.M., Borglum, A.D., Daly, M.J., 2019. Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. Nat Neurosci 22(12), 1961-1965.

Schizophrenia Working Group of the Psychiatric Genomics, C., 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511(7510), 421-427.

Schormair, B., Kemlink, D., Roeske, D., Eckstein, G., Xiong, L., Lichtner, P., Ripke, S., Trenkwalder, C., Zimprich, A., Stiasny-Kolster, K., Oertel, W., Bachmann, C.G., Paulus, W., Hogl, B., Frauscher, B., Gschliesser, V., Poewe, W., Peglau, I., Vodicka, P., Vavrova, J., Sonka, K., Nevsimalova, S., Montplaisir, J., Turecki, G., Rouleau, G., Gieger, C., Illig, T., Wichmann, H.E., Holsboer, F., Muller-Myhsok, B., Meitinger, T., Winkelmann, J., 2008. PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. Nat Genet 40(8), 946-948.

Schormair, B., Zhao, C., Bell, S., Tilch, E., Salminen, A.V., Putz, B., Dauvilliers, Y., Stefani, A., Hogl, B., Poewe, W., Kemlink, D., Sonka, K., Bachmann, C.G., Paulus, W., Trenkwalder, C., Oertel, W.H., Hornyak, M., Teder-Laving, M., Metspalu, A., Hadjigeorgiou, G.M., Polo, O., Fietze, I., Ross, O.A., Wszolek, Z., Butterworth, A.S., Soranzo, N., Ouwehand, W.H., Roberts, D.J., Danesh, J., Allen, R.P., Earley, C.J., Ondo, W.G., Xiong, L., Montplaisir, J., Gan-Or, Z., Perola, M., Vodicka, P., Dina, C., Franke, A., Tittmann, L., Stewart, A.F.R., Shah, S.H., Gieger, C., Peters, A., Rouleau, G.A., Berger, K., Oexle, K., Di Angelantonio, E., Hinds, D.A., Muller-Myhsok, B., Winkelmann, J., andMe Research, T., group, D.s., 2017. Identification of novel risk loci for restless legs syndrome in genome-wide association studies in individuals of European ancestry: a meta-analysis. Lancet Neurol 16(11), 898-907.

Schulte, E.C., 2013. Genetic variants in restless legs syndrome and Parkinson's disease: the rare, the common and everything in between. Technical University Munich.

Schulte, E.C., Claussen, M.C., Jochim, A., Haack, T., Hartig, M., Hempel, M., Prokisch, H., Haun-Junger, U., Winkelmann, J., Hemmer, B., Forschler, A., Ilg, R., 2013a. Mitochondrial membrane protein associated neurodegenration: a novel variant of neurodegeneration with brain iron accumulation. Mov Disord 28(2), 224-227.

Schulte, E.C., Ellwanger, D.C., Dihanich, S., Manzoni, C., Stangl, K., Schormair, B., Graf, E., Eck, S., Mollenhauer, B., Haubenberger, D., Pirker, W., Zimprich, A., Brucke, T., Lichtner, P., Peters, A., Gieger, C., Trenkwalder, C., Mewes, H.W., Meitinger, T., Lewis, P.A., Klunemann, H.H., Winkelmann, J., 2014a. Rare variants in LRRK1 and Parkinson's disease. Neurogenetics 15(1), 49-57.

Schulte, E.C., Fukumori, A., Mollenhauer, B., Hor, H., Arzberger, T., Perneczky, R., Kurz, A., Diehl-Schmid, J., Hull, M., Lichtner, P., Eckstein, G., Zimprich, A., Haubenberger, D., Pirker, W., Brucke, T., Bereznai, B., Molnar, M.J., Lorenzo-Betancor, O., Pastor, P., Peters, A., Gieger, C., Estivill, X., Meitinger, T., Kretzschmar, H.A., Trenkwalder, C., Haass, C., Winkelmann, J., 2015a. Rare variants in beta-Amyloid precursor protein (APP) and Parkinson's disease. Eur J Hum Genet 23(10), 1328-1333.

Schulte, E.C., Knauf, F., Kemlink, D., Schormair, B., Lichtner, P., Gieger, C., Meitinger, T., Winkelmann, J., 2011. Variant screening of the coding regions of MEIS1 in patients with restless legs syndrome. Neurology 76(12), 1106-1108.

Schulte, E.C., Kousi, M., Tan, P.L., Tilch, E., Knauf, F., Lichtner, P., Trenkwalder, C., Hogl, B., Frauscher, B., Berger, K., Fietze, I., Hornyak, M., Oertel, W.H., Bachmann, C.G., Zimprich, A., Peters, A., Gieger, C., Meitinger, T., Muller-Myhsok, B., Katsanis, N., Winkelmann, J., 2014b. Targeted resequencing and systematic in vivo functional

testing identifies rare variants in MEIS1 as significant contributors to restless legs syndrome. Am J Hum Genet 95(1), 85-95.

Schulte, E.C., Kurz, A., Alexopoulos, P., Hampel, H., Peters, A., Gieger, C., Rujescu, D., Diehl-Schmid, J., Winkelmann, J., 2015b. Excess of rare coding variants in PLD3 in late- but not early-onset Alzheimer's disease. Hum Genome Var 2, 14028.

Schulte, E.C., Mollenhauer, B., Zimprich, A., Bereznai, B., Lichtner, P., Haubenberger, D., Pirker, W., Brucke, T., Molnar, M.J., Peters, A., Gieger, C., Trenkwalder, C., Winkelmann, J., 2012. Variants in eukaryotic translation initiation factor 4G1 in sporadic Parkinson's disease. Neurogenetics 13(3), 281-285.

Schulte, E.C., Stahl, I., Czamara, D., Ellwanger, D.C., Eck, S., Graf, E., Mollenhauer, B., Zimprich, A., Lichtner, P., Haubenberger, D., Pirker, W., Brucke, T., Bereznai, B., Molnar, M.J., Peters, A., Gieger, C., Muller-Myhsok, B., Trenkwalder, C., Winkelmann, J., 2013b. Rare variants in PLXNA4 and Parkinson's disease. PLoS One 8(11), e79145.

Schulte, E.C.W., J., 2015. Clinical Phenotype and Genetics of Restless Legs Syndrome, in: LeDoux, M.S. (Ed.), Second Edition ed. Elsevier, New York.

Schwarz, J.M., Rodelsperger, C., Schuelke, M., Seelow, D., 2010. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 7(8), 575-576.

Siitonen, A., Majounie, E., Federoff, M., Ding, J., Majamaa, K., Singleton, A.B., 2013. Mutations in EIF4G1 are not a common cause of Parkinson's disease. Eur J Neurol 20(4), e59.

Singh, T.N., B.M.; Daly, M., 2020. Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia.

Smeland, O.B., Frei, O., Dale, A.M., Andreassen, O.A., 2020. The polygenic architecture of schizophrenia - rethinking pathogenesis and nosology. Nat Rev Neurol 16(7), 366-379.

Smoller, J.W., 2017. A Quarter Century of Progress in Psychiatric Genetics. Harv Rev Psychiatry 25(6), 256-258.

Solyom, S., Kazazian, H.H., Jr., 2012. Mobile elements in the human genome: implications for disease. Genome Med 4(2), 12.

