Aus der Klinik Translationale Psychiatrie
Klinikum der Ludwig-Maximilians-Universität München

Dissertation

zum Erwerb des Doctor of Philosophy (Ph.D.)

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität München

Identification of in vitro model systems capable of capturing the polygenic basis of mental illness

vorgelegt von:

Laura Teresa Jiménez Barrón

aus:

Mexico City

Jahr:

2024

Mit Genehmigung der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

Erstes Gutachten von:	Prof. Dr. Nikolaos Koutsouleris
Zweites Gutachten von:	Prof. Dr. Michael Ziller
Drittes Gutachten von:	Prof. Dr. Angela Abicht
Viertes Gutachtes:	Priv. Doz. Dr. Florence Bareyre

Dekan: **Prof. Dr. med. Thomas Gudermann**

Datum der Verteidigung:

20.02.2024

Contents

1 ABSTRACT

Large genome wide association studies have identified thousands of genetic variants associated with psychiatric diseases. These variants are likely to act in a highly condition and cell type specific fashion. However, at this point their cellular context and developmental of action remains poorly understood. Moreover, it remains unclear to what extent in vivo and in vitro model systems can approximate the human cellular conditions where polygenic psychiatric disease risk is operational.

One of the application scenarios of genome wide association studies results, is the calculation of polygenic risk scores to identify subjects at risk of developing disease or to stratify a cohort with a given trait or disease, in this thesis I performed genomic imputation and traditional polygenic risk score calculation of two case/control cohorts for schizophrenia at different P-value thresholds used to subset the GWAS derived SNP associations considered in the scoring. The power of polygenic risk scores in the clinical setting for psychiatric disorders is limited, the fact that the hundreds of loci that contribute to disease liability are likely to act in a cell type specific manner, supports the need to develop cell type specific polygenic risk scores, for this purpose it is crucial to understand how the disease risk affects specific cell types.

One of the objectives of this thesis was to identify which cell types are vulnerable to psychiatric disease associated polygenic risk, I did this by using stratified LD score regression for partitioning heritability from GWAS summary statistics while accounting for linkage disequilibrium, this allowed me to identify which cell type groups are enriched for psychiatric disorders heritability. I performed this partitioned heritability analysis in the transcriptomic and chromatin accessibility profiles of 10 neuronal and non neuronal cerebral cell types derived from human post-mortem brain tissue of the prefrontal cortex.

LD score regression was performed using the identified active elements and GWAS derived summary statistics for various psychiatric disorders as well as control traits. Cell types including excitatory neurons of the cortical layer 2-3, corticothalamic neurons and inhibitory neurons as well as microglial cells and oligodendrocyte progenitor cells were significantly enriched for schizophrenia and bipolar disorder, many more significant positive associations were found between some of these cell types and

traits corresponding to psychiatric disorders and Central Nervous System traits. All of this in line with previous research where these cell types had been found to be relevant to susceptibility to variants associated to the above mentioned traits. I performed the same analyses on ATACseq and RNAseq having ATACseq yielding finer and more specific results than RNAseq, which failed to identify some relevant cell types.

ATACseq was used in assessing the validity of different model systems to see if they could capture the polygenic architecture of psychiatric diseases, this model systems included iPSC derived neurons, cerebral organoids, post-natal mouse cortical cells and fetal cortical neurons. The results were consistent with the postmortem tissue findings.

These results allowed us to benchmark the used *in vitro* models that maintain the heritability enrichment of postmortem tissues, opening the possibility for a robust use of scATACseq data derived from these models to reliably identify cell type specific elements relevant for disease and leverage this information in building better predictive tools like cell-type specific polygenic risk scores.

2 LIST OF FIGURES

List of Figures

3 LIST OF ABBREVIATIONS

ATACseq	assay for transposase-accessible chromatin with sequencing
BD	Bipolar Disorder
CAD	Coronary Artery Disease
CHR	Chromosome
CNS	Central Nervous System
CNVs	Copy Number Variants
CThPNs	Corticothalamic Projection Neurons
DHSs	DNase I hypersensitive sites
DLPFC	Dorsolateral Prefrontal Cortex
DSM	Diagnostic and Statistical Manual of Mental Disorders
Disc1	Disrupted in schizophrenia-1
ENCODE	EN- Cyclopedia Of DNA Elements
GWAS	Genome Wide Association Study
LD	Linkage Disequilibrium
LDL	Low-Density Lipoprotein
LDSC	Stratified LD-Score regression
MDS	Multidimensional Scaling
MHC	Major Histocompatibility Complex
NPCs	Neural Progenitor Cells
OPC	Oligodendrocyte Progenitor Cells
PFC	Pre-Frontal Cortex
PN	Projection Neurons
PRS	Polygenic Risk Score
PVALB	Parvalbumin
SCZ	Schizophrenia
SNP	Single Nucleotide Polymorphism
SST	Somatostatin
ScPNs	Subcerebral Projection Neurons
VIP	Vasoactive Intestinal Polypeptide
eQTL	expression Quantitative Trait Loci
hKCl	high potassium chloride
iPSCs	induced Pluripotent Stem Cells
scRNAseq	single-cell Ribonucleic acid sequencing
snRNAseq	single-nucleus Ribonucleic acid sequencing
scATACseq	single cell Deoxiribonucleic Acid sequencing

4 INTRODUCTION

4.1 Psychiatric disorders and their causes

Psychiatric disorders or mental illnesses are described as health conditions that include behavioral, cognitive, and/or emotional changes that negatively impact an individual's life by interfering with their daily social or work activities, these changes must not be an expected or isolated response to an isolated traumatic event and rather a behavioral pattern [61]. They affect more than 25% of the world's population and therefore, are one of the leading causes of disability [19].

Examples of psychiatric disorders are schizophrenia (SCZ), bipolar disorder (BP), Major depressive disorder, anxiety, anorexia nervosa, autism spectrum disorder, attention-deficit/hyperactivity disorder, Tourette syndrome, and obsessive-compulsive disorder[19]. Given that psychiatric disorders are often observed to run in families, many familial and twin genetic studies have been carried out, resulting in high amounts of heritability, which is the amount of variation from parents to offspring explained by genetic factors [66].

Historically, these disorders have been studied and diagnosed based on experts' clinical observations resulting in tools like the diagnostic and statistical manual of mental disorders (DSM), a manual containing a series of criteria and checklists of symptoms, signs of illness, and their duration, this manual aims to objectively assess whether an individual meets the minimal criteria to be diagnosed with a mental disorder and therefore is widely used by clinicians to help them carry out a diagnosis. However, the scope of the DSM to stratify a group of people with mental illness or to discern one disease from another has been challenged by recent genomic studies [57] [63].

Genomic studies have shown that psychiatric disorders are not exempt from pleiotropy, a phenomenon where a genetic variant has an effect on two different traits or multiple variants have an effect on two different traits, this challenges the idea that psychiatric disorders have well-defined boundaries and are mutually exclusive. Cross-disorder studies have yielded evidence for pleiotropic effects of copy number variants (CNVs) where they are directly associated with multiple disorders

like schizophrenia, intellectual disability, and autism just to mention a few. Another example is a schizophrenia and bipolar disorder cross-study where it was shown how a risk score derived from a Schizophrenia study is directly associated with the risk of having bipolar disorder, suggesting a shared heritability between both of them [31][57].

Genomic studies have also shown that psychiatric disorders are highly polygenic and arise from a combination of many common variants that have small effects combined with rare variants with much bigger effects, making it challenging to select individual variants, genes, or pathways for downstream functional analyses. [66][56]

In spite of the before-mentioned high heritability of psychiatric disorders, there are many non-genetic factors that contribute to the risk of the development of disease [67]. These factors include but are not limited to the urban environment, childhood stress, and the use of drugs. Considering both, genetic and environmental factors and how they interact is crucial for a better understanding of mental illnesses [67] [42].

4.1.1 Schizophrenia and Bipolar Disorder, clinical aspects and their genomic and Environmental Liability

Clinical and molecular evidence suggests that psychiatric disorders that were once considered independent conditions, actually share associated genomic loci and biological pathways blurring their boundaries, this phenomenon is known as pleiotropy and its effects in variants associated to various psychiatric diseases make it difficult to draw a clear line between two diseases [33].

A recent analysis of pleiotropy identified 109 genomic loci that affected two or more psychiatric disorders, it also found 23 different loci that affected four or more disorders. These findings allowed them to. Identify groups of psychiatric diseases that shared a genetic background. The first group is composed by anorexia nervosa, obsessive compulsive disorder and tourette syndrome, which unsurprisingly also share clinical aspects like compulsive behaviors. The second group formed by major depressive disorder, bipolar disorders and schizophrenia which are all considered to be mood and psychotic disorders. The last group they identified that shared genomic liability belonged autism spectrum disorders and attention-deficit/hyperactivity disorder [19].Therefore we can see depression, bipolar disorder, and schizophrenia are

often seen as a continuum [33]. In this section, the differences and commonalities between bipolar disorder and schizophrenia relevant to their nosology and causes will be described.

Schizophrenia and bipolar disorder are two psychiatric disorders that according to the diagnostic and statistical manual of mental disorders (DSM) have an average lifetime prevalence of under 1% and 2% respectively [26][33] However, prevalence has a high amount of variation among populations across the world, by way of illustration, the Finnish population has been reported to have a schizophrenia prevalence of 0.87%, supporting the idea that population and environmental factors have a big impact on someone's risk to develop a mental illness [26]. Additionally, heterogeneity in diagnosis and sub-classification of these diseases contribute to varying prevalence numbers. For instance, when considering the most studied subtypes of bipolar disorder, the global lifetime prevalence goes down to 1% [33].

Bipolar Disorder (BD) can in itself be looked at as a spectrum of conditions that range from BD type 1 which must include at least one maniac episode, BD type 2 with at least one depressive and one hypomania episode, and cyclothymic disorder where both depressive and hypomania milder symptoms last for a period of at least two years [33]. Schizophrenia is characterized by different symptoms that are divided into three different categories; cognitive dysfunctions, positive symptoms like hallucinations and delusions that lead to the loss of contact with reality, and negative symptoms like anhedonia, reduced energy, and social withdrawal just to mention some. People diagnosed with these diseases can either fully recover or need treatment for life and their life quality and expectancy significantly go down compared to the rest of the population, mainly due to suicide and cardiovascular disease. [26]

There are a series of aspects of an individual's life that confer a greater risk of developing schizophrenia and bipolar disorder compared to the general population, for instance, complications during pregnancy and birth as well as a higher paternal age, all correlate with a higher risk of developing schizophrenia, which is interestingly more frequent and severe in men than women who also tend to develop the disease at a later age. Other environmental factors that have been shown to have an important impact when calculating someone's risk are lower social cohesion, growing up in large urban areas, being a migrant, or belonging to a migrant family, drug abuse, and social adversity [26]. Childhood maltreatment is a factor that is particularly

common in people who develop bipolar disorder [35]. The exact mechanisms through which these factors act are still not fully understood [26].

Both diseases can be observed to cluster in families and different studies across different populations have shown high degrees of heritability. The schizophrenia global heritability is estimated at 80%, and monozygotic twin concordance at 45% [67]. A Danish nationwide study of their twin registry reported a schizophrenia heritability of 79% whereas a Finnish population twin study estimated schizophrenia heritability to be 83% [21][9]. Bipolar disorder has a similarly high heritability of about 70% [35]. These observations have led researchers to carry out a series of genetic studies, which have contributed a vast amount of evidence pointing to variants involved in molecular mechanisms associated with these diseases [26]. Some rare and many common variants have been found to contribute to the increased liability of developing both disorders [33] [26] [40].

Some schizophrenia-associated variants are harbored in the major histocompatibility complex which contains genes involved in the regulation of the immune response, others are in genes like DRD2 which encodes a dopamine receptor, and some others are in the glutamate receptors mGluR3, GluN2A, GluA1 and their respective receptor components GRM3, GRIN2A, and GRIA1 [26]. Some of the schizophrenia-associated genes have a higher expression during fetal development, this together with the risk factor of pregnancy complications, suggests that schizophrenia might already start developing much earlier than thought. Besides these genes, there are many more genomic loci that contain risk alleles, so far a total of 270 loci have been identified for schizophrenia of which 130 contain candidate causal genes for European populations [65]. For bipolar disorder, 30 loci coding for ion channels, neurotransmitter transporters, and synaptic components have been identified [60]. Individually, these loci with common variants confer only small risks and even cumulatively they contribute only a small fraction of the total liability [26]. 8 different rare copy number variants have been identified to have a strong association to schizophrenia, including the 22q11.2 deletion which confers a 20-fold risk of developing the disease [32]. In summary common variants contribute only a small fraction of the total risk while rarer variants contribute much more [26] [32].

Many of the common risk loci belong to non-coding sequences pointing to the contribution of epigenetic mechanisms, which are those that contribute to changes in gene expression in space and time without having to change the genetic code itself

[59]. There is evidence that states that bipolar disease-associated variants can be found in enhancers that are active in cortical projection neurons and corticothalamic projection neurons, [33]. Schizophrenia as well as bipolar disorder have enhancers active in inhibitory GABAergic parvalbumin-expressing interneurons. Furthermore, postmortem and neuroimaging studies have also been widely used to better understand schizophrenia, finding brain regions with abnormal volumes and activation patterns, some of these results have been helpful for connecting genetic risk to brain function, this is done by looking at neuroimaging differences among subjects with different genetic risk loci for schizophrenia, including loci in genes like COMT, NRG1, and DISC1. Gene set studies point risk loci presence to synaptic disruption by associations with synaptic plasticity, targets of FMRP, voltage gated calcium ion channel complexes [32].