Song, C.X., Yi, C., He, C., 2012. Mapping recently identified nucleotide variants in the genome and transcriptome. Nat Biotechnol 30(11), 1107-1116.

Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J.R.I., Gaspar, H.A., de Leeuw, C.A., Steinberg, S., Pavlides, J.M.W., Trzaskowski, M., Byrne, E.M., Pers, T.H., Holmans, P.A., Richards, A.L., Abbott, L., Agerbo, E., Akil, H., Albani, D., Alliey-Rodriguez, N., Als, T.D., Anjorin, A., Antilla, V., Awasthi, S., Badner, J.A., Baekvad-Hansen, M., Barchas, J.D., Bass, N., Bauer, M., Belliveau, R., Bergen, S.E., Pedersen, C.B., Boen, E., Boks, M.P., Boocock, J., Budde, M., Bunney, W., Burmeister, M., Bybjerg-Grauholm, J., Byerley, W., Casas, M., Cerrato,

F., Cervantes, P., Chambert, K., Charney, A.W., Chen, D., Churchhouse, C., Clarke, T.K., Coryell, W., Craig, D.W., Cruceanu, C., Curtis, D., Czerski, P.M., Dale, A.M., de Jong, S., Degenhardt, F., Del-Favero, J., DePaulo, J.R., Djurovic, S., Dobbyn, A.L., Dumont, A., Elvsashagen, T., Escott-Price, V., Fan, C.C., Fischer, S.B., Flickinger, M., Foroud, T.M., Forty, L., Frank, J., Fraser, C., Freimer, N.B., Frisen, L., Gade, K., Gage, D., Garnham, J., Giambartolomei, C., Pedersen, M.G., Goldstein, J., Gordon, S.D., Gordon-Smith, K., Green, E.K., Green, M.J., Greenwood, T.A., Grove, J., Guan, W., Guzman-Parra, J., Hamshere, M.L., Hautzinger, M., Heilbronner, U., Herms, S., Hipolito, M., Hoffmann, P., Holland, D., Huckins, L., Jamain, S., Johnson, J.S., Jureus, A., Kandaswamy, R., Karlsson, R., Kennedy, J.L., Kittel-Schneider, S., Knowles, J.A., Kogevinas, M., Koller, A.C., Kupka, R., Lavebratt, C., Lawrence, J., Lawson, W.B., Leber, M., Lee, P.H., Levy, S.E., Li, J.Z., Liu, C., Lucae, S., Maaser, A., MacIntyre, D.J., Mahon, P.B., Maier, W., Martinsson, L., McCarroll, S., McGuffin, P., McInnis, M.G., McKay, J.D., Medeiros, H., Medland, S.E., Meng, F., Milani, L., Montgomery, G.W., Morris, D.W., Muhleisen, T.W., Mullins, N., Nguyen, H., Nievergelt, C.M., Adolfsson, A.N., Nwulia, E.A., O'Donovan, C., Loohuis, L.M.O., Ori, A.P.S., Oruc, L., Osby, U., Perlis, R.H., Perry, A., Pfennig, A., Potash, J.B., Purcell, S.M., Regeer, E.J., Reif, A., Reinbold, C.S., Rice, J.P., Rivas, F., Rivera, M., Roussos, P., Ruderfer, D.M., Ryu, E., Sanchez-Mora, C., Schatzberg, A.F., Scheftner, W.A., Schork, N.J., Shannon Weickert, C., Shehktman, T., Shilling, P.D., Sigurdsson, E., Slaney, C., Smeland, O.B., Sobell, J.L., Soholm Hansen, C., Spijker, A.T., St Clair, D., Steffens, M., Strauss, J.S., Streit, F., Strohmaier, J., Szelinger, S., Thompson, R.C., Thorgeirsson, T.E., Treutlein, J., Vedder, H., Wang, W., Watson, S.J., Weickert, T.W., Witt, S.H., Xi, S., Xu, W., Young, A.H., Zandi, P., Zhang, P., Zollner, S., e, Q.C., Consortium, B., Adolfsson, R., Agartz, I., Alda, M., Backlund, L., Baune, B.T., Bellivier, F., Berrettini, W.H., Biernacka, J.M., Blackwood, D.H.R., Boehnke, M., Borglum, A.D., Corvin, A., Craddock, N., Daly, M.J., Dannlowski, U., Esko, T., Etain, B., Frye, M., Fullerton, J.M., Gershon, E.S., Gill, M., Goes, F., Grigoroiu-Serbanescu, M., Hauser, J., Hougaard, D.M., Hultman, C.M., Jones, I., Jones, L.A., Kahn, R.S., Kirov, G., Landen, M., Leboyer, M., Lewis, C.M., Li, Q.S., Lissowska, J., Martin, N.G., Mayoral, F., McElroy, S.L., McIntosh, A.M., McMahon, F.J., Melle, I., Metspalu, A., Mitchell, P.B., Morken, G., Mors, O., Mortensen, P.B., Muller-Myhsok, B., Myers, R.M., Neale, B.M., Nimgaonkar, V., Nordentoft, M., Nothen, M.M., O'Donovan, M.C., Oedegaard, K.J., Owen, M.J., Paciga, S.A., Pato, C., Pato, M.T., Posthuma, D., Ramos-Quiroga, J.A., Ribases, M., Rietschel, M., Rouleau, G.A., Schalling, M., Schofield, P.R., Schulze, T.G., Serretti, A., Smoller, J.W., Stefansson, H., Stefansson, K., Stordal, E., Sullivan, P.F., Turecki, G., Vaaler, A.E., Vieta, E., Vincent, J.B., Werge, T., Nurnberger, J.I., Wray, N.R., Di Florio, A., Edenberg, H.J., Cichon, S., Ophoff, R.A., Scott, L.J., Andreassen, O.A., Kelsoe, J., Sklar, P., Bipolar Disorder Working Group of the Psychiatric Genomics, C., 2019. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat Genet 51(5), 793-803.

Stevenson, A., Akena, D., Stroud, R.E., Atwoli, L., Campbell, M.M., Chibnik, L.B., Kwobah, E., Kariuki, S.M., Martin, A.R., de Menil, V., Newton, C., Sibeko, G., Stein, D.J., Teferra, S., Zingela, Z., Koenen, K.C., 2019. Neuropsychiatric Genetics of African Populations-Psychosis (NeuroGAP-Psychosis): a case-control study protocol and GWAS in Ethiopia, Kenya, South Africa and Uganda. BMJ Open 9(2), e025469.

Strachan, T.R., 2011. Human Molecular Genetics. Garland Science, New York City, NY.

Sullivan, P.F., Geschwind, D.H., 2019. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. Cell 177(1), 162-183.

Tang, F.L., Erion, J.R., Tian, Y., Liu, W., Yin, D.M., Ye, J., Tang, B., Mei, L., Xiong, W.C., 2015. VPS35 in Dopamine Neurons Is Required for Endosome-to-Golgi Retrieval of Lamp2a, a Receptor of Chaperone-Mediated Autophagy That Is Critical for alpha-Synuclein Degradation and Prevention of Pathogenesis of Parkinson's Disease. J Neurosci 35(29), 10613-10628.

Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., Kang, H.M., Jordan, D., Leal, S.M., Gabriel, S., Rieder, M.J., Abecasis, G., Altshuler, D., Nickerson, D.A., Boerwinkle, E., Sunyaev, S., Bustamante, C.D., Bamshad, M.J., Akey, J.M., Broad, G.O., Seattle, G.O., Project, N.E.S., 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337(6090), 64-69.

Tolosa, E., Vila, M., Klein, C., Rascol, O., 2020. LRRK2 in Parkinson disease: challenges of clinical trials. Nat Rev Neurol 16(2), 97-107.

Vilarino-Guell, C., Chai, H., Keeling, B.H., Young, J.E., Rajput, A., Lynch, T., Aasly, J.O., Uitti, R.J., Wszolek, Z.K., Farrer, M.J., Lin, S.C., 2009. MEIS1 p.R272H in familial restless legs syndrome. Neurology 73(3), 243-245.

Vilarino-Guell, C., Wider, C., Ross, O.A., Dachsel, J.C., Kachergus, J.M., Lincoln, S.J., Soto-Ortolaza, A.I., Cobb, S.A., Wilhoite, G.J., Bacon, J.A., Behrouz, B., Melrose, H.L., Hentati, E., Puschmann, A., Evans, D.M., Conibear, E., Wasserman, W.W., Aasly, J.O., Burkhard, P.R., Djaldetti, R., Ghika, J., Hentati, F., Krygowska-Wajs, A., Lynch, T., Melamed, E., Rajput, A., Rajput, A.H., Solida, A., Wu, R.M., Uitti, R.J., Wszolek, Z.K., Vingerhoets, F., Farrer, M.J., 2011. VPS35 mutations in Parkinson disease. Am J Hum Genet 89(1), 162-167.

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., Yang, J., 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 101(1), 5-22.

Wang, W., Wang, X., Fujioka, H., Hoppel, C., Whone, A.L., Caldwell, M.A., Cullen, P.J., Liu, J., Zhu, X., 2016. Parkinson's disease-associated mutant VPS35 causes mitochondrial dysfunction by recycling DLP1 complexes. Nat Med 22(1), 54-63.

Waskiewicz, A.J., Rikhof, H.A., Hernandez, R.E., Moens, C.B., 2001. Zebrafish Meis functions to stabilize Pbx proteins and regulate hindbrain patterning. Development 128(21), 4139-4151.

Winkelmann, J., Czamara, D., Schormair, B., Knauf, F., Schulte, E.C., Trenkwalder, C., Dauvilliers, Y., Polo, O., Hogl, B., Berger, K., Fuhs, A., Gross, N., Stiasny-Kolster, K., Oertel, W., Bachmann, C.G., Paulus, W., Xiong, L., Montplaisir, J., Rouleau, G.A., Fietze, I., Vavrova, J., Kemlink, D., Sonka, K., Nevsimalova, S., Lin, S.C., Wszolek, Z., Vilarino-Guell, C., Farrer, M.J., Gschliesser, V., Frauscher, B., Falkenstetter, T., Poewe, W., Allen, R.P., Earley, C.J., Ondo, W.G., Le, W.D., Spieler, D., Kaffe, M., Zimprich, A., Kettunen, J., Perola, M., Silander, K., Cournu-Rebeix, I., Francavilla, M., Fontenille, C., Fontaine, B., Vodicka, P., Prokisch, H., Lichtner, P., Peppard, P., Faraco, J., Mignot, E.,

Gieger, C., Illig, T., Wichmann, H.E., Muller-Myhsok, B., Meitinger, T., 2011. Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. PLoS Genet 7(7), e1002171.

Winkelmann, J., Schormair, B., Lichtner, P., Ripke, S., Xiong, L., Jalilzadeh, S., Fulda, S., Putz, B., Eckstein, G., Hauk, S., Trenkwalder, C., Zimprich, A., Stiasny-Kolster, K., Oertel, W., Bachmann, C.G., Paulus, W., Peglau, I., Eisensehr, I., Montplaisir, J., Turecki, G., Rouleau, G., Gieger, C., Illig, T., Wichmann, H.E., Holsboer, F., Muller-Myhsok, B., Meitinger, T., 2007. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. Nat Genet 39(8), 1000-1006.

Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., Bacanu, S.A., Baekvad-Hansen, M., Beekman, A.F.T., Bigdeli, T.B., Binder, E.B., Blackwood, D.R.H., Bryois, J., Buttenschon, H.N., Bybjerg-Grauholm, J., Cai, N., Castelao, E., Christensen, J.H., Clarke, T.K., Coleman, J.I.R., Colodro-Conde, L., Couvy-Duchesne, B., Craddock, N., Crawford, G.E., Crowley, C.A., Dashti, H.S., Davies, G., Deary, I.J., Degenhardt, F., Derks, E.M., Direk, N., Dolan, C.V., Dunn, E.C., Eley, T.C., Eriksson, N., Escott-Price, V., Kiadeh, F.H.F., Finucane, H.K., Forstner, A.J., Frank, J., Gaspar, H.A., Gill, M., Giusti-Rodriguez, P., Goes, F.S., Gordon, S.D., Grove, J., Hall, L.S., Hannon, E., Hansen, C.S., Hansen, T.F., Herms, S., Hickie, I.B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J.J., Hougaard, D.M., Hu, M., Hyde, C.L., Ising, M., Jansen, R., Jin, F., Jorgenson, E., Knowles, J.A., Kohane, I.S., Kraft, J., Kretzschmar, W.W., Krogh, J., Kutalik, Z., Lane, J.M., Li, Y., Li, Y., Lind, P.A., Liu, X., Lu, L., MacIntyre, D.J., MacKinnon, D.F., Maier, R.M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland, S.E., Mehta, D., Middeldorp, C.M., Mihailov, E., Milaneschi, Y., Milani, L., Mill, J., Mondimore, F.M., Montgomery, G.W., Mostafavi, S., Mullins, N., Nauck, M., Ng, B., Nivard, M.G., Nyholt, D.R., O'Reilly, P.F., Oskarsson, H., Owen, M.J., Painter, J.N., Pedersen, C.B., Pedersen, M.G., Peterson, R.E., Pettersson, E., Peyrot, W.J., Pistis, G., Posthuma, D., Purcell, S.M., Quiroz, J.A., Qvist, P., Rice, J.P., Riley, B.P., Rivera, M., Saeed Mirza, S., Saxena, R., Schoevers, R., Schulte, E.C., Shen, L., Shi, J., Shyn, S.I., Sigurdsson, E., Sinnamon, G.B.C., Smit, J.H., Smith, D.J., Stefansson, H., Steinberg, S., Stockmeier, C.A., Streit, F., Strohmaier, J., Tansey, K.E., Teismann, H., Teumer, A., Thompson, W., Thomson, P.A., Thorgeirsson, T.E., Tian, C., Traylor, M., Treutlein, J., Trubetskoy, V., Uitterlinden, A.G., Umbricht, D., Van der Auwera, S., van Hemert, A.M., Viktorin, A., Visscher, P.M., Wang, Y., Webb, B.T., Weinsheimer, S.M., Wellmann, J., Willemsen, G., Witt, S.H., Wu, Y., Xi, H.S., Yang, J., Zhang, F., eQtlgen, andMe, Arolt, V., Baune, B.T., Berger, K., Boomsma, D.I., Cichon, S., Dannlowski, U., de Geus, E.C.J., DePaulo, J.R., Domenici, E., Domschke, K., Esko, T., Grabe, H.J., Hamilton, S.P., Hayward, C., Heath, A.C., Hinds, D.A., Kendler, K.S., Kloiber, S., Lewis, G., Li, Q.S., Lucae, S., Madden, P.F.A., Magnusson, P.K., Martin, N.G., McIntosh, A.M., Metspalu, A., Mors, O., Mortensen, P.B., Muller-Myhsok, B., Nordentoft, M., Nothen, M.M., O'Donovan, M.C., Paciga, S.A., Pedersen, N.L., Penninx, B., Perlis, R.H., Porteous, D.J., Potash, J.B., Preisig, M., Rietschel, M., Schaefer, C., Schulze, T.G., Smoller, J.W., Stefansson, K., Tiemeier, H., Uher, R., Volzke, H., Weissman, M.M., Werge, T., Winslow, A.R., Lewis, C.M., Levinson, D.F., Breen, G., Borglum, A.D., Sullivan, P.F., Major Depressive Disorder Working Group of the Psychiatric Genomics, C., 2018. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 50(5), 668-681.

Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., Lin, X., 2011. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 89(1), 82-93.

Xiong, L., Catoire, H., Dion, P., Gaspar, C., Lafreniere, R.G., Girard, S.L., Levchenko, A., Riviere, J.B., Fiori, L., St-Onge, J., Bachand, I., Thibodeau, P., Allen, R., Earley, C., Turecki, G., Montplaisir, J., Rouleau, G.A., 2009. MEIS1 intronic risk haplotype associated with restless legs syndrome affects its mRNA and protein expression levels. Hum Mol Genet 18(6), 1065-1074.

Zavodszky, E., Seaman, M.N., Moreau, K., Jimenez-Sanchez, M., Breusegem, S.Y., Harbour, M.E., Rubinsztein, D.C., 2014. Mutation in VPS35 associated with Parkinson's disease impairs WASH complex association and inhibits autophagy. Nat Commun 5, 3828.

Zech, M., Nubling, G., Castrop, F., Jochim, A., Schulte, E.C., Mollenhauer, B., Lichtner, P., Peters, A., Gieger, C., Marquardt, T., Vanier, M.T., Latour, P., Klunemann, H., Trenkwalder, C., Diehl-Schmid, J., Perneczky, R., Meitinger, T., Oexle, K., Haslinger, B., Lorenzl, S., Winkelmann, J., 2013. Niemann-Pick C disease gene mutations and age-related neurodegenerative disorders. PLoS One 8(12), e82879.

Zeggini, E., Gloyn, A.L., Barton, A.C., Wain, L.V., 2019. Translational genomics and precision medicine: Moving from the lab to the clinic. Science 365(6460), 1409-1413.

Zhang, Y., Qi, G., Park, J.H., Chatterjee, N., 2018. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. Nat Genet 50(9), 1318-1326.

Zhao, Y., Ho, P., Prakash, K.M., Foo, J.N., Liu, J.J., Au, W.L., Tan, L.C., Tan, E.K., 2013. Analysis of EIF4G1 in Parkinson's disease among Asians. Neurobiol Aging 34(4), 1311 e1315-1316.

Zimprich, A., Benet-Pages, A., Struhal, W., Graf, E., Eck, S.H., Offman, M.N., Haubenberger, D., Spielberger, S., Schulte, E.C., Lichtner, P., Rossle, S.C., Klopp, N., Wolf, E., Seppi, K., Pirker, W., Presslauer, S., Mollenhauer, B., Katzenschlager, R., Foki, T., Hotzy, C., Reinthaler, E., Harutyunyan, A., Kralovics, R., Peters, A., Zimprich, F., Brucke, T., Poewe, W., Auff, E., Trenkwalder, C., Rost, B., Ransmayr, G., Winkelmann, J., Meitinger, T., Strom, T.M., 2011. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. Am J Hum Genet 89(1), 168-175.

7 COMPLETE LIST OF OWN PUBLICATIONS

1. **Schulte EC***, Kousi M*, Tan PL, Tilch E, Knauf F, Lichtner P, Trenkwalder C, Högl B, Frauscher B, Berger K, Fietze I, Hornyak M, Oertel WH, Bachmann CG, Zimprich A, Peters A, Gieger C, Meitinger T, Müller-Myhsok B, Katsanis N, Winkelmann J. Targeted resequencing and systematic in vivo functional testing identifies rare variants in *MEIS1* as significant contributors to restless legs syndrome. Am J Hum Genet 2014, 95:85-95. PMID: 24995868 (IF 2014: 12,6)

2. Heilbronner U, Streit F, Vogl T, Senner F, Schaupp SK, Reich-Erkelenz D, Papiol S, Oraki Kohshour M, Klöhn-Saghatolislam F, Kalman JL, Heilbronner M, Gade K, Comes AL, Budde M, Andlauer TFM, Anderson-Schmidt H, Adorjan K, Stürmer T, Loerbroks A, Amelang M, Poisel E, Foo J, Heilmann-Heimbach S, Forstner AJ, Degenhardt F, Zimmermann J, Wiltfang J, von Hagen M, Spitzer C, Schmauss M, Reininghaus E, Reimer J, Konrad C, Juckel G, Lang FU, Jäger M, Figge C, Fallgatter AJ, Dietrich DE, Dannlowski U, Baune BT, Arolt V, Anghelescu IG, Nöthen MM, Witt SH, Andreassen OA, Chen CH, Falkai P, Rietschel M, Schulze TG, **Schulte EC.** Interplay between the genetics of personality traits, severe psychiatric disorders and COVID-19 host genetics in the susceptibility to SARS-CoV-2 infection. BJ Psych Open 2021, 7:e188. PMID: 34659794 (IF 2021: 3,2)

3. **Schulte EC**, Sellner J. SARS-CoV-2 vaccination in multiple sclerosis: A clearer picture for the time point during CD20 depleting therapy. EBioMed 2021, 73:103635. PMID: 34656879 (IF 2020: 8,1)

4. **Schulte EC**, Hauer L, Kunz AB, Sellner J. Systematic Review of Cases of Acute Myelitis in Individuals with COVID-19. Eur J Neurol 2021, in press. (IF 2020: 6,1)

5. **Schulte EC,** Fukumori A, Mollenhauer B, Hor H, Arzberger T, Perneczky R, Kurz A, Diehl-Schmid J, Hüll M, Lichtner P, Eckstein G, Zimprich A, Haubenberger D, Pirker W, Brücke T, Bereznai B, Molnar MJ, Lorenzo-Betancor O, Pastor P, Peters A, Gieger C, Estivill X, Meitinger T, Kretzschmar HA, Trenkwalder C, Haass C, Winkelmann J. Rare variants in *beta-amyloid precursor protein* (*APP*) and Parkinson`s disease. Eur J Hum Genet 2015,23:1328-1333. PMID: 25604855. (IF 2015: 4,6)

6. **Schulte EC**, Kurz A, Alexopoulos P, Hampel H, Peters A, Gieger C, Rujescu D, Diehl-Schmid J, Winkelmann J. Excess of rare coding variants in *PLD3* in late- but not early-onset Alzheimer`s disease. Hum Genome Var 2015, 2:14028. PMID: 27081517. (IF 2015: 1,3)

7. **Schulte EC***, Altmaier E*, Berger HS, Do KT, Kastenmüller G, Wahl S, Adamski J, Peters A, Krumsiek J, Suhre K, Haslinger B, Ceballos-Baumann A, Gieger C, Winkelmann J. Alterations in Lipid and Inositol Metabolisms in Two Dopaminergic Disorders. PLoS One 2016: e0147129. PMID: 26808974. (IF 2016: 2,8)

8. **Schulte EC**, Schramm K, Schurmann C, Lichtner P, Herder C, Roden M, Gieger C, Peters A, Trenkwalder C, Högl B, Frauscher B, Berger K, Fietze I, Gross N, Stiasny-Kolster K, Oertel W, Bachmann CG, Paulus W, Zimprich A, Völzke H, Schminke U, Nauck M, Illig T, Meitinger T, Müller-Myhsok B, Prokisch H, Winkelmann J. Blood cis-eQTL analysis fails to identify novel association signals among sub-threshold candidates from genome-wide association studies in restless legs syndrome. PLoS One 2014, 9:e98092. PMID: 24875634 (IF 2014: 3,2)

9. **Schulte EC**, Ellwanger DC, Dihanich S, Manzoni C, Stangl K, Schormair B, Graf E, Eck S, Mollenhauer B, Haubenberger D, Pirker W, Zimprich A, Brücke T, Lichtner P, Peters A, Gieger C, Trenkwalder C, Mewes HW, Meitinger T, Lewis PA, Klünemann HH, Winkelmann J. Rare variants in LRRK1 in Parkinson`s disease. Neurogenetics 2014,15:49-57. PMID: 24241507 (IF 2014: 3,5)

10. **Schulte EC**, Kaffe M, Schormair B, Winkelmann J. Iron in Restless Legs Syndrome. Mov Disord Clin Pract 2014, 1:161-172. PMID: 30363981 (IF 2014: 1,0)

11. **Schulte EC**, Gross N, Slawik H, Winkelmann J. "Malignant restless legs syndrome"—A curse or a blessing ? Sleep Med 2014,15:155-156. PMID: 24325808 (IF 2014: 4,5)

12. **Schulte EC***, Claussen MC*, Jochim A, Haack T, Hartig M, Hempel M, Prokisch H, Haun-Jünger U, Winkelmann J, Hemmer B, Förschler A, Ilg R. Mitochondrial membrane protein associated neurodegeneration: a novel variant of neurodegeneration with brain iron accumulation. Mov Disord 2013,28:224-227. PMID: 23436634 (IF 2013: 5,6)

13. **Schulte EC**, Stahl I, Czamara D, Ellwanger DC, Eck S, Graf E, Mollenhauer B, Zimprich A, Lichtner P, Haubenberger D, Pirker W, Brücke T, Bereznai B, Molnar MJ, Peters A, Gieger C, Müller-Myhsok B, Trenkwalder C, Winkelmann J. Rare Variants in PLXNA4 and Parkinson`s Disease. PLoS One 2013,8:e79145. PMID: 24244438 (IF 2013: 3,5)

14. **Schulte EC**, Gross N, Slawik H, Winkelmann J. When restless legs syndrome turns malignant. Sleep Med 2013,14:575-577. PMID: 23643657 (IF 2013: 4,4)

15. **Schulte EC**, Mollenhauer B, Zimprich A, Bereznai B, Lichtner P, Haubenberger D, Pirker W, Brücke T, Molnar MJ, Peters A, Gieger C, Trenkwalder C, Winkelmann J. Variants in eurkaryotic translation initiation factor 4G1 in sporadic Parkinson's disease. Neurogenetics 2012,13:281-285. PMID: 22707335 (IF 2012: 3,4)

16. **Schulte EC**, Knauf F, Kemlink D, Schormair B, Gieger C, Lichtner P, Meitinger T, Winkelmann J. Variant Screening of the Coding Regions of MEIS1 in Patients with Restless Legs Syndrome. Neurology 2011,76:1106-1108. PMID: 21422461. (IF 2011: 8,3)

17. **Schulte EC**, Spieler D, Winkelmann J. Restless-legs syndrome and cardiovascular diseases, Der Nervenarzt 2011,82:1006-1011. PMID: 21174071 (IF 2011: 0,4)

18. **Schulte EC**, Winkelmann J. When Parkinson`s disease patients go to sleep. J Neurol 2011,258:S328-335. PMID: 21560064 (IF 2011: 2,9)

19. **Schulte EC**, Sauerbrei A, Hoffmann D, Zimmer C, Hemmer B, Mühlau M. Acylcovir Resistance in Herpes Simplex Encephalitis. Ann Neurol 2010,67:830-833. PMID: 20517946 (IF 2010: 8,2)

20. **Schulte EC**, Slawik H, Schüle R, Gunther T, Hüll M. Alterations in excitotoxicity and prostaglandin metabolism in a transgenic mouse model of Alzheimer's disease. Neurochem Int 2009,55:689-696. PMID: 19560505 (IF 2009: 4,2)

21. **Schulte EC**, Winkelmann J. Family- and Genome-Wide Association Studies of Restless Legs Syndrome in *Genetics of Sleep and Sleep Disorders*. (Ed.: Thorpy M, Shaw P, Tafti M). Cambridge University Press, Cambridge, UK. 2013.

22. **Schulte EC**, Winkelmann J. Clinical Phenotype and Genetics of Restless Legs Syndrome in *Movement Disorders: Genetics and Models.* (Ed.: LeDoux MS). Elsevier, San Diego, CA, USA. 2014.

23. **Schulte EC,** Winkelmann J. Update on RLS Genetics in *Movement Disorder Genetics*. (Ed.: Schneider SA, Bras J). Springer, Heidelberg, Germany. 2015.

24. **Schulte EC,** Oraki Koshour M, Khaitovich P, Schulze TG. Multiomic Approaches to Psychiatry in *Tasman's Psychiatry*, fifth edition. (Ed: Tasman A, Riba MB, Schulze TG, Ng CH, Alfonso CA, Lecic-Tosevski D, Kanba S, Alarcon RD, Ndetei DM). Springer, New York City, NY, USA. in press

25. Stahl K, Adorjan K, Anderson-Schmidt H, Budde M, Comes AL, Gade K, Heilbronner M, Kalman JL, Klöhn-Saghatolislam F, Oraki Kohshour M, Papiol S, Reich-Erkelenz D, Schaupp SK, **Schulte EC**, Senner F, Vogl T, Wiltfang J, Reininghaus E, Falkai P, Schulze TG, Bickeböller H, Heilbronner U. Stability over time of scores on psychiatric rating scales, questionnaires and cognitive tests in healthy controls. BJPsych Open 2022, 8, e55. PMID: 35232513 (IF 2021: 3,2)

26. Van Assche E, **Schulte EC**, Andreassen OA, Smeland OB, Luykx JJ. Editorial: Cross-disorder Genetics in Neuropsychiatry. Front Neurosci 2022, 16, 826300. PMID: 35221906 (IF 2021: 4,5)