Bipolar disorder age of onset varies from childhood to late adulthood, having its peak at around 25 years, there are currently no reliable biomarkers that allow early detection [35]. Formerly thought of as an early-adulthood-onset disease, schizophrenia is usually only diagnosed after the first psychotic episode, however, there is clinical and genomic-derived evidence pointing to a much earlier onset in adolescence, making it important to develop tools and biomarkers to be able to detect individuals at a higher risk of developing schizophrenia for early intervention and possibly prevent the disease from developing [26].

4.2 Heritability

Understanding how genes contribute to complex traits has long been a subject of interest for biologists, pioneering work by Ronald Fisher in 1918 established an explanation for the observations made of correlations of trait measurements among relatives. His aim was to attribute these correlations to the influence of numerous genetic factors. Fisher introduced the concept of variance and proposed its decomposition into genetic and environmental components, which laid the foundation for the concept of heritability[51] Fisher suggested a decomposition of the phenotypic variance, denoted as P, into the combined genotypic variance G and the environmental variance E: var(P) = var (G) + var (E) Although Fisher did not explicitly employ the term "heritability" in his 1918 paper, he emphasized the significance of the ratio var(G)/var(P), which aligns precisely with what we now refer to as the broad-sense heritability H2. Therefore,

H^2 = var(G) / var (P)

Where H^2 measures the amount of phenotypic variance attributable to genotypic variance[51].

The genotypic variance G can be additionally broken down into its constituent components: additive A, dominance D, and epistasis Ep. The ratio of the additive genetic variance, which accounts for the combined effects of the two alleles at each genetic locus, to the phenotypic variance is referred to as "narrow-sense heritability" [51].

h^2 = var (A) / var (P)

Over the past years, many methods have been developed to estimate the heritability of SNPs to measure the proportion of phenotypic variance explained by a set of SNPs interrogated in an experiment. The estimation of SNP heritability provides insights into the extent to which genetic variants contribute to phenotypes, thereby enhancing our comprehension of the genetic makeup of complex traits. In this article, we examine the latest approaches and widely employed methods for estimating SNP heritability in relation to continuous and binary phenotypes. Our focus is on exploring the underlying model assumptions and parameter optimization strategies utilized in these methods.[51]

4.3 Genome-Wide Association Studies

Genome-Wide Association Studies (GWAS), serve the purpose of identifying a genomic locus or loci with a specific genotype associated with a specific phenotype or trait. A group of people (cohort) is carefully selected and drawn from a population, these are classified in cases and controls when carrying out disease-related studies that can be modeled with a binary status or divided into more categories when the trait is quantitative. The cohort's phenotypic information is collected and genotypes for each individual are obtained either by sequencing their genome or by using a combination of genotyping arrays plus genomic imputation. This results in a collection of Copy Number Variants (CNVs), small insertions or deletions (Indels), or single-nucleotide polymorphisms (SNPs). The latter is the most commonly tested type of variant in GWAS. Then, the allele frequencies of both cohorts are compared by performing an association test on every single genomic variant to see whether a

specific allele is significantly more common to be found in the cases group versus the control group [66].

Linkage disequilibrium is the non-independent association of two alleles in a population, this can happen because of the physical arrangement of chromosomes and the mechanism through which they recombine, which is not entirely random making two physically close SNPs more likely to be inherited together and therefore be in linkage disequilibrium, this is a well-known phenomenon that affects GWAS associations and that can modify the strength of an association or make a variant seem to be associated with the disease or trait being analyzed when its association is a result of physical distance to the true associated SNP [66]

4.3.1 Biological interpretation and function of GWAS associations

Cell identity is driven by the activation of selective regulatory DNA that governs the gene expression patterns of each cell type. Regulation of DNA is mediated by sequence-specific binding of regulatory elements like transcription factors, this results in chromatin-remodeled states that make specific DNA regions accessible and sensitive to nucleases, these accessible regions are known as DNase I hypersensitive sites (DHSs). Taking advantage of this feature, DNase I hypersensitivity assays have been widely used to profile the genome-wide chromatin landscape of different cell lineages at different life stages, the resulting maps of regulatory DNA have helped to show connections between chromatin accessibility and transcription. [64]. The ENCyclopedia Of DNA Elements (ENCODE) project is a consortium that has mapped, annotated, and cataloged cis-regulatory functional elements encoded in the human and mouse genomes, [27] Genome-wide association studies test associations for gene-contained SNPs and SNPs contained in non-coding genomic regions, where many significant associations have also been found. Thanks to projects like ENCODE, it has been possible to annotate these disease-associated non-coding variants and see that they cluster in genomic regulatory DNA elements.

Interestingly 88% of DNase I hypersensitive sites in many GWAS correspond to fetal development active regions suggesting a role of disease-associated non-coding variants in development. Not only are these DHSs active in a time-specific manner but are also found in specific tissues where they can modulate insulators, silencers, locus control regions enhancers, and promoters. [64] [1] To confirm whether these SNPs are not simply tagging neighboring SNPs in coding regions, linkage disequilib-

rium analyses are performed and in some studies, it has been observed that about three quarters of the investigated SNPs were in a DHS or were in perfect linkage disequilibrium with a DHS.[64] The chromatin accessibility landscape has a major role during development [1] and many disorders have a link to gestational exposure to environmental factors [64], this can directly link GWAS disease-associated non-coding variants to specific cell types and developmental or life stages. For instance, DHSs maps corresponding to brain cell types in fetal stages overlap with variants associated to neurological disorders, indicating that mapping associated variants to functional DNA can point to disease-relevant cell lineages [64] GWAS disease-associated non-coding variants can alter transcription factors recognition motifs and modify the local chromatin structure and gene expression, these variants are considered as expression quantitative trait loci (eQTL) and can explain part of the variation in the expression of a gene which results in an imbalanced expression of the two different alleles for a heterozygous individual [64]. Not only do disease-associated variants cluster in DHSs but they also converge in transcriptional regulatory pathways, this can be observed when analyzing the interaction network of transcription factors that bear disease-associated variants from a disease or a group of related diseases, also contributing evidence for the shared liability of related diseases like psychiatric ones [64].

Since GWAS have a very stringent statistical P−value correction for multiple testing, variants with a small effect, might not reach significance and therefore not be considered in post-GWAS analyses, more systematic and conditional association methods on regulatory DNA, can point to specific associations for these variants [64]. Furthermore, there are many DHSs harboring GWAS SNPs that have been connected to distal regions shedding light on how the 3D chromatin structure also plays a role in regulating gene expression and explaining some SNP-disease associations [64].

4.3.2 GWAS scope, biases and limitations

GWAS can be used for several different purposes, like identifying genetic risk loci for a disease or being the source of the necessary information to stratify a population and calculate someone's liability for a disease [66].

Despite being a method that shed light on many genomic loci relevant to a vast amount of traits, GWAS have several limitations that must be considered when inter-

preting their results. GWAS often only identify loci with small effect sizes, meaning that they can only explain a limited proportion of the studied trait heritability [62].

Besides the above-mentioned limitations, many GWAS are subject to several biases that should also be considered when interpreting and further using their results. Performing a GWAS involves the recruitment of a very large sample, which is a very costly and time-consuming task, for this reason, there are many consortia and public efforts that have launched projects that aim to recruit a big number of people who are willing to contribute their genotypes as well as their phenotypic data, examples of these are the UK Biobank or the Mexico City Prospective Study [3].

The strategies that these consortia use to collect the data, can bias the entire cohort in different ways like containing a disproportionate amount of people who are healthier or with a specific life quality, all of which may not be representative of the population one aims to study, drawing conclusions from studies must always be done with caution and always considering these limitations [66][62].

4.4 Polygenic Risk Scores

Polygenic risk scores have been proven to be useful to study many different traits or diseases and are particularly useful for psychiatric disorders where it has been shown that many common variants contribute with a small effect, therefore it becomes important to consider all genomic variation at once rather than performing studies on a single variant or gene, allowing to capture a lager amount of the heritable variance [57]

A polygenic risk score (PRS) is a number that is calculated for a given individual using the resulting summary statistics of a given GWAS, it is calculated by summing the effect size of all the risk alleles corresponding to a trait [10]. PRS have a handful of applications, among them is assessing the shared heritability between related traits or diseases, polygenic risk scores have already proven useful for demonstrating the shared genetic risk between psychiatric disorders like schizophrenia and bipolar disorder, in 2009 the International Schizophrenia Consortium [43] showed via PRS that the genetic risk for schizophrenia is a predictor for bipolar disorder. polygenic risk scores can also be used to identify target individuals who lie in the extreme ends of the score range to perform further studies on them or their tissues, helping to shed

light on the cellular and molecular mechanisms of disease. In the clinical setting, the development and improvement of polygenic risk scores are of great interest as a tool to identify individuals at risk of developing a disease or to stratify a cohort of diagnosed individuals. Even tho there are other methods that can accurately estimate and study heritability, polygenic risk scores are the only ones that allow researchers to predict the risk of disease for any given individual [66][10].

There are many strategies to calculate a polygenic risk score, the simplest and most widely used one is a genome-wide summation of the effect sizes corresponding to each allele for a given sample, these effect sizes are part of the result of a genome-wide association study which must be performed on a base cohort, which is different from the target cohort on which the PRS will be calculated, both cohorts must belong to the same population but researches must make sure there is no sample overlap between the two cohorts, otherwise, this can result in an inflated association between the trait and the calculated polygenic risk score. Another aspect that can result in an inflated association is the presence of related individuals between the base and target datasets so highly related individuals should be excluded [10]. The power and accuracy of a polygenic risk score are heavily dependent on the quality control performed on both the base and target cohorts, calculating chip-heritability from the GWAS summary statistics and making sure it is higher than 0.005 is a critical QC step to assert the predictive power of the polygenic risk score [10][47].

Not every single tested variant in the GWAS is always considered in the calculation of a PRS; a subset of variants is often created to consider only the ones that are informative and thus contribute to an individual's higher or lower risk of developing a disease [10]. Furthermore, given that the GWAS SNP effects are calculated with uncertainty, the effect size must be adjusted to avoid high standard errors, this can be achieved by shrinkage of effect size estimates or by using P value thresholding as the criteria to include a certain SNP or not, the second method ultimately shrinks the effect size from all excluded SNPs to zero but leaves the included ones unchanged. To address the effect of linkage disequilibrium on polygenic risk scores, SNPs can be clumped with the goal of keeping SNPs that are independent of each other and whose effects can be summed without inflating the result [10][15].

After calculating polygenic risk scores for a given trait or disease, a regression is performed on the target sample using the PRS as a predictor for the trait or disease. When using P value thresholding, a regression for the PRS corresponding

to the different sets of SNPs can be performed in order to choose the threshold that has the best fit that explains the most phenotypic variance R^2. Something to keep in mind is that while choosing the predictive threshold that produces the polygenic risk scores with the best fit, this could be a result of overfitting the target data which in turn produces inflated results, the best way to avoid overfitting is to perform out-of-sample prediction where the parameters are first optimized by using a training sample and then tested in a validation dataset [10][15].

4.5 Functional genomics

Between individual lab efforts and big consortia projects, the amount of omics datasets that have been generated in the past decade has been massive. Identifying genomic associations to psychiatric diseases is very informative but also limited when it comes to providing mechanistic insights for these diseases. For this purpose, a number of omic technologies have been developed to study multiple omes including the methylome, transcriptome, epigenome, proteome, metabolome, and microbiome, all of which can help to annotate disease-associated SNPs and genomic regions [14].

Some of the consortia efforts include the ENCODE project, initially launched to characterize the regulatory function of the human genome [1], the Roadmap Epigenomics project (now included under the ENCODE data portal) was an effort to collect RNA-seq, ChIP-seq, DNase-seq and methylation data from human blood and 22 tissue types. Along these lines, the PsychENCODE project was launched with the specific objective of investigating genomics and epigenomics data related to the human brain in order to study neuropsychiatric disorders. This project has the largest collection of brains with psychiatric disorders and has characterized them using ChIP-seq, ATAC-seq, Ribo-seq, proteomics, DNA methylation, Hi-C data are available for some tissues and genotyping all samples. The Allen Brain Atlas initially collected microarray expression and MRI measurements from approximately 900 neuroanatomical brain slices from two subjects, their results showcased the correlation between gene expression and spatial localization, now they have expanded to more modern technologies like single-cell RNA-sequencing performed in different brain tissues. The CommonMind Consortium developed a publicly accessible repository of functional genomic data derived from the dorsolateral prefrontal cortex (DLPFC) encompassing Brodmann areas 9 and 46. This comprehensive dataset comprises information from approximately 1000 individuals sourced from four dis-

tinct brain banks, including healthy individuals, schizophrenia, and bipolar disorder patients. The genomic data encompasses RNA-seq and SNP genotypes as well as ATAC-seq data [14]. This thesis focuses on the use of two of these omic technologies, RNA sequencing and ATAC sequencing to study schizophrenia in post-mortem brain tissue as well as in iPSCs.

Whole genome and exome sequencing have shed light on the polygenic nature of psychiatric disorders, showing that their genetic liability is distributed across the genome, and that cell identity is defined by its transcriptome, it is reasonable to look at the gene expression profiles and active regulatory elements in different neuronal cell types to find out which ones are relevant for the study of these disorders and elucidate the functional consequences of disease-associated variants [69] .

To profile the set of transcripts of a cell type, various technologies have been developed including microarrays and Sanger sequencing, later when high throughput sequencing technologies were created, a technique called RNA-seq was developed, which involves capturing the bulk of mRNA molecules of a cell population and using a reverse transcriptase enzyme to generate a library of complementary DNA (cDNA) that corresponds to the captured mRNA, this library can then be sequenced using next generation sequencing technologies and mapped to the reference genome, this makes it possible to know which genes and in which quantity are being transcribed in a population of cells [37].

RNA-seq has been widely used to profile various organ tissues in the human body, however, the human brain has an extra layer of complexity where many different cell types coexist in a small space, making bulk RNA sequencing a suboptimal technique to profile brain cell types with enough resolution. A more recent technique where a single cell can be sequenced opens the possibility of accurately profiling the vast diversity of brain cell types. Neuronal cell types are typically identified by looking at cell morphology, their localization within the brain, their electrophysiological activity connectivity to other cells, and the expression of specific marker genes, however a comprehensive atlas of brain cell types that includes their transcriptional signatures in different conditions like developmental stages or diseases, has only started to develop, this thanks to the advancement of single-cell sequencing technologies like single-cell RNA sequencing (scRNA-seq) where different platforms for capturing individual cells or strategies for single-cell demultiplexing have been designed [36].

Single-nucleus isolation and sequencing methods have also been developed, examples of them are single-nucleus RNA sequencing (sNuc-seq) which has demonstrated to agree with the results yielded by scRNAseq and having the advantage of not needing intact cells and being able to keep different cell types regardless of their size differences, which is common in brain cell tissue samples and a known problem in scRNAseq cell sorting steps. Yet another advantage of sNuc-seq over scRNAseq is that human tissue banks usually use chemical fixation and freeze their samples, making it hard to preserve cells as needed for scRNAseq techniques [36].

Once the RNA sequencing has been performed, a series of computational analyses have to be performed, a typical RNAseq bioinformatics workflow consists of preprocessing the raw sequencing data which includes cleaning it, demultiplexing it, mapping reads to the reference genome, and counting how many reads map to a specific gene, the resulting gene expression matrix is then normalized to remove batch effects. Once these steps are done, the data goes through a dimensionality reduction step to capture the main signal and be able to assign each different cell to a cluster, a differential gene expression analysis is then performed between the different clusters reflecting the different cell lineages or different cell states [36].

Overall, this high-throughput technology has helped to reveal the spatiotemporal diversity of brain cell types, to resolve heterogeneity in brain tumors, to track the dynamic landscape of transcription in development as well as in aging, and identify cell types associated with diseases. As this technology advances making it possible to sequence a higher number of cells in a single experiment with more than half a million single cells, fast and memory-efficient bioinformatic tools like Seurat or Scanpy are needed [36].

The transcription profiles captured by the described RNA sequencing technologies above are controlled by cis-acting DNA regulatory elements like promoters and enhancers which are tightly associated with local changes in chromatin state and chromatin accessibility, profiling these changes can be very informative into the dynamics of a cell's transcription regulation. Some methods to perform chromatin profiling include the pioneering DNase-seq method by Boyle, et al [58] which takes advantage of the enhanced accessibility of open chromatin regions to DNase I [53].

Another method to profile chromatin accessibility is the assay for transposase-accessible chromatin using sequencing (ATAC-seq) [8] which uses the Tn5 transpose

enzyme that contains specific 19 base pairs long flanking sequences called end sequences, naturally these end sequences are recognized in the transposon by Tn5 which cuts the DNA and inserts it into a new position, a modified hyperactive Tn5 cuts accessible DNA and attaches adapter sequences that are later used to amplify, tag and sequence this DNA using high throughput sequencing technologies [53].

Just like RNA sequencing and other omic technologies, ATAC-seq can be performed on high throughput single cells experiments, which can be carried out by either utilizing a microfluidic platform to physically isolate individual cells or by sorting nuclei and applying combinatorial indexing, a method known as combinatorial cellular indexing of sorted nuclei (sciATAC-seq). Both techniques are adaptations of the original ATAC-seq protocol that employs a hyperactive Tn5 transposase to simultaneously cleave and label accessible chromatin [16].

Due to the physical limitations, there are to study the human brain of a live subject, studying brain tissue of post-mortem brain samples has been critical to understanding the biological mechanisms of many diseases [69]. Despite the limited availability of post-mortem brains, a number of studies have been performed and collected valuable information about psychiatric disorders.

To mention some, RNAseq of post-mortem brains of schizophrenia cases and controls has been performed on various brain regions and identified 144 differentially expressed genes in the hippocampus [24], their findings indicated the upregulation of immune/inflammation-related genes, including IFITM1, IFITM2, IFITM3, APOL1, ADORA2A, IGFBP4, and CD163, potentially contributing to the pathophysiology of schizophrenia. Sinclair et al. [54] performed RNA-Seq analysis on the prefrontal cortex samples from both schizophrenia and bipolar disorder subjects, revealing abnormal expression of FKBP5, PTGES3, BAG1, and glucocorticoid receptor genes. Another RNA-Seq study demonstrated altered expression of gene transcripts related to neuroplasticity (PROM1, ABCG2, FLI1) and circadian rhythms (OSBPL3, GANAB, SRSF5, RFX4) in the dorsolateral prefrontal cortex of individuals with BD. Furthermore, enrichment analysis of genome-wide association study (GWAS) data suggested the potential involvement of genes associated with GTPase binding in bipolar disorder. Kohen et al. [28] performed an RNA-Seq survey on dentate gyrus granule cells isolated from individuals with mental illnesses such as schizophrenia, BD, and major depression, revealing disrupted hippocampal miR-182 signaling compared to controls. Another recent RNA-Seq study identified three differen-

tially expressed genes, RXFP1, SSTR2, and CHRM2, belonging to class A of the G protein-coupled receptor family in the anterior cingulate of BD patients, suggesting dysregulation of G protein-coupled receptors in Bipolar disorder [28].

These studies show the value of studying post-mortem brain tissue and the potential of making use of omic technologies to further understand how specific brain regions, cell types, genes, and genomic regulatory elements are implicated in health and disease.

4.6 Linkage Disequilibrium Score Regression

As referenced in the previous sections, many of the genetic liabilities of polygenic traits or diseases like schizophrenia lie in markers with weak associations [64][66] that often don't meet the significance threshold and therefore are ignored in many post-GWAS functional studies [66]. Added to this phenomenon, some functional studies control for linkage disequilibrium by considering only one causal SNP per locus or don't consider it at all. Finucane et al. [17] developed a method that allows partitioning genome-wide SNP heritability for a specific functional category throughout the genome and using all SNPs while accounting for linkage disequilibrium, this method is referred to as stratified LD-score regression method or LDSC [17].

To investigate whether a given genomic functional category is enriched for the heritability of a specific trait or disease, this method uses GWAS summary statistics for the investigated trait or disease and linkage disequilibrium information from a panel that matches the population of the GWAS [17]. This method uses the fact that the x^2 association statistic for a specific SNP incorporates the effects of all other SNPs that are linked to it. Consequently, when examining a polygenic trait, SNPs with a higher LD score tend to exhibit, on average, higher x^2 statistics compared to SNPs with a lower LD score. This observation may stem from these SNPs having a greater likelihood of tagging a single SNP with a substantial effect or capturing multiple SNPs with weaker effects. By categorizing SNPs into different functional groups based on their contributions to heritability, how LD affects the x^2 statistic of a SNP can be assessed. Specifically, if LD is higher in a heritability-enriched category, the x^2 statistic of a SNP will increase more significantly compared to LD in a category that has minimal impact on heritability. In summary, this method identifies a category of SNPs as enriched for heritability when SNPs with high LD

to that category exhibit higher x^2 statistics than those with low LD to that category, to apply LDSC to a functional category, a baseline model with 24 publicly available non-cell-type-specific annotations is used, these annotations include coding, UTR, promoter and intronic regions, histone marks, DNase I hypersensitivity regions and enhancers, Z scores, P values, and false discovery rates are calculated, the P value tests if the annotation being evaluated contributes significantly to heritability after controlling for the categories in the baseline model [17].

4.6.1 Identifying Cell types relevant to schizophrenia heritability

Besides using stratified LDSC regression for testing SNP heritability enrichment on general functional categories like histone marks or evolutionarily conserved regions, it is also possible to use the stratified LD-score regression method to identify trait or disease-relevant tissues or cell types. For many diseases including psychiatric disorders, it is unclear which tissues or cell types are involved in their pathophysiology, however, information like which genes and which parts of the genome are open is now widely available, and leveraging this information together with GWAS results offers the opportunity to have a clearer picture of which tissues or cell types and at which time points or conditions are relevant for a disease. Some methods have already attempted to do this but fail to include all SNPs because they simply don't meet the required significance threshold, thus losing information. In contrast, the stratified LDSC regression method uses the association information for all SNPs in a GWAS while explicitly modeling linkage disequilibrium which controls for the inflated associations that the other methods are trying to avoid by simply removing SNPs [18][17].

By using the transcriptome as a functional category together with the GWAS summary statistics of the trait or disease of interest. Finucane et al. [18] tested for enrichment of heritability in different tissues according to their specific gene expression profiles, this was tested for 48 diseases and traits and showed tissue-specific significant enrichment for 34 of the tested traits. Among these, they found significant enrichment of inhibitory over excitatory neurons when tested for bipolar disorder heritability and excitatory over inhibitory neurons when tested for schizophrenia and body mass index [18]. This shows how gene expression data can yield very accurate enrichment results for polygenic traits like psychiatric disorders [18].

An independent study by Scene et al. also in 2018 [55] investigated the impli-

cation of different brain cell types in schizophrenia by mapping their unique transcriptional signatures obtained by scRNA-seq to SNPs associated with the disease. They use both, the stratified LD-score regression method and MAGMA, a method that evaluates whether gene level genetic associations with a trait increase linearly with cell type transcription profiles. Both methods control for gene size and linkage disequilibrium in different ways but they required both methods to give strong evidence of a cell-type to schizophrenia association to consider it true. Their results found a handful of cell types relevant to schizophrenia liability including hippocampal CA1 pyramidal cells, stratal MSNs, neocortical somatosensory pyramidal cells, and cortical interneurons, the associated genetic risk was bigger for mature cells in comparison to embryonic or progenitor cells. When dividing these cell types into subtypes, stratal MSNs expressing Drd1 or Drd2 as well as Pvalb-expressing interneurons were significantly associated, in hippocampal CA1 pyramidal cells, both major subgroups were significantly associated, in neocortical somatosensory cells, cortical layers 2/3,4,5 and 6 were also significant [55].

These findings provide compelling evidence that this polygenic approach represents a powerful method for utilizing gene expression data in order to interpret GWAS associations, making it useful for selecting the appropriate tissue or cell type for conducting in vitro experiments which is crucial in order to gain deeper insights into molecular mechanisms.[18].

4.7 Model systems for the study of psychiatric disorders

Studying psychiatric disorders in humans has the physical limitation of the inaccessibility of the brain [22], Therefore, there is a need to develop suitable model systems that accurately replicate the pathophysiological mechanisms of the disorder [12]. Throughout years of psychiatric research, different model systems like ex vivo brain tissue, human neuroblastoma cells, induced pluripotent stem cells (iPSCs), or animal models have been proven useful to reveal insights into the pathophysiology of different disorders, each of them having strengths and limitations but all useful to study specific aspects of a disease [12].

Animal models have long been used in many areas of disease research with the objective of obtaining biological insights into the studied disease. However, using animals for modeling psychiatric disorders has been very challenging, due to the symptom heterogeneity and overlap across different psychiatric disorders, this is

summed to the fact that there are no widely used reliable biomarkers, and therefore diagnostic tools are reduced to behavioral evaluations, however, it is hard to replicate disease aspects of human psychiatric disorders like hallucinations or anhedonia in animals. Even animal behaviors that are believed to correspond to human ones cannot be fully trusted as an equivalent since emotional and cognitive function cannot be assessed [39].

Nevertheless, some animal models for psychiatric disorders like schizophrenia have been developed, resulting from the introduction of known highly penetrant mutations associated with the disease, this is the case of schizophrenia mouse models for human 22q11.2 microdeletions which were created by deleting genes within the homologous region. Another example is the Disrupted in schizophrenia-1 (Disc1) mutant mouse that is believed to mimic some of the behavioral abnormalities of schizophrenia [39].

Given that, only a small fraction of schizophrenia-diagnosed cases can be linked to highly penetrant variants like Disc1 mutations or 22q11.2 microdeletions, and rather caused by a combination of the contribution of many common variants with a small effect size, these animal models have a very limited scope [39][57].

Induced pluripotent stem cells derived from patients with schizophrenia have shown promising evidence to support their use as in-vitro model systems to study specific cellular aspects of psychiatric disorders. iPSCs have provided insights into how genetic risk factors and environmental insults contribute to the onset of schizophrenia [46], iPSCs just like animal models cannot directly reflect the behavioral aspects of a psychiatric disorder, however, these cells can be differentiated into different brain cell types and have a number of experiments performed on them, having the advantage that their genomes come from actual patients so instead of studying a single isolated gene or mutation, one can study the cumulative effect of all the contributing loci. Besides the potential of identifying pathogenic mechanisms, patient-derived iPSCs can also be useful for finding pharmacological targets [12].

The neurodevelopment hypothesis of schizophrenia states that risk genetic and environmental factors of schizophrenia and bipolar disorder have an effect on the human brain during early development [22] [46] when cortical neurogenesis, axonal pathfinding, and neuronal functional development take place, these abnormalities during gestation are later followed by a gradual decrease of synapses during ado-

lescence, which is carried out by microglial cells, this process is known as synaptic pruning and it continues making changes in the cortical excitatory and inhibitory systems during early adulthood (the 20s-30s), which coincides with the age of onset of schizophrenia. During brain development, cortical neurogenesis represents a critical phase that is particularly susceptible to disruptions. The rate of gene expression alterations during this period is estimated to exceed 100 times the rate observed in the adult brain, genes associated with schizophrenia peak at prenatal weeks 16–19, and risk genes for bipolar disorder peak around weeks 19-22 [46].