27. Kalman JL, Yoshida T, Andlauer TFM, **Schulte EC**, Adorjan K, Alda M, Ardau R, Aubry JM, Brosch K, Budde M, Chillotti C, Czerski PM, DePaulo RJ, Forstner A, Goes FS, Grigoroiu-Serbanescu M, Grof P, Grotegerd D, Hahn T, Heilbronner M, Hasler R, Heilbronner U, Heilmann-Heimbach S, Kapelski P, Kato T, Kohshour MO, Meinert S, Meller T, Nenadić I, Nöthen MM, Novak T, Opel N, Pawlak J, Pfarr JK, Potash JB, Reich-Erkelenz D, Repple J, Richard-Lepouriel H, Rietschel M, Ringwald KG, Rouleau G, Schaupp S, Senner F, Severino G, Squassina A, Stein F, Stopkova P, Streit F, Thiel K, Thomas-Odenthal F, Turecki G, Twarowska-Hauser J, Winter A, Zandi PP, Kelsoe JR; Consortium on Lithium Genetics (ConLiGen), PsyCourse, Falkai P, Dannlowski U, Kircher T, Schulze TG, Papiol S. Investigating the phenotypic

and genetic associations between personality traits and suicidal behavior across major mental health diagnoses. Eur Arch Psychiatry Clin Neurosci 2022, online ahead of print. PMID: 35146571 (IF 2021: 5,3)

28. Kirchner SK, Lauseker M, Adorjan K, Anderson-Schmidt H, Anghelescu IG, Baune BT, Budde M, Dannlowski U, Dietrich DE, Fallgatter AJ, Falkai P, Figge C, Gade K, Heilbronner U, Hiendl L, Juckel G, Kalman JL, Klöhn-Saghatolislam F, Konrad C, Lang FU, Oraki Kohshour M, Papiol S, Reich-Erkelenz D, Reimer J, Reininghaus EZ, Schaupp SK, Schmauß M, Schmitt A, **Schulte EC**, Senner S, Spitzer C, Vogl T, Zimmermann J, Hasan A, Schulze TG, Senner F. Medication Adherence in a Cross-Diagnostic Sample of Patients From the Affective-to-Psychotic Spectrum: Results From the PsyCourse Study. Front Psychiatry 2022, 12, 713060. PMID: 35126191 (IF 2021: 3,8)

29. Fallerini C, Picchiotti N, Baldassarri M, Zguro K, Daga S, Fava F, Benetti E, Amitrano S, Bruttini M, Palmieri M, Croci S, Lista M, Beligni G, Valentino F, Meloni I, Tanfoni M, Minnai F, Colombo F, Cabri E, Fratelli M, Gabbi C, Mantovani S, Frullanti E, Gori M, Crawley FP, Butler-Laporte G, Richards B, Zeberg H, Lipcsey M, Hultström M, Ludwig KU, **Schulte EC**, Pairo-Castineira E, Baillie JK, Schmidt A, Frithiof R; WES/WGS Working Group Within the HGI; GenOMICC Consortium; GEN-COVID Multicenter Study, Mari F, Renieri A, Furini S. Common, low-frequency, rare, and ultra-rare coding variants contribute to COVID-19 severity. Hum Genet 2021, epub ahead of print. PMID 34889978. (IF 2020: 4,1)

30. Smigielski L, Papiol S, Theodoridou A, Heekeren K, Gerstenberg M, Wotruba D, Buechler R, Hoffmann P, Herms S, Adorjan K, Anderson-Schmidt H, Budde M, Comes AL, Gade K, Heilbronner M, Heilbronner U, Kalman JL, Klöhn-Saghatolislam F, Reich-Erkelenz D, Schaupp SK, **Schulte EC**, Senner F, Anghelescu IG, Arolt V, Baune BT, Dannlowski U, Dietrich DE, Fallgatter AJ, Figge C, Jäger M, Juckel G, Konrad C, Nieratschker V, Reimer J, Reininghaus E, Schmauß M, Spitzer C, von Hagen M, Wiltfang J, Zimmermann J, Gryaznova A, Flatau-Nagel L, Reitt M, Meyers M, Emons B, Haußleiter IS, Lang FU, Becker T, Wigand ME, Witt SH, Degenhardt F, Forstner AJ, Rietschel M, Nöthen MM, Andlauer TFM, Rössler W, Walitza S, Falkai P, Schulze TG, Grünblatt E. Polygenic risk scores across the extended psychosis spectrum. Transl Psychiatry 2021, epub ahead of print. PMID: 34836939. (IF 2020: 6,2)

31. Kalman JL, Papiol S, Grigoroiu-Serbanescu M, Adorjan K, Anderson-Schmidt H, Brosch K, Budde M, Comes AL, Gade K, Forstner A, Grotegerd D, Hahn T, Heilbronner M, Heilbronner U, Heilmann-Heimbach S, Klöhn-Saghatolislam F, Kohshour MO, Meinert S, Meller T, Mullins N, Nenadić I, Nöthen MM, Pfarr JK, Reich-Erkelenz D, Rietschel M, Ringwald KG, Schaupp S, **Schulte EC**, Senner F, Stein F, Streit F, Vogl T, Falkai P, Dannlowski U, Kircher T, Schulze TG, Andlauer TFM. Genetic risk for psychiatric illness is associated with the number of hospitalizations of bipolar disorder patients. J Affect Disord 2021, epub ahead of print. PMID: 34656040. (IF 2020: 4,8)

32. Nakanishi T, Pigazzini S, Degenhardt F, …, **Schulte EC**, et al. Age-dependent impact of the major common genetic risk factor for COVID-19 on severity and mortality. J Clin Invest 2021, epub ahead of print. PMID:34597274. (IF 2020: 14,8)

33. Niemi MEK, Karjalainen J, Liao RG, …, **Schulte EC**, et al. Mapping the human genetic architecture of COVID-19 by worldwide meta-analysis. Nature 2021, epub ahead of print. PMID:34237774. (IF 2020: 42,8)

34. Mullins N, Forstner AJ, O´Connell KS …, **Schulte EC**, et al. Genome-wide association study of over 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet 2021, 53(6):817-829. PMID:34002096. (IF 2020: 26,7)

35. Wendel B, Papiol S, Andlauer TFM, …, **Schulte EC**, et al. A genome-wide association study of the longitudinal course of executive functions. Transl Psychiatry 2021, 11(1):386. PMID: 34247186 (IF 2020: 5.3)

36. Kendall KM, Van Assche E, Andlauer TFM, Choi KW, Luykx JJ, **Schulte EC**, Lu Y. The genetic basis of major depression. Psychol Med 2021, epub ahead of print. PMID: 33682643. (IF 2020: 5,8)

37. Baldini T, Asioli GM, Romoli M, Carvalho Dias M, **Schulte EC**, Hauer L, Aguiar De Sousa D, Sellner J, Zini A. Cerebral venous thrombosis and severe acute respiratory syndrome coronavirus-2 infection: A systematic review and meta-analysis. Eur J Neurol 2021, epub ahead of print. PMID: 33426733. (IF 2020: 6.1)

38. Dwyer DB, Kalman JL, Budde M, …, **Schulte EC**, *et al*. An Investigation of Psychosis Subgroups With Prognostic Validation and Exploration of Genetic Underpinings: The PsyCourse Study. JAMA Psychiatry 2020, 77(5):523-533. PMID:32049274. (IF 2020: 17,5)