Schizophrenia risk genes which are expressed during prenatal development, are often associated with cell fate determination and morphogenesis. [46] Studies utilizing postmortem analysis and brain imaging in individuals with schizophrenia, have identified abnormalities in cortical cell distribution and overall tissue organization as well as reduced density in synaptic spines causing thinning of the temporal and prefrontal cortices [22], potentially originating from disrupted brain development. Decreased density of parvalbumin (PV)-expressing interneurons in the prefrontal cortex (PFC), reduced thickness of the superficial cortical layers, and an increase in the volume of lateral ventricles have also been observed in these studies. However, the mechanisms through which these factors act, are not yet fully understood [46].

iPSCs-derived neurons can recapitulate some aspects of prenatal brain development and since they can be generated from patients, the associated genetic risk can also be studied, this makes iPSCs a valuable tool for studying schizophrenia and bipolar disorder among other psychiatric disorders[46]. iPSCs have already contributed valuable insights regarding the pathophysiology of psychiatric disorders, for instance, thanks to these models changes in cortical morphogenesis and cell type composition in schizophrenia patient-specific neuronal models have been associated with abnormal proliferation and differentiation of neural progenitor cells (NPCs)[46]. Among these alterations, an abnormal expression of WNT signaling pathway components in NPCs has also been observed in schizophrenia and bipolar disorder, suggesting a link with the altered proliferation of NPCs. Some studies using iPSCs have shown altered expression of ephrin-A ligands and receptors in cortical neurons and also in astrocytes suggesting that in schizophrenia, the varying expression of axonal guidance molecules by developing brain cells disrupts the proper connectivity of axons [46].

iPSCs studies have also shown that neurons derived from individuals with schizophre-

nia exhibit changes in the expression of GABA-synthesizing enzymes and variations in the expression of GABA receptor subunits, suggesting there might be a deficit in both glutamate and GABA-mediated neurotransmission in schizophrenia[46]. Also, abnormal excitation or inhibition has been observed in electrophysiological recordings of iPSC-derived neurons, depending on the patient's background. Altered miRNA expression, impaired mitochondrial function, and oxidative stress are also among the schizophrenia-related findings that have been possible thanks to research carried out on iPSCs derived from schizophrenia patients. Very similar findings have been collected for bipolar disorder where there's disruption of signaling pathways as well as abnormal mitochondrial function and miRNA expression and NPC proliferation [46][22]. One aspect that differentiates bipolar disorder from schizophrenia in iPSCs-derived neurons is their hyper-excitability, which can be brought back to normal after treatment with Lithium, a commonly prescribed drug for bipolar disorder-diagnosed patients but to which not everyone reacts.

Even tho studying iPSCs-derived neurons of a single type has been useful, the interaction between different cell types also needs to be taken into account, circuit-scale models including excitatory and inhibitory neuronal subtypes will be crucial, most current models can often only asses the early stages of neurogenesis, NPC development, and differentiation, also only a few reach initial stages of neural network formation . An in vitro model system has been developed to assess the circuit formation of iPSC-derived hippocampal DG and CA3 neurons, comparing these co-cultures from schizophrenia patients versus controls, the amount of spontaneous spike and network bursts were found to be smaller in the schizophrenia corresponding co-cultres [22]. Additionally, the integration of glial cell types to iPSCs-derived neuronal cultures could shed light on the neuronal malfunction, due to glial cells functioning as pacers for synaptic maturation [46][22]

Altogether iPSCs have provided valuable insights into the developmental origins of schizophrenia, we have now a better understanding of how disrupted signaling pathways are involved in the effective wiring of neuronal circuits as well as which neuronal cell subtypes take part. However, many cellular traits inherent to more mature stages and circuits like myelination and hippocampal circuit formation are yet to be further studied in vitro [46][22][12].

Figure 1 illustrates how different levels of genomic profiling relate and can be used together to gain insights into the biological mechanisms through which a disease is developed. Genome wide association studies are a great way to shed light into the genomic areas that contain variants associated to the trait or disease being analysed [66], however, this signal is often affected by the existance of linkage disequilibrium (LD) blocks which are regions in the genome that tend to be inherited together and move as a unit during recombination, therefore the fact that a SNP has a high association score (high $-log10P$) does not necessarily mean that it is a causal variant, it could simply be a neighboring SNP within the same LD block. Given that not all variants in the genome are equally relevant for all tissues or cell types, technologies like RNA sequencing and ATAC sequencing which capture the transcriptomic and chromatin landscape respectively of a cell or tissue, have been widely used to find out the mechanism of action of disease associated variants [69], this can done by simply overlapping disease associated variants that pass a specific P value threshold for association with a functional category of a cell type, this has already shed some light into what variants are relevant for what tissues. However more recent methods like the LD score regression method not only overlaps some variants onto functional categories but uses all GWAS interrogated variants while accounting for LD blocks, making sure to consider variants that don't pass the established P-value threshold but nevertheless contribute to the heritability of the trait or disease, at the same time this method explicitly models linkage disequilibrium. Leveraging this method and the rich availability of functional genomic annotations will be key to advance out understanding of polygenic disorders [18].

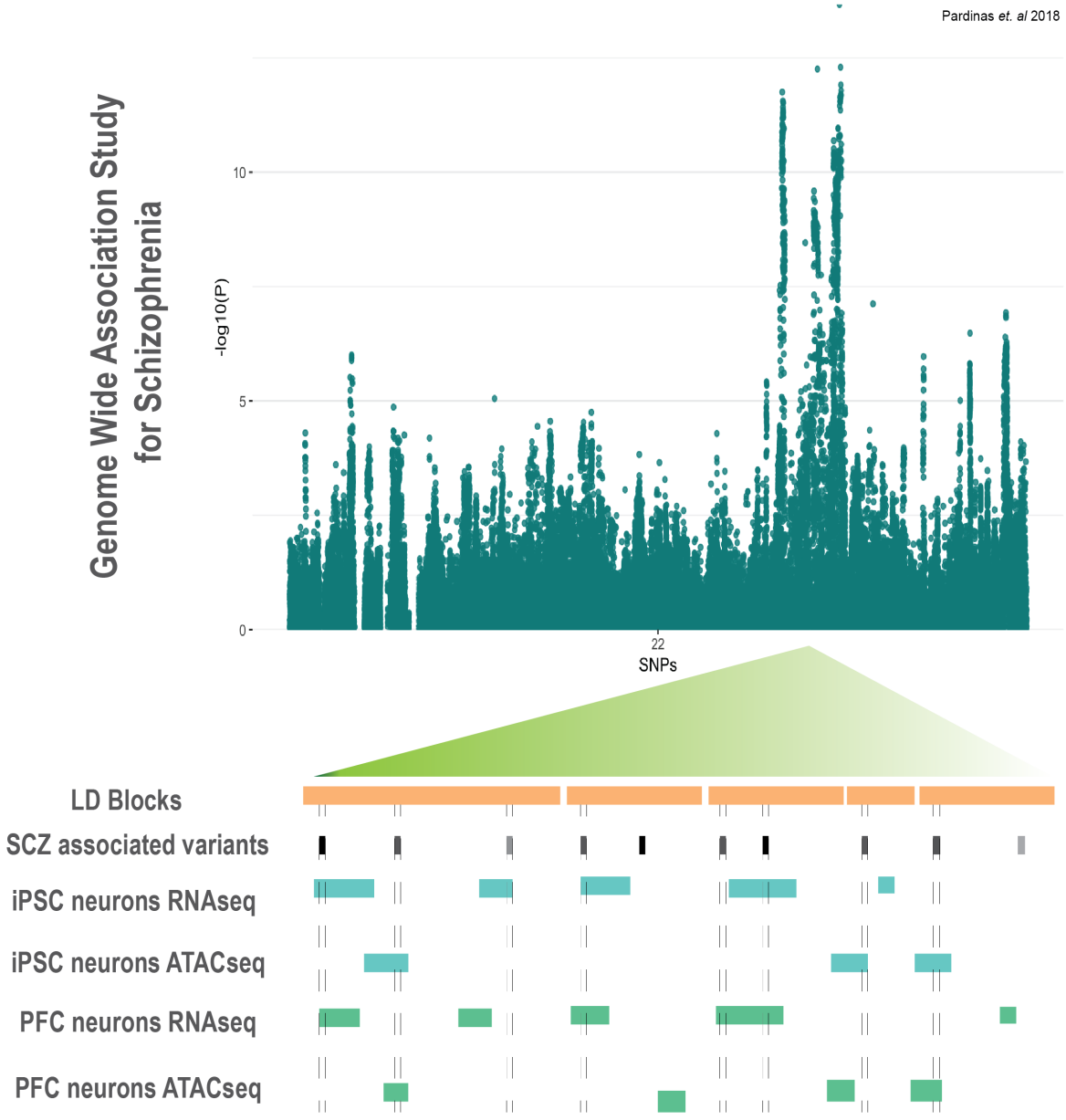

Figure 1: Overlap between GWAS variants, their corresponding LD blocks and the different methods to capture their function.

5 AIMS

The global aim of this doctoral thesis was to gain biological insights from psychiatric disorders by identifying relevant cell-types vulnerable to genetic risk and leverage this findings to benchmark suitable in vitro model systems capable of capturing the polygenic basis of mental illness. In more detail, this thesis aimed to

A) Identify which cell types are vulnerable to psychiatric disease-associated polygenic risk by making use of single-cell assay for transposase accessible chromatin sequencing (ATACseq) and RNA sequencing (RNAseq) signature genes and regulatory elements from specific neural subtypes obtained from post-mortem tissue to apply stratified LD−score regression for partitioning heritability from GWAS summary statistics while accounting for linkage disequilibrium.

B) Compare the results from the post-mortem snRANseq and scATACseq in point A) and utilize the most suitable technique for the benchmarking of in vitro model systems suitable to study cell-type specific SNP heritability.

C) Establish a genomic imputation pipeline and its application to case/control cohorts to calculate their polygenic risk for schizophrenia.

Since the available patient data only included diagnosis for schizophrenia and bipolar disorder, all analyses were focused on either of these two psychiatric diseases.

6 MATERIAL AND METHODS

6.1 Cohorts

6.1.1 PRONIA Cohort

The Personalised Prognostic Tools for Early Psychosis Management (PRONIA) is a study that collects information for healthy controls as well as people having clinical high risk (CHR) syndromes, recent onset psychosis, or recent onset depression (ROD) [52]. The PRONIA sample we had available through a collaboration with Prof. Dr. Nikolaos Koutsouleris, was composed by 342 Healthy controls, 245 subjects with Recent Onset psychosis, 237 with Clinical high risk symptoms, 2019 with Recent Onset Depression and 33 with unknown status (Fig2).

6.1.2 MIMICS plus Inhouse combined cohort

An In–House cohort for studying schizophrenia and bipolar disorder was put together resulting from a collaboration between the research groups of Prof. Dr. Michael Ziller at the Max Planck Institute of Psychiatry and Prof. Dr. Moritz Rossner at the LMU Klinikum. This resulted in cohort with a total of 101 controls and 91 subjects diagnosed with schizophrenia (Fig3).

6.2 Genomic imputation

All selected samples from both cohorts were genotyped using an Illumina Global Screening Array (GSA), containing a total of 654,027 interrogated genomic loci including 294,578 common variants, 68,431 rare variants and 3053 not present in the 1000 genomes project[25].

GSA array genotyping data for PRONIA cohort and MIMICS-In-House chorot samples corresponding to the case/control cohort used in this study was separately processed by Illumina Genome Studio, PLINK1.9 and PLINK2 were used for all quality control assessments, wrapper scripts for quality control and imputation were provided by Dr. Till Andlauer.

A per SNP and per sample missingness count was performed with parameters set

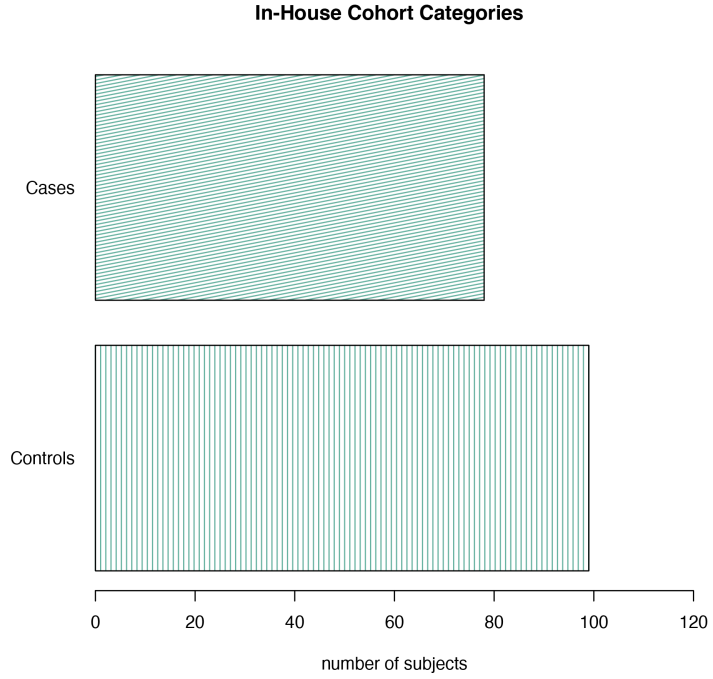

Figure 2: MIMICS Inhouse Cohort number of subjects per category

at —maf 0.01 —geno 0.02 —mind 0.02. The dataset was then filtered for sample repeats, identity by descent was used to identify close relatives which were temporarily removed (PI HAT threshold = 0.0625) to assess population stratification and identify population outliers. The latter was done by first pruning the SNPs by —geno 0.02 —hwe 1e–3 —indep–pairwise 200 100 0.2 —maf 0.05 —set–hh–missing and excluding the major histocompatibility complex (MHC) as well as the interstitial inverted duplication INV8, this was followed by creating an multidimensional scaling (MDS) clustering using the following parameters: —cluster —mds–plot 10 eigendecomp. Outliers were discarded using a threshold of 4 standard deviations and relatives were added back to the dataset. SNPs were evaluated and filtered based on Hardy–Weinberg equilibrium with a threshold of 1e–6. SNP names were updated to match those of the 1000 Genomes Phase 3 which was the reference panel used for imputing the samples. Duplicated SNPs were removed and remaining SNPs were checked and strand orientation corrected if needed.

After performing basic quality control on the original PRONIA cohort data 201108 SNPs and 6 subjects were removed due to low genotyping not passing the threshold both set to 80%. 8 samples were duplicated of another 8, and the samples with the lowest genotyping missing rate were kept, all these samples belong to the control group. A total of 18 samples were found to have a relative within the cohort.

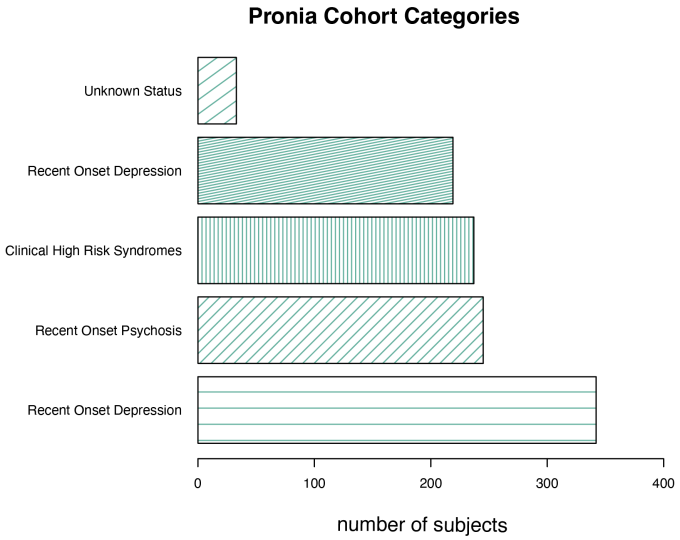

Figure 3: PRONIA Inhouse Cohort number of subjects per category

After obtaining the biological sex from analyzing heterozygosity for the sex chromosomes XY, 1 sample was reported as male and a SNP sex corresponding to female, when checking the number of heterozygous haploid SNPs (hh-freq 200) this same sample is found to likely be XXY. After performing a multidimensional analysis on the population components of the cohort, 18 samples that did not cluster with the rest of the samples were removed, figure 4 shows the first two components of the multidimensional analysis performed on the inbreeding coefficients for the PRONIA cohort and the 1000 genomes European sub−population, this was done to confirm the QCed cohort corresponded to the reported ancestry.

Each dataset was then phased using shapeit v2.r837 and windows of 5Mb were imputed using IMPUTE2 version 2.3.2 using the parameters –pgs–miss –filt–rules–l –buffer 500. Post–imputation quality control was applied by identifying empty or low SNP count blocks which were either discarded or reimputed by moving the window location and/or size. In order to filter high quality imputed SNPs, an info score great than 0.8 and minor allele frequency threshold of 0.01 were used to select the final SNPs [23].

6.3 Polygenic Risk Scores calculation

After genomic imputation was performed, the software package PRSice [15] was used to calculate standard polygenic risk scores for schizophrenia on all samples from the MIMICS-In-House cohort (target data set 1). A susbset of the PRONIA

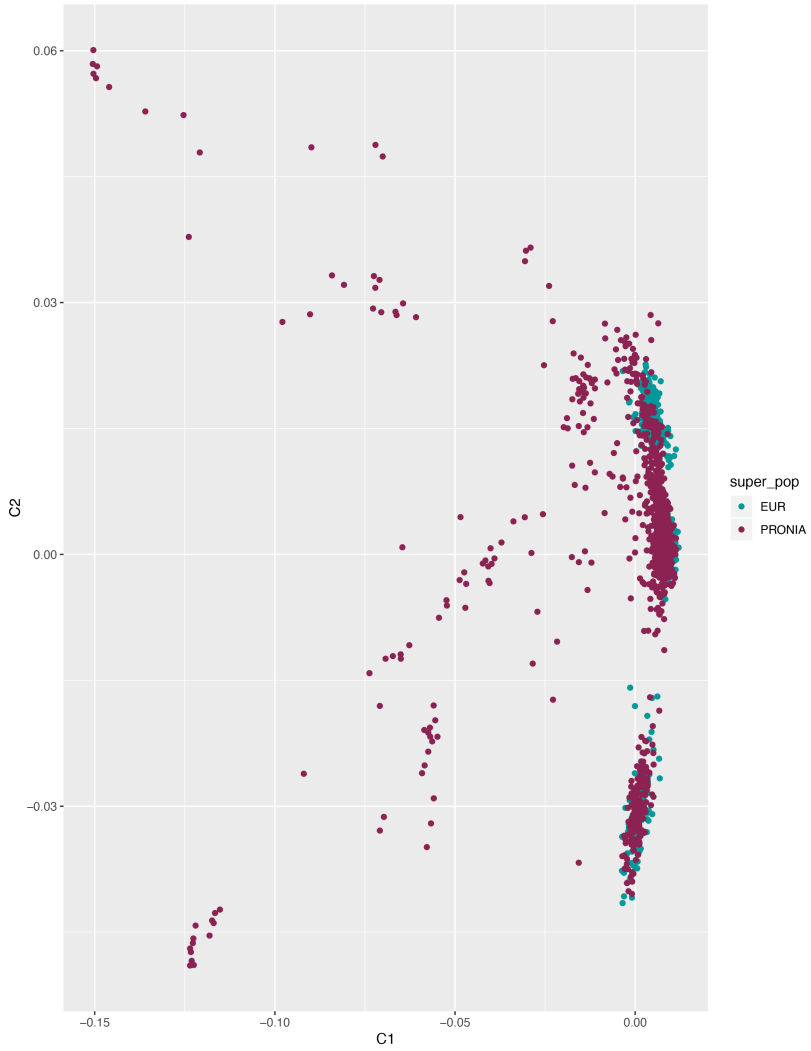

Figure 4: Complete PRONIA Cohort Multidimensional analysis plot overlapped with the 1000 genomes European population.

cohort was selected for polygenic risk score calculation, the included samples were the 305 healthy controls and the 195 diagnosed with recent onset psychosis (target data set 2).

As the "base phenotype" I used the GWAS summary statistics from the schizophrenia study from Pardiñas et. al 2018 [40] which were obtained from the Alkes research group at the BROAD institue in the following link:

https://data.broadinstitute.org/alkesgroup/

Each of the target data sets were matched to the base phenotype data and those SNPs which were absent in one of the data sets were removed. Polygenic risk scores were then calculated with report.individual.scores set to true to report all calculated scores for all thresholds for each individual, the –binary-target parameter was also set to True indicating this is a binary disease where the sample is either a case or a control, –stat was set to OR to indicate the column from the base phenotype file that contains the effect sizes, the –clump.kb flag was set to 500 kilobases as the window size to define a linkage disequilibrium independent SNP. A bar plot chart showing the R2 at the different evaluated p–values was generated (see results) and only the best fit threshold derived Polygenic risk scores were assessed. The fastscore flag was set to true so scores would be evaluate at 7 different thresholds between 0.0001 and 0.5, which are the default boundaries. A second run was performed without the fastscore flag in order to calculate High resolution polygenic risk scores at all P-value thresholds, a plot showing the model fit at the broad P-value thresholds was generated (see results).

6.4 Copy Number Variant Analysis

Given that a known number of rare copy number variants can be directly associated with causing schizophrenia [48] , copy number variants were called on all the samples for which iPSCs were generated to make sure non of these cases were casued by any known schizphrenia associated copy number variant.

Copy number variants (CNVs) were called by Dr. Lucia Trastulla using a pipeline she developed that includes a BCFtools wraper of their variant caller algorithm [2].

CNVs were called in all our samples and later compared to the known disease associated CNVs compilated by Marschall et al. 2016 [48], as they report the CNVs

in the genomic assebly hg18, I first lifted over the corresponding genomic coordinates from hg18 to hg19 genome assembly using UCSC liftover. The resulting coordinates were then compared to the ones called on our cohort by overlapping their genomic coordinates, this was done using bedtools intersect v2.26.0, a match was only called when a minimum overlap of 90% of both CNVs being compared was covered.

6.5 Number of schizophrenia associated alleles per sample in the MIMICS-In-House Cohort

A genome-wide selection of schizophrenia associated SNPs from the 2014 genome-wide association study [50] with a P-value threshold, resulting in a set of variants and the corresponding schizophrenia assciated allle per GRE. The number of schizophrenia-associated alleles in these sets was assessed in all samples from the MIMICS-In-House Cohort, the goal was to find out whether there was a significant difference of the number of schizophrenia associated alleles in cases versus healthy controls. Figure 5 shows an example of the overlap between the GWAS variants and the selected subset across chromosome 1. The heatmap shows the number of schizophrenia associated alleles per sample for all selected variants, this ranges from 0 to 2 with white meaning that the locus was not genotyped and therefore there is no allele information available.

6.6 Human post-mortem brain tissue characterization and heritability assessment

6.6.1 Post mortem brain tissue snRNAseq and ATACseq

All experiments in post-mortem brain tissue performed by Dr. Miriam Gagliardi, from which she generated scRNA-seq and scATACseq data from the pre frontal cortex of 4 adult individuals corresponding to two controls and two schizophrenia cases RNAseq and ATACseq data was processed by Dr. Ziller.

6.6.2 Stratified LD–score regression for cell type specific heritability enrichment assessment in post-mortem brain tissue

After obtaining the expressed gene list for each identified cell type resulting from snRNAseq data, the corresponding gene coordinates were obtained using ENSEMBL biomaRt with hsapiens–gene–ensembl dataset GRCh37, a 100–kb window was added to each side of the transcribed region. These resulting extended gene coordinates

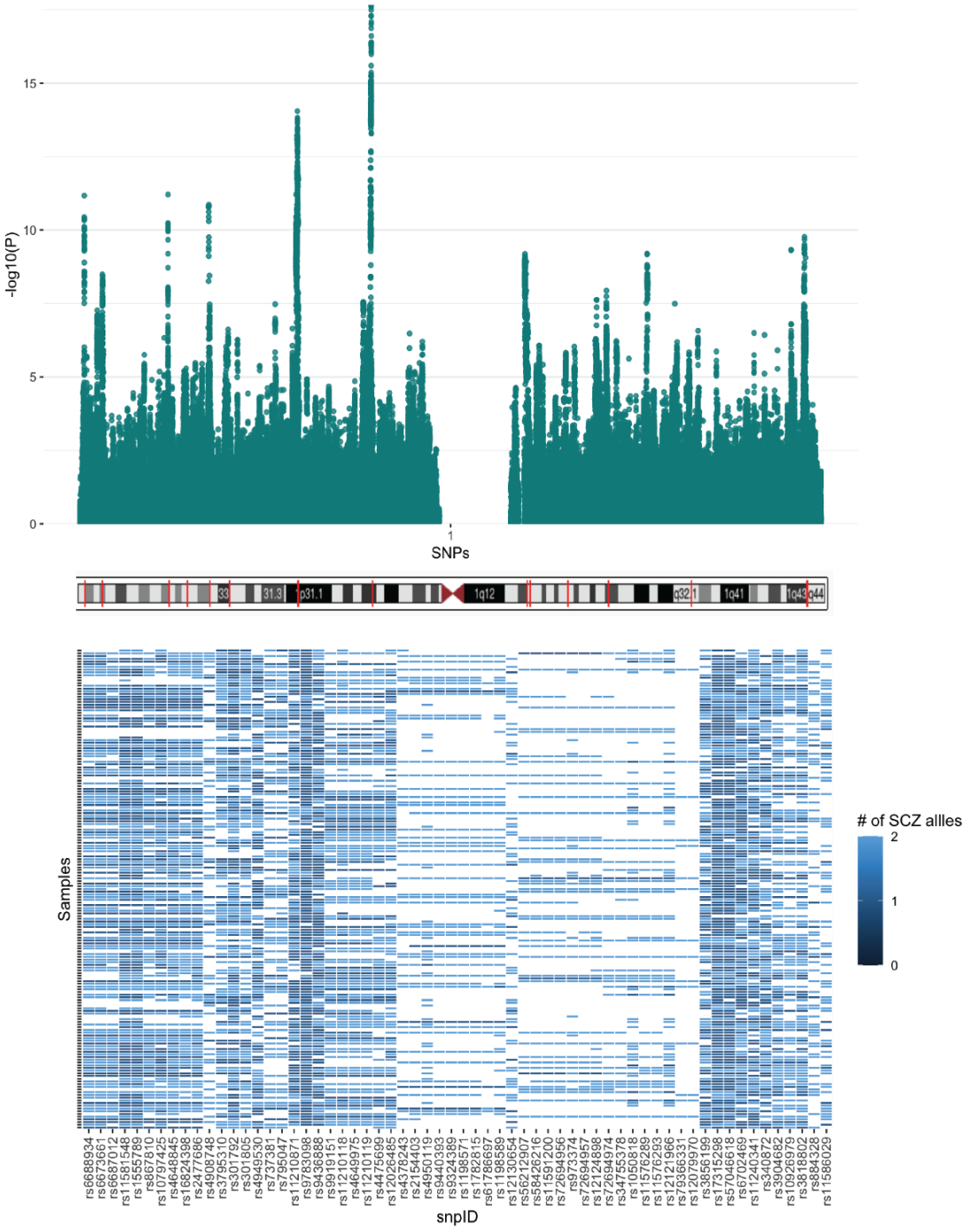

Figure 5: All tested SNPs for chromosome 1 in the Schizphrenia GWAS by Ripke et al.[50] are displayed in the manhattan plot at the top, representative SNPs determined to be significantly associated with schizophreia are shown as red lines in the chromosome cytoband plot in the middle to show their position within the chromosome. The heatmap at the botton shows the number of schizophrenia-asscoiated alleles for each of the selected SNPs for which their SNP IDs are displayed on the X-axis of the plot, each row of the heatmap represents a sample. Each square represents the number of schizophrenia associated alleles for a specific SNP in a given sample, it ranges from $0 - 2$ and a missing genotype is represented by a white colored square.

were then overlapped with the LD scores files from the reference panel using bedtools Version: v2.27.1–1–gb87c465. and formatted as input to the S–LDSC software. LD scores were then calculated separately for each cell type using LD Score Regression (LDSC) Version 1.0.0 [17]

All cell types were then analyzed together by performing a cell–type–specific analysis to estimate disease association for 19 different polygenic traits divided into 4 classifications. For psychiatric disorders, I analyzed enrichment in schizophrenia, bipolar disorder, autism spectrum disorders, and major depressive disorder, corresponding to Central Nervous System (CNS) traits I included the polygenic traits of antidepresant treatment resistance, any psychotic experience, clozapine levels, norclozapine levels, clozapine to norclozapine ratio, years of education, and Neuroticism. Two different groups of controls were assessed, one control group corresponded to polygenic diseases which included Coronary Artery Disease (CAD), celiac disease, Type I and type II diabetes, for the second group I included polygenic traits including height, fasting glucose levels, if ever smoked, and low-density lipoprotein (LDL) cholesterol levels. This was carried out using each disease's corresponding GWAS summary statistics obtained from:

https://data.broadinstitute.org/alkesgroup/

A modified baseline model v1.2 containing 52 categories was built to run this analysis. The modified baseline model contains an extra category as a control that includes all analyzed genes in our RNA–seq or ATACseq, similar to Finucane et al 2018 [18].To account for multiple testing, significant enrichment was determined after correcting using FDR < 0.05 over the 10 included cell type categories.