39. Comes AL, Czamara D, Adorjan K, …, **Schulte EC**, *et al*. The role of environmental stress and DNA methylation in the longitudinal course of bipolar disorder. Int J Bipolar Disord 2020, 8, 9. PMID: 32048126. (IF 2020: 3,0)

40. Schwabe I, Milaneschi Y, Gerring Z, …, **Schulte E**, *et al*. Unraveling the Genetic Architecture of Major Depressive Disorder: Merits and Pitfalls of the Approaches Used in Genome-Wide Association Studies. Psychol Med 2019, 49, 2646-2656. PMID:31559935 (IF 2019: 5,8)

41. Comes AL, Senner F, Budde M, …, **Schulte EC**, *et al*. The genetic relationship between educational attainment and cognitive performance in major psychiatric disorders. Transl Psychiatry 2019, 9, 210. PMID:31462630. (IF 2018: 5,2)

42. Hauer L, Pikija S, **Schulte EC**, *et al*. Cerebrovascular Manifestations of Herpes Simplex Virus Infection of the Central Nervous System: A Systematic Review. J Neuroinflammation 2019, 16, 19. PMID: 30696448 (IF 2018: 5,2)

43. Budde M, Anderson-Schmidt H, Gade K, …, **Schulte EC**, *et al*. A longitudinal approach to biological psychiatric research: The PsyCourse study. Am J Med Genet B Neuropsychiatr Genet 2019, 180, 89-102. PMID: 30070057 (IF 2018: 3,1)

44. Kalman JL, Papiol S, Forstner AJ, …, **Schulte EC**, *et al*. Investigating polygenic burden in age at disease onset in bipolar disorder: Findings from an international multicentric study. Bipolar Disord 2019, 21, 68-75. PMID: 29956436 (IF 2018: 4,9)

45. Wray NR, Ripke S, Mattheisen M, …, **Schulte EC**, *et al*. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 2018, 50, 668-681. PMID: 29700475. (IF 2018: 25,5)

46. Winkelmann J, Schormair B, Czamara D, Knauf F, **Schulte EC**, *et al.* Genome-wide association study identifies novel Restless Legs Syndrome susceptibility loci on 2p14 and 16q12.1.PLoS Genetics 2011,7:e1002171. PMID: 21779176 (IF 2011: 10,2)

47. Zimprich A, Benet-Pages A, Struhal W, Graf E, Eck SH, Offman MN, Haubenberger D, Spielberger S, **Schulte EC**, Lichtner P, Rossle SC, Klopp N, Wolf E, Seppi K, Pirker W, Presslauer S, Mollenhauer B, Katzenschlager R, Foki T, Hotzy C, Reinthaler E, Harutyunyan A, Kralovics R, Peters A, Zimprich F, Brücke T, Poewe W, Auff E, Trenkwalder C, Rost B, Ransmayr G, Winkelmann J, Meitinger T, Strom TM. Exome sequencing reveals mutations in the retromer protein VPS35 as cause for Parkinson`s disease. Am J Hum Genet 2011,89:168-175. PMID: 21763483 (IF 2011: 10,6)

48. Zech M, Nübling G, Castrop F, Jochim A, **Schulte EC**, Mollenhauer B, Lichtner P, Peters A, Gieger C, Marquardt T, Vanier MT, Latour P, Klünemann HH, Trenkwalder C, Diehl-Schmid J, Perneczky R, Meitinger T, Oexle K, Haslinger B, Lorenzl S, Winkelmann J. Niemann-Pick C disease gene mutations and age-related neurodegenerative disorders. PLoS One 2013,8:e82879. PMID: 24386122 (IF 2013: 3,5)

49. Hopfner F, **Schulte EC**, Mollenhauer B, Bereznai B, Knauf F, Lichtner P, Zimprich A, Haubenberger D, Pirker W, Brücke T, Peters A, Gieger C, Kuhlenbäumer G, Trenkwalder C, Winkelmann J. The role of SCARB2 as susceptibility factor in Parkinson`s disease. Mov Disord 2013,28:538-540. PMID: 23408458 (IF 2013: 5,6)

50. Wines-Samuelson M, **Schulte EC**, Smith MJ, Aoki C, Liu X, Kelleher RJ III, Shen J. Characterization of age-dependent and progressive cortical neuronal degeneration in presenilin conditional mutant mice. PLoS ONE 2010,5:e10195. PMID: 20419112 (IF 2010: 4,4)

51. Mullins N, Kang J, Campos AI, …, **Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium**, et al. Dissecting the Shared Genetic Architecture of Suicide Attempt, Psychiatric Disorders, and Known Risk Factors. Biol Psychiatry 2021, epub ahead of print. PMID: 34861974 (IF 2020: 12,1)

52. Giannakopoulou O, Lin K, Meng X, …, **Major Depressive Disorder Working Group of the Psychiatric Genomcis Consortium**. The Genetic Architecture of Depression in Individuals of East Asian Ancestry: A Genome-Wide Association Study. JAMA Psychiatry 2021, epub ahead of print. PMID: 34586374 (IF 2020: 17.5)

53. Ni G, Zeng J, Revez JA, …, **Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium**, et al. A Comparison of Ten Polygenic Score Methods for Psychiatric Disorders Applied Across Multiple Cohorts. Biol Psychiatry, 2021, online ahead of print. PMID: 34304866 (IF 2020: 12,1)

54. Blokland GAM, Grove J, Chen CY, …, **MDD Working Group of the Psychiatric Genomics Consortium**, et al. Sex-Dependent Shared and Nonshared Genetic Architecture Across Mood and Psychotic Disorders. Biol Psychiatry 2021, epub ahead of print. PMID: 34099189 (IF 2020: 12.1)

55. Shen X, Howard DM, Adams MJ, …, **MDD Working Group of the Psychiatric Genomics Consortium**, et al. A phenome-wide association and Mendelian Randomisation study of polygenic risk for depression in UK Biobank. Nat Commun, 2020, 11:2301. PMID: 32385265. (IF 2020: 12,1)

56. Amare AT, Schubert KO, Hou L, …, **MDD Working Group of the Psychiatric Genomics Consortium**, *et al.* Association of Polygenic Score for Major Depression with Response to Lithium in Patients with Bipolar Disorder. Mol Psychiatry 2020, epub ahead of print. PMID:32203155 (IF 2020: 12,4)

57. Coleman JRI, Peyrot WJ, Purves KL, …, **MDD Working Group of the Psychiatric Genomics Consortium**, *et al*. Genome-wide gene-environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK Biobank. Mol Psychiatry 2020, 25(7):1430-1446. PMID:31969693. (IF 2020: 12,4)

58. Glanville KP, Coleman JRI, Hanscombe KB, …, **MDD Working Group of the Psychiatric Genomics Consortium**, *et al*. Classical Human Leukocyte Antigen Alleles and C4 Haplotypes Are Not Significantly Associated With Depression. Biol Psychiatry 2020, 87, 419-430. PMID:31570195. (IF 2020: 12,1)