The cell–type–specific analysis output includes a regression coefficient, coefficient standard error and a coefficient P value from which Z–scores were calculated and plotted in heat maps for all categories and traits. using R version 3.5.2 (2018–12–20) — "Eggshell Igloo"

6.7 Model Systems to capture the polygenic basis of mental illness

6.7.1 Charachterization chromatin accessibility profiles of cerebral organoid-derived cells by scATACseq

Cerebral organoids and their chromatin accessibility characterization are a courtesy of the Dr. Silvia Cappello's research group at Max Planck Institue for Psychiatry. Data processing including label transfer and uniform manifold approximation and projection clustering was performed by Prof. Dr. Michael Ziller and provided to me as genomic tracks in bed format.

6.7.2 Stratified LD–score regression for cell type specific heritability enrichment assessment in cerebral organoids

Genomic tracks in bed format were provided by Dr. Micahel Ziller and peak genomic locations were intersected with the LDSC annotation files using bedtools Version: v2.27.1–1–gb87c465. and formatted as input to the S–LDSC software. LD scores were then calculated separately for each category (positive and negative peaks) using LD Score Regression (LDSC) Version 1.0.0 [17]

All organoid clusters were then analyzed together by performing a cell–type–specific analysis to estimate disease association for 13 different polygenic traits divided into 4 classifications. For psychiatric disorders, I analyzed enrichment in schizophrenia, bipolar disorder, autism spectrum disorders, and major depressive disorder, corresponding to Central Nervous System (CNS) traits I included the polygenic traits intelligence, educational attainment (EA), and Neuroticism. Two different groups of controls were assessed, one control group corresponded to polygenic diseases which included Cardiovascular Disease, celiac disease, and asthma, for the second group I included polygenic traits including height, lymphocyte count, and low-density lipoprotein (LDL) cholesterol levels. This was carried out using each disease's corresponding GWAS summary statistics obtained from

https://data.broadinstitute.org/alkesgroup/

To account for multiple testing, significant enrichment was determined after correcting using FDR < 0.05 over the 10 included cell type categories.

The cell–type–specific analysis output includes a regression coefficient, coeffi-

cient standard error and a coefficient P value from which Z–scores were calculated and plotted in heat maps for all categories and traits including iPSC differentiated iNeurons for comparison, plotting was done using R version 3.5.2 (2018–12–20) — "Eggshell Igloo"

6.7.3 iPSC–derived cortical neurons

iPSC–derived cortical neurons were obtained by Dr. Ruhel Ahmad at Max Planck Institue for Psychiatry based on the protocol from Qi et. al (2017). To briefly describe the procedure. The iPSCs were propagated on the cell culture dishes coated with Matrigel (1:100) (Corning) diluted with DMEM/F12 (Thermo Fisher Scientific) in StemMACS iPS–Brew XF medium (Miltenyi Biotec). For neuronal differentiation, iPSCs were disassociated with Accutase (Sigma–Aldrich) at 37°C for 5 minutes. After washing and centrifugation cells were plated on the Matrigel (1:30) coated tissue culture dishes in StemMACS iPS–Brew XF medium with RevitaCell™ Supplement (Thermo Fisher Scientific) at a density of 200,000 cells/cm2. On the next day (day 0) differentiation was started by changing the medium to KSR medium containing 410 ml of Knockout DMEM, 75 ml Knockout Serum Replacement, 1 mM L–glutamine, 100 µM MEM nonessential amino acids, and 0.1 mM beta–mercaptoethanol (all from Thermo Fisher Scientific) including small molecules LDN193189 (250 nM; Miltenyi Biotec), SB431542 (10 µM; Miltenyi Biotec), XAV939 (5 µM; Tocris), PD0325901 (1 µM; Miltenyi Biotec), SU5402 (5 µM; Tocris), DAPT (10 µM; Miltenyi Biotec). LSB+X were added from day 0–7, and P/S/D were added from day 2–7. N2 and B27 supplements (Life Technologies) were added in increasing 1/3 increments every other day from day 4: 1/3 N2/B27 for days 4 and 5, 2/3 N2/B27 for days 6, 7. For initiating Neuronal differentiation from NPCs. On day 8 NPCs were dissociated to single cells with Accutase (Sigma–Aldrich). Cells were resuspended in neural differentiation media containing neurobasal (Thermo Fisher Scientific) supplemented with B27, BDNF (20 ng/ml; Miltenyi Biotec), cAMP (0.5 mM; Sigma–Aldrich), laminin (1 µg/ml; Sigma–Aldrich), and ascorbic acid (0.2 mM; Sigma–Aldrich). After centrifugation cells were plated on polyornithine (PO; 15 µg/ml), laminin (1 µg/ml), fibronectin (2 µg/ml) (all from Sigma–Aldrich) at 300,000 cells/cm2 in neural differentiation media. On day 14 mouse astrocytes were added to the neurons at a density of 5000 cells/cm2. On day 21 changed media which included Ara–C (4µM; Sigma–Aldrich). Until day 63 medium was changed every 3–4 days.

6.7.4 iPSC–derived GABAergic neuron

For the derivation of GABAergic neuron (iG) from iPSCs, obtained by Dr. Ruhel Ahmad at Max Planck Institue for Psychiatry he followed Yang e.t al. (2017) differentiation protocol. Briefly, iPSCs were propagated on Matrigel (1:100) (Corning) coated cell culture dishes in StemMACS iPS–Brew XF medium. iG were derived by dissociating iPSCs with accutase and infected with lentivirus. The following lentivirus constructs were used: FUW M2rtTA expressing the reverse tetracycline transactivator (M2rtTA) FUW–TetO–Ascl1–T2A puromycin expressing Ascl1 gene, FUW–TetO–Dlx2–IRES–hygromycin expressing Dlx2 gene. Cells were resuspended in Stem Brew supplemented with Polybrene (6ug/ml) & 1x RevitaCell and seed cells onto Matrigel–coated dishes at a density of 20000 cells/ cm2. After 16–18 hours, remove the virus–containing media and add N2 media (DMEM/F12 medium and 5 ml 100x N2 supplement) with Doxycycline 2 µg/ml. After 24 hours, change the media with N2 media with Doxycycline and antibiotics (puromycin 10 µg/ml and hygromycin 5 µg/ml). On day 4 Change the media with N2 media with Doxycycline then again on day 7 change the media with N2 media with Doxycycline and Ara–C (4µM). Cells were dissociated on Day 8 and seed onto Poly–Ornithine/Laminin/Fibronectin coated plates in Growth Medium with Doxycycline (2×105 cells/cm^2). Growth medium: Neurobasal Medium, 1 mM L–glutamine, 10 mL B–27 Supplement, 5% FBS and BDNF (10 ng/ml). On Day 11 mouse astrocytes were added to the iG culture at a density of 5000 cells/cm^2. On Day 15 changed half of the Growth Medium with Doxycycline and Ara–C (4µM). From Day 18 onwards Doxycycline was removed from the Growth Medium. Half media was changed every 3–4 days until day 42.

6.7.5 Oligodendroglial differentiation from human stem cells

Neural induction and oligodendroglial differentiation was performed by Dr. Florian Rabee at the Department of Psychiatry and Psychotherapy at the LMU Klinikum adapted to previously published protocols (García–León et al., 2018). Briefly, hiPSCs were cultivated in feeder–free conditions with iPS–Brew on Vitronectin. Transfection with lentiviral construct, based on the pINDUCER21–puro–Gateway–3xFLAG (Addgene plasmid #172981) that contained SOX10–P2A–OLIG2–T2A–NKX6.2 (SON) under the control of the doxycycline–inducible operator (tetO) and constitutive expression units for the reversed tetracycline transactivator (rtTA) and a puromycin selection cassette, was performed with single cells in iPS–Brew supplemented with 1 µM Y–27632. 48 h after transfection, selection with 1 ug/mL puromycin was per-

formed to generate stable SON–hiPSCs. For neuronal induction and oligodendroglial prepatterning (sample day13), hiPSC, cultured in mTeSR1, were first plated on Matrigel. For passaging mTeSR1 was supplemented with RevitaCell. Two days later, media was changed to N2B27, which consists of DMEM/F–12, 1x GlutaMAX, 1x N2, 1x NEAA, 50 µM Mercaptoethanol and 25 µg/mL Insulin, supplemented with 100 nM retinoic acid (RA)and 10 µM SB431542, 1 µM LDN193189. Media was changed every day. After 8 days, SB and LDN were replaced by 1 µM SAG until day 13. For directed oligodendrocyte differentiation, cells after neural induction were plated on 50 µg/mL poly–L–ornithine/10 µg/ml laminin, on the next day (day+0) media directed OL differentiation was initiated changing medium to Oligodendrocyte differentiation medium (OL–DM), which consists of N2B27 supplemented with 10 ng/mL PDGF–AA, 10 ng/mL IGF1, 5 ng/mL HGF, 10 ng/mL NT3, 0.1 ng/mL Biotin, 1 mM dbcAMP, 60 ng/mL T3 and 1 µg/mL doxycycline. Media was changed every second day and puromycin (1µg/ml) selection was performed from Day +2 to Day +4. To enrich for O4+ cells, we performed O4–microbead purification to the manufacturer's instructions at day +10 and cells were cultivated until day +16.

6.7.6 Characterizing patient-derived iNeurons

Primary human cells from healthy controls and schizophrenia cases were used by Christine Rummel to generate iPSCs, which she then differentiated into excitatory forebrain neurons using a modified protocol by Zhang et al. [72] cell's identity characterization were assessed using immunocytochemistry, bulk, and single-cell RNA-Seq. Electrophysiological activity of the generated iNeurons derived from both patients and controls was also recorded and showed network activity and recurrent bursting behavior. Neurite and cortical layer markers were also measured to further confirm neuronal identity.

6.7.7 Stratified LD–score regression for cell type specific heritability enrichment assessment in patient-derived iPSCs

scATACseq data from pateint-derived iPSCs differentiated cells was obatined and processed by members of the Ziller research group at Max Planck institute for psychiatry.

Genomic locations of peaks from scATACseq data were overlapped with the LD scores files from the reference panel using bedtools Version: v2.27.1–1–gb87c465. and formatted as input to the S–LDSC software. LD scores were then calculated

separately for each cell type using LD Score Regression (LDSC) Version 1.0.0 [17]

All cell types were then analyzed together by performing a cell–type–specific analysis to estimate disease association for 13 different polygenic traits divided into 4 classifications. For psychiatric disorders, I analyzed enrichment in schizophrenia, bipolar disorder, autism spectrum disorders, and major depressive disorder, corresponding to Central Nervous System (CNS) traits I included the polygenic traits intelligence, educational attainment (EA), and Neuroticism. Two different groups of controls were assessed, one control group corresponded to polygenic diseases which included Cardiovascular Disease, celiac disease, and asthma, for the second group I included polygenic traits including height, lymphocyte count, and low-density lipoprotein (LDL) cholesterol levels. This was carried out using each disease's corresponding GWAS summary statistics obtained from

https://data.broadinstitute.org/alkesgroup/

A modified baseline model v1.2 containing 52 categories was built to run this analysis. The modified baseline model contains an extra category as a control that included the DNase I hypersensitive sites from the ENCODE project, similar to the Finucane et al 2018 [18].To account for multiple testing, significant enrichment was determined after correcting using FDR < 0.05 over the 10 included cell type categories.

The cell–type–specific analysis output includes a regression coefficient, coefficient standard error and a coefficient P value from which Z–scores were calculated and plotted in heat maps for all categories and traits. using R version 3.5.2 (2018–12–20) — "Eggshell Igloo"

6.7.8 Charachterization chromatin accessibility profiles of mouse prefrontal cortical cells

scATACseq data from mouse prefrontal cortex at post-natal days 1,7 and 21 (P1, P7 P21) was obtained from Yuan et al [71] available through the Gene Expression Omnibus (GEO SuperSeries GSE204851). The data set includes different brain cell types, corresonding to P1, Astrocytes, oligodendrocyte progenitor cells, corticothalamic projection neurons, cortical projection neurons, and Inhibitory MGE and CGE neurons are included in this analysis, For P7 sub-cerebral projection neurons, neuropeptide Y-expressing neurons, cortical projection neurons, corticothalamic pro-

jection neurons, astrocytes, somatostatin expressing neurons, and oligodendrocyte progenitor cells, for P21 neuropeptide Y-expressing neurons, parvalbumin-positive neurons, somatostatin expressing neurons, cortical projection neurons, NFOL neurons, sub-cerebral projection neurons, corticothalamic projection neurons, layer IV neurons and Astrocytes, and MFOL expressing neurons were analysed.

6.7.9 Stratified LD–score regression for cell type specific heritability enrichment assessment in murine prefrontal cortical cells

Peak genomic locations in bed format were intersected with the LDSC annotation files using bedtools Version: v2.27.1–1–gb87c465. and formatted as input to the S–LDSC software. LD scores were then calculated separately for each of the murine cortical cell types at the 3 different post-natal stages using LD Score Regression (LDSC) Version 1.0.0

A cell–type–specific analysis was performed on each cell-type chromatin accessibility profile to estimate disease association for the following polygenic traits and diseases: schizophrenia, bipolar disorder, autism spectrum disorders, major depressive disorder, intelligence, educational attainment (EA), and neuroticism. Control traits and diseases were also analyzed including cardiovascular Disease, celiac disease, asthma, height, lymphocyte count, and low-density lipoprotein (LDL) cholesterol levels. This was carried out using each disease's corresponding GWAS summary statistics and a modified baseline model v1.1 which contains 52 categories. The modified baseline model contains an extra category as a control that included the mouse DNase I hypersensitive sites from the ENCODE project. Summary statistics and baseline model were downloaded from:

https://data.broadinstitute.org/alkesgroup/

To account for multiple testing, significant enrichment was determined after correcting using FDR < 0.05

The cell–type–specific analysis output includes a regression coefficient, coefficient standard error and a coefficient P value from which Z–scores were calculated and plotted in heat maps for all categories and traits. using R version 3.5.2 (2018–12–20) — "Eggshell Igloo"

6.7.10 Mouse and human fetal callosal neurons of the developing neocortex

Zuccaro et al. [73] generated a longitudinal transcriptional map of excitatory projection neuron (PN) and inhibitory interneuron (IN) subtypes of the cerebral cortex, across a timeline of mouse embryonic and postnatal development, as well as fetal human cortex and human cortical organoids. Major subtypes of cortical excitatory and inhibitory neurons were profiled at different time points including excitatory projection neurons, corticothalamic projection neurons (CThPNs) of layer 6, subcerebral projection neurons (ScPNs) of layer 5, and cortical projection neurons in the mouse cortex. The profiled cell types were identified by the genetic markers Bcl11b, Tle4, and Satb2 that charachterize ScPNs, CThPNs, and CPNs, respectively.

6.7.11 Stratified LD–score regression for cell type specific heritability enrichment assessment in callosal neurons of the developing neocortex

To evaluate the association of different neuronal signatures with neuropsychiatric disease risk loci, I performed stratified LD-score regression on cell-type specific chromatin profiles obtained by single-cell ATACseq from mouse embryonic and postnatal development, as well as fetal human cortex and human cortical organoids.

6.7.12 Stratified LD–score regression for heritability enrichment assessment in hKCl depolarizing treated iNeurons VS untreated iNeurons

High potassium chloride (hKCl) depolarizing treatment was added to iNeurons and the resulting chromatin accessibility profile was obtained with ATAC−Seq. These experiments were performed by Dr. Christine Rummel, scATACseq data was processed by Prof. Dr. Michael Ziller and provided to me as genomic tracks in bed format.

hKCL peak genomic locations in bed format were intersected with the LDSC annotation files using bedtools Version: v2.27.1–1–gb87c465. and formatted as input to the S–LDSC software. LD scores were then calculated separately for each category (positive and negative peaks) using LD Score Regression (LDSC) Version 1.0.0

For both categories, a cell–type–specific analysis was performed to estimate disease association for the following traits: Major Depressive Disorder, Antidepressant

treatment resist, any psychotic experience, Schizophrenia, Ever Smoked, Autism Spectrum Disorder, Bipolar Disorder, Coronary Artery, Disease, Height, Type 1 Diabetes, Type 2 Diabetes, Neuroticism, LDL, Celiac, Fasting Glucose, Years of Education. This was carried out using each disease's corresponding GWAS summary statistics and a modified baseline model v1.1 which contains 52 categories. The modified baseline model contains an extra category as a control that included the DNase I hypersensitive sites from the ENCODE project, similar to the Finucane et al 2018 [18]. To account for multiple testing, significant enrichment was determined after correcting using FDR < 0.05

The cell–type–specific analysis output includes a regression coefficient, coefficient standard error and a coefficient P value from which Z–scores were calculated and plotted in heat maps for all categories and traits. using R version 3.5.2 (2018–12–20) — "Eggshell Igloo"

7 RESULTS

7.1 Genomic imputation

7.1.1 PRONIA

A total of 7.2M variants resulted from the imputation of the PRONIA cohort, Figure 6 shows the relative variance explained by each of the components based on the population structure of the PRONIA cohort, this was evaluated by applying multidimensional scaling to the inbreeding coefficients of the cohort, the first two components explain the majority of the variance and therefore could be used to determine if any and which samples needed to be discarded. The observed and expected autosomal homozygous genotype counts were computed for all samples, Figure 7 shows the method-of-moments F coefficient estimates, all samples are expected to be within the range delineated by the red lines, samples that don't meet these values need to be discarded, a total of 2 control samples were discarded. Figure 8 shows the first 2 components of the population-based multidimensional analysis for the after-imputation PRONIA cohort, each dot corresponds to a sample and they are colored according to biological sex, confirming there is no clustering according to this variable. SNP-based as well as individual-based genotyping missingness was calculated, the after-imputation missingness results are shown in Figure 9 as the cumulative distribution of individual and SNP call rates, the Y axis indicates the genotyping call rate, and the X-axis represents the quantiles to which each individual or SNP belong, both call rate measurements are above 0.9 showing that more than 90% of the total SNPs were called for each individual.

7.1.2 MIMICS InHouse

A total of 7M variants resulted from the imputation of the MIMICS-InHouse cohort. The relative variance corresponding to each of the components was calculated for the population structure within the cohort using multidimensional scaling and it is shown in Figure 10, components 1 and 2 explain the majority of the observed variance. Heterozygosity was calculated for all samples counting the observed and expected autosomal homozygous genotype and reported using the method-of-moments F coefficient estimates, Figure 11 shows the sample counts (Y-axis) for the different F coefficient, two samples that fell outside the range delineated by the red vertical lines were taken out of the analysis due to abnormal heterozygosity. Figure

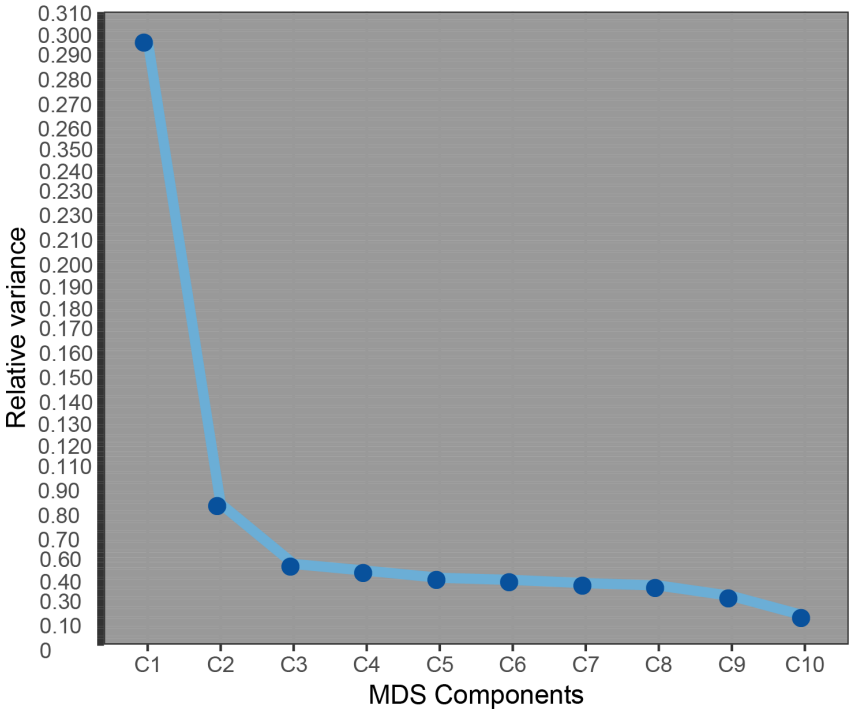

Figure 6: Relative variance explained by each component calculated on the population structre of the PRONIA Cohort.

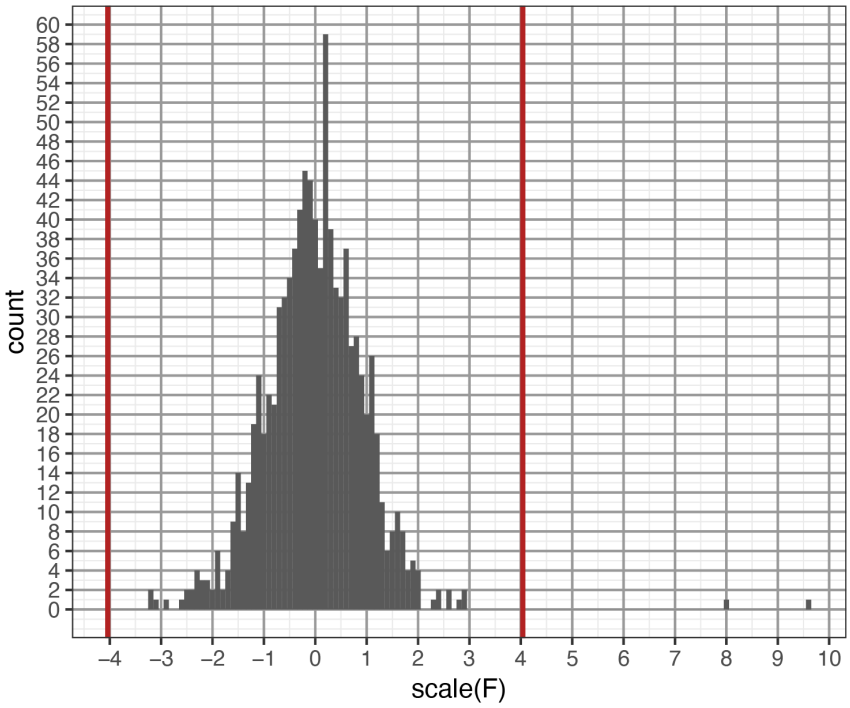

Figure 7: F coefficient estimates of heterozygosity for the PRONIA Cohort.

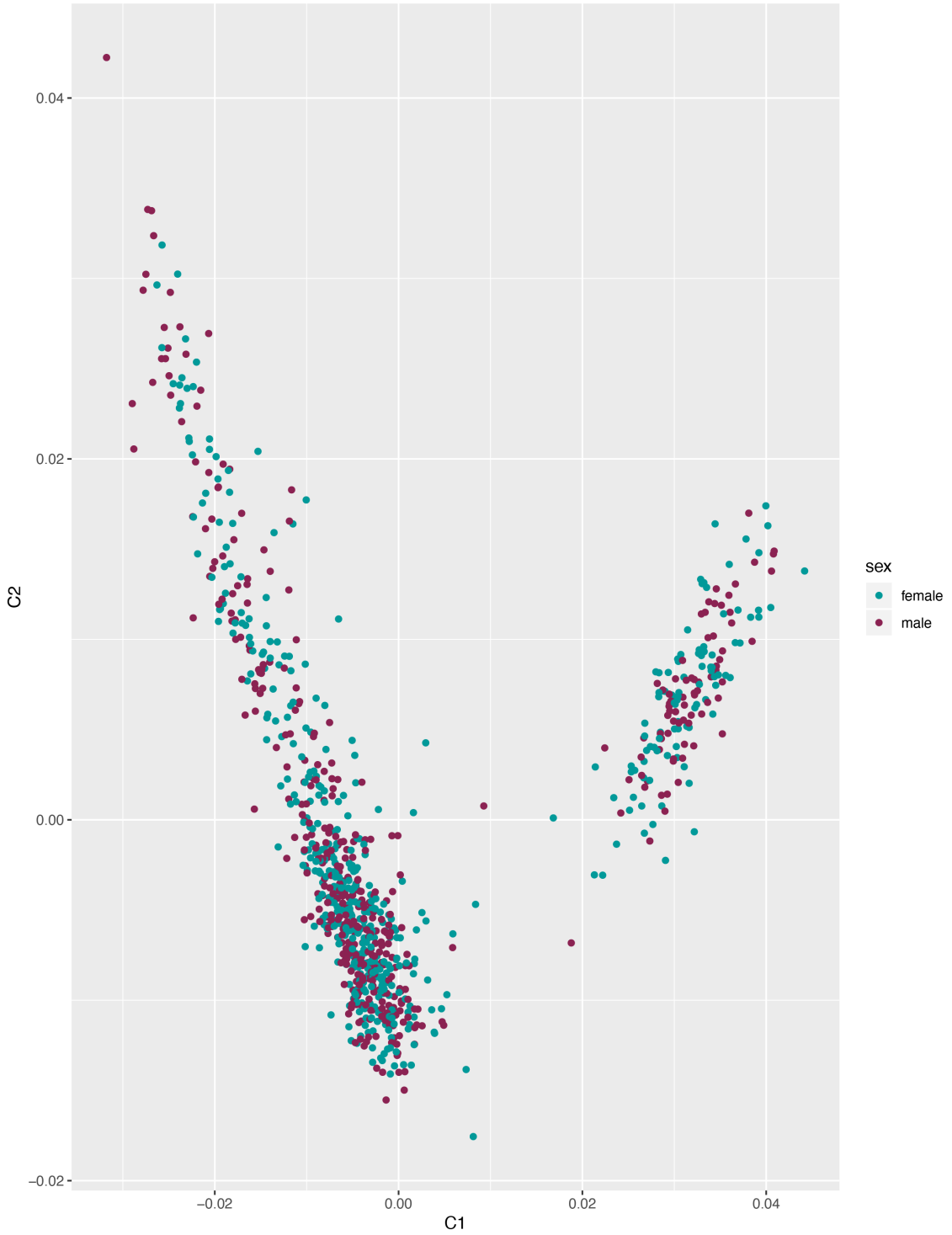

Figure 8: First two components of the inbreeding coefficients multidimensional analysis of the PRONIA cohort after imputation, the color code corresponds to biological sex.