59. Coleman JRI, Gaspar HA, Bryois J, …, **MDD Working Group of the Psychiatric Genomics Consortium**, *et al*. The Genetics of the Mood Disorder Spectrum: Genome-wide Association Analyses of More Than 185,000 Cases and 439,000 Controls. Biol Psychiatry 2020, 88(2):169-184. PMID: 31926635 (IF 2020: 12,1)

60. Andlauer TFM, Guzman-Parra J, Streit F, …, **MDD Working Group of the Psychiatric Genomics Consortium**, *et al*. Bipolar multiplex families have an increased burden of common risk variants for psychiatric disorders. Mol Psychiatry 2021, 26(4):1286-1298. PMID:31712721. (IF 2020: 12,4)

61. Czamara D, Eraslan G, Page CM, …, **MDD Working Group of the Psychiatric Genomics Consortium**, *et al*. Integrated analysis of environmental and genetic influences on cord blood DNA methylation in new-borns. Nat Commun 2019, 10, 2548. PMID: 31186427. (IF 2018: 11,9)

62. Grove J, Ripke S, Als TD, …, **MDD Working Group of the Psychiatric Genomics Consortium**, *et al*. Identificaiton of common genetic risk variants for autism spectrum disorder. Nat Genet 2019, 51, 431-444. PMID: 30804558 (IF 2018: 25,5)

63. De Jong S, Diniz MJA, Saloma A, …, **MDD Working Group of the Psychiatric Genomics Consortium**, *et al*. Applying Polygenic Risk Scoring for Psychiatric Disorders to a Large Family With Bipolar Disorder and Major Depressive Disorder. Commun Biol 2018,1,163. PMID: 30320231 (IF 2019: 4,0)

64. Barbu MC, Zeng Y, Shen X, …, **MDD Working Group of the Psychiatric Genomics Consortium**, *et al*. Association of Whole-Genome and NETRIN1 Signaling Pathway-Derived Polygenic Risk Scores for Major Depressive Disorder and White Matter Microstructure in the UK Biobank. Biol Psychiatry Cogn Neurosci Neuroimaging 2019, 4, 91-100. PMID: 30197049 (IF 2018: 4,6)

65. Peyrot WJ, Van der Auwera S, Milaneschi Y, …, **MDD Working Group of the Psychiatric Genomics Consortium,** *et al*. Does Childhood Trauma Moderate Polygenic Risk for Depression? A Meta-analysis of 5765 Subjects From the Psychiatric Genomics Consortium. Biol Psychiatry 2018, 84, 138-147. PMID: 29129318 (IF 2018: 11,5)

66. Milaneschi Y, Lamers F, Peyrot WJ, …, **MDD Working Group of the Psychiatric Genomics Consortium**, *et al*. Genetic Association of Major Depression, et al. With Atypical Features and Obesity-Related Immunometabolic Dysregulations. JAMA Psychiatry 2017, 74, 1214-1225. PMID: 29049554 (IF 2017: 15,3)

67. **Cross-Disorder Group of the Psychiatric Genomics Consortium.** Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell 2019, 179, 1469-1482. PMID:31835028 (IF 2020: 38,6)

68. Krämer B, Knoll R, Bonaguro L, …, **Deutsche COVID-19 OMICS Initiative (DeCOI)**, et al. Early IFN-alpha signatures and persistent dysfunction are distinguishing features of NK cells in severe COVID-19. Immunity 2021, online ahead of print. PMID: 34592166 (IF 2020: 22,6)

69. Singh Y, Trautwein C, Fendel R, …, **Deutsche COVID-19 OMICS Initiative (DeCOI).** SARS-CoV-2 infection paralyzes cytotoxic and metabolic functions in the immune cells. Heliyon 2021, e07147. PMID: 34075347 (IF 2020: 1,9)

70. Warnat-Herresthal S, Schultze H, Shastry KL, …, **Deutsche COVID-19 Omics Initiative (DeCOI)**, *et al*. Swarm Learning for decentralized and confidential clinical machine learning. Nature 2021, in press. PMID:34040261 (IF 2020: 42,8)

71. Aschenbrenner AC, Mouktaroudi M, Krämer B, …, **German COVID-19 Omics Initiative (DeCOI)**. Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients. Genome Med 2021, 13, 7. PMID:33441124 (IF 2020: 10,7)

72. Bernardes JP, Mishra N, Tran F, …, **Deutsche COVID-19 Omics Initiative (DeCOI)**. Longitudinal Multi-omics Analyses Identify Responses of Megakaryocytes, Erythroid Cells, and Plasmablasts as Hallmarks of Severe COVID-19. Immunity 2020, 53, 1296-1314. PMID:33296687 (IF 2020: 22,6)

73. Schulte-Schrepping J, Reusch N, Paclik D, …, **Deutsche COVID-19 Omics Initiative (DeCOI)**. Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. Cell 2020, 182, 1419-1440. PMID:32810438 (IF 2020: 38,6)

8 ACKNOWLEDGEMENTS

Mein Dank gilt an allererster Stelle all meinen Mentorinnen und Mentoren, die meine wissenschaftliche Laufbahn über die Jahre durch all ihre Höhen und Tiefen wahrhaft unermütlich unterstützt und mir mit Rat und Tat und einem stets offenen Ohr zur Seite gestanden haben. Besonders bedanken möchte ich mich bei ihnen für all ihre Förderung und dafür, dass ich auch immer viele eigene Wünsche bezüglich meiner Projekte habe verwirklichen können. Danke, lieber Thomas Schulze, lieber Peter Falkai, liebe Ulla Protzer, liebe Juliane Winkelmann, lieber Thomas Meitinger, lieber Nico Katsanis, liebe Jie Shen und lieber Michael Hüll ! Und es würde mich sehr freuen, wenn mir diese Unterstützung auch in Zukunft noch lange erhalten bliebe und ich mich auch weiterhin von ihren motivierten, motivierenden, intelligenten, ehrlichen und kritischen Herangehensweisen inspirieren lassen könnte. Zudem danke ich Frau Prof. Ortrud Steinlein dafür, dass sie sich als Teil meines Fachmentorats zur Verfügung gestellt hat.

Mein ganz großer Dank gebührt auch allen Kollaborationspartnern in München, in Weihenstephan, in Freiburg, in Bonn, in Kassel, in London, in Wien, in Innsbruck, in Barcelona, in Prag, in Budapest, in Durham, in Palo Alto und weltweit im Psychiatric Genomics Consortium sowie den zig tausenden Studienteilnehmerinnen und -teilnehmern. Ohne sie wäre keines der hier dargestellten Projekte möglich gewesen.

Leider reicht der Platz nicht aus, um all diejenigen namentlich zu nennen, die meinen Arbeitsalltag—in München am Helmholtz Zentrum und an der LMU sowie im Katsanis Lab an Duke und im Winkelmann Lab an Stanford—geprägt und durch viele interessante Diskussionen bereichert, durch Hilfestellungen und Erklärungen erleichtert und durch ihre Unterhaltung sehr abwechslungsreich gestaltet haben. Vielen lieben Dank für die schönen Stunden und all Eure Hilfe ! Und, ja, am Ende hat mir das alles auch sehr viel Spaß gemacht.

Der allergrößte Dank an dieser Stelle gebührt natürlich meiner Familie, ohne deren liebevolle Unterstützung während der letzten 40 Jahre all dies nicht möglich gewesen wäre.