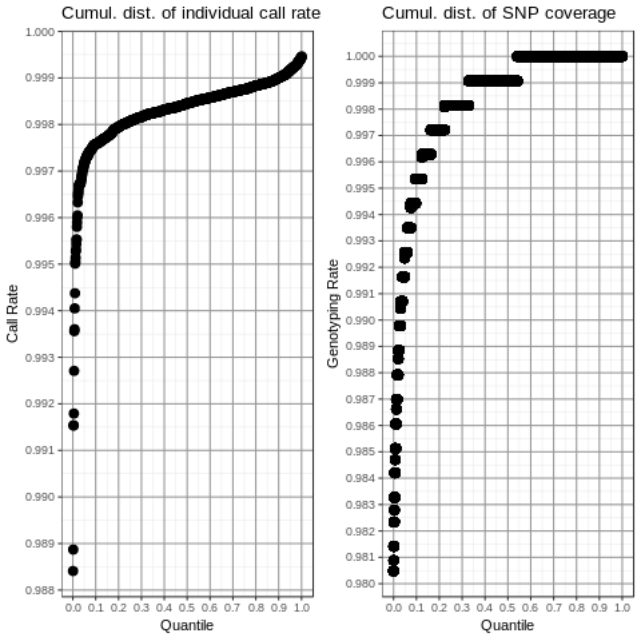

Figure 9: Missingness per sample and SNP post-imputation of the PRONIA cohort.

12 shows the first 2 components of the population-based multidimensional analysis for the after-imputation MIMICS InHouse cohort, each dot corresponds to a sample and they are colored according to biological sex, confirming there is no clustering according to this variable. Some samples do not cluster with the main cluster but we decided to keep them given that they are only 2 standard deviations higher than the usual threshold of 4 standard deviations. In Figure 13 the post imputation cumulative distribution of individual call rate and SNP call rate are plotted. On the left, each dot represents a subject from the MIMICS-In-house cohort, and on the right each dot represents a SNP, both call rate measurements are above 0.9 indicating that more than 90% of the total SNPs were called for each sample.

7.2 Polygenic Risk Scores calculation

Standard polygenic risk scores for schizophrenia were calculated for all imputed samples in the PRONIA cohort as well as for all samples in the MIMICS - InHouse cohort. Figures 14 and 16 show the goodness of fit for the association between the polygenic risk scores calculated at different P-value thresholds and schizophrenia for PRONIA and MIMICS-InHouse respectively. Each bar corresponds to the result of a tested polygenic risk score that included the SNPs passing a given P-value threshold in the schizophrenia GWAS. The P-value used for each PRS (bar) is indicated

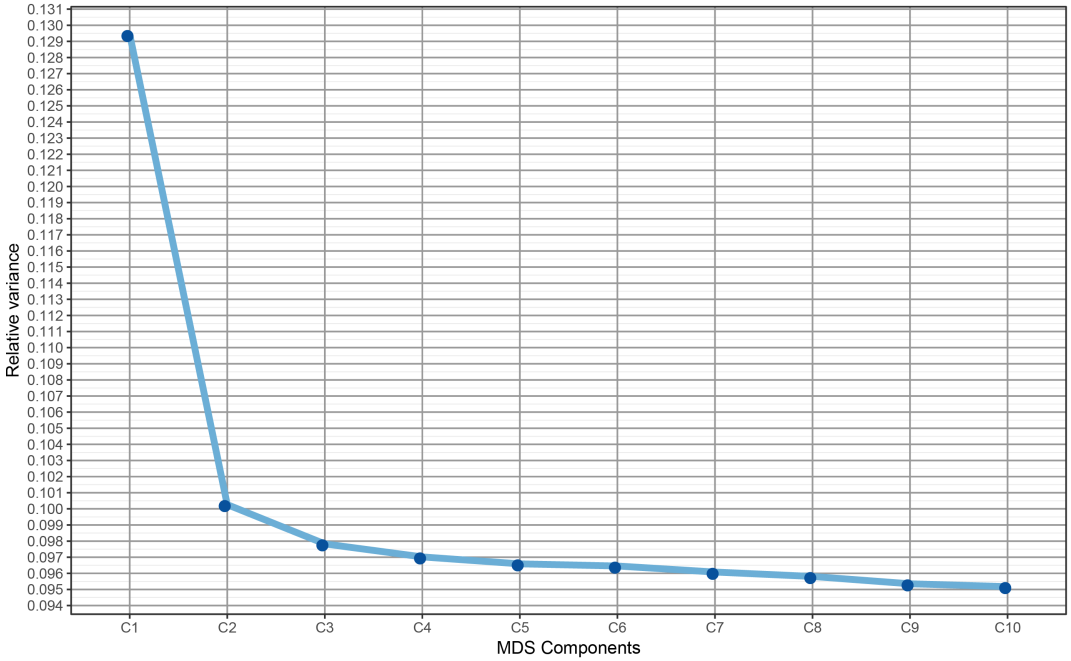

Figure 10: Relative variance explained by each component calculated on the population structre of the MIMICS-In-House cohort

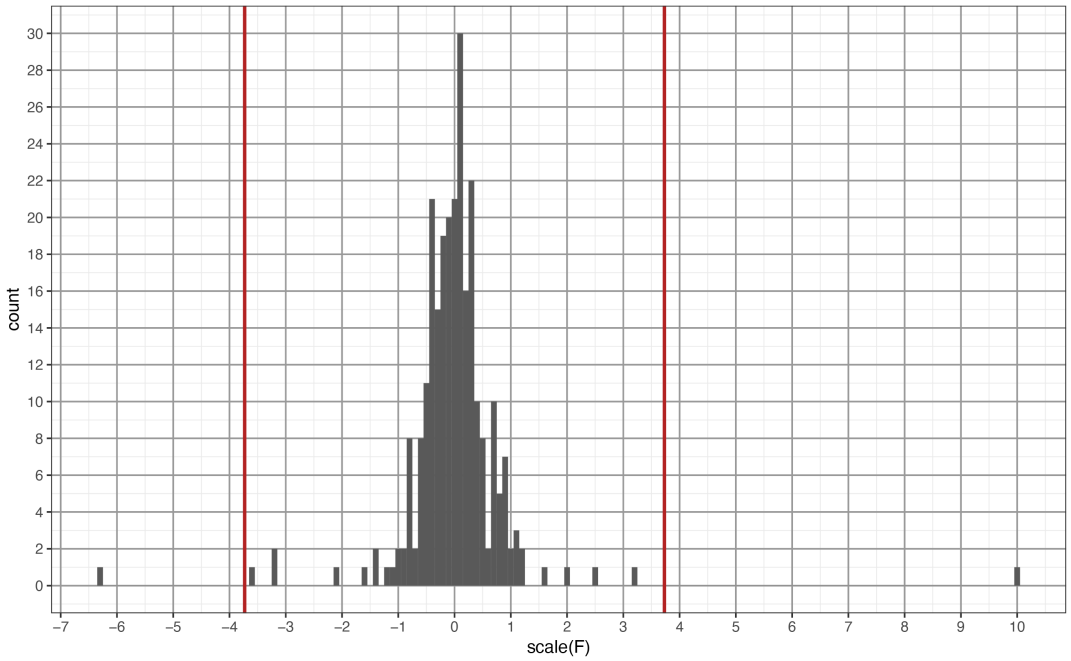

Figure 11: F coefficient estimates of heterozygosity for the MIMICS-In-House cohort

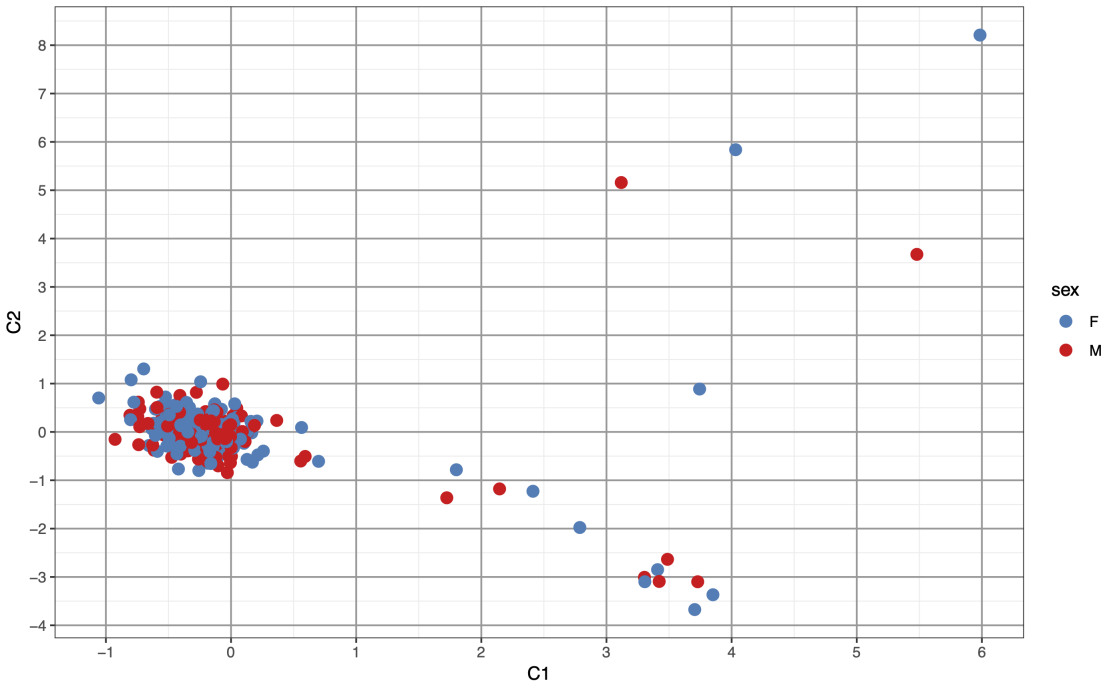

Figure 12: Multidimensional analysis of the identity by descent matrix from the MIMICS InHouse cohort, color corresponds to biological sex

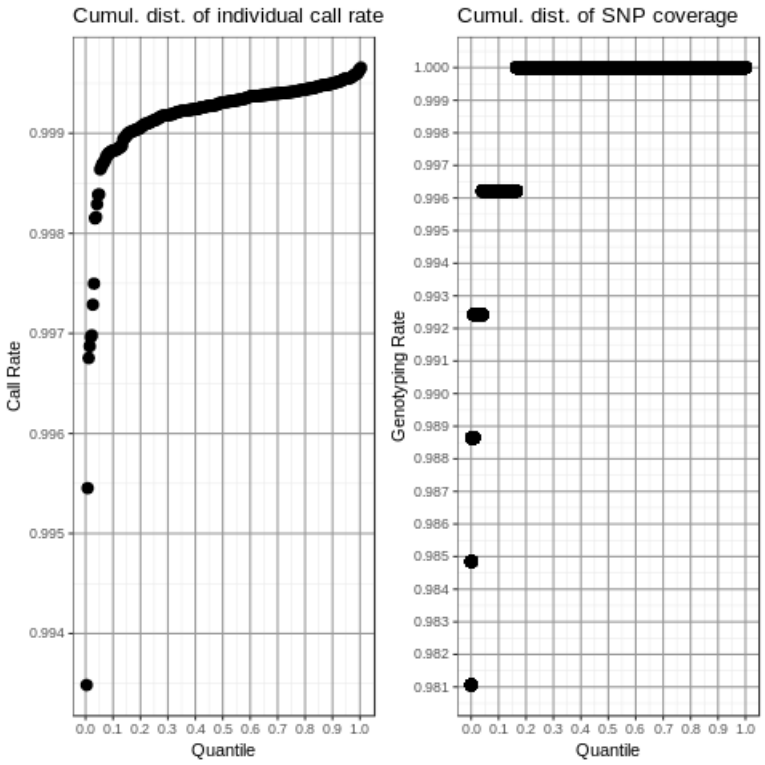

Figure 13: Missingness per sample and SNP post-imputation of the MIMICS-In-House cohort

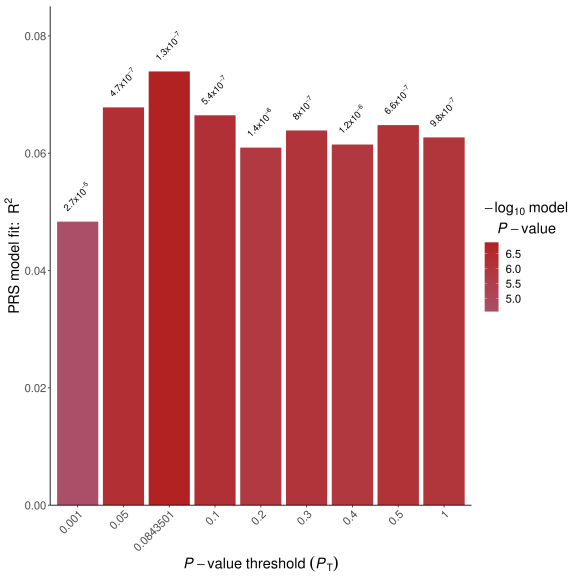

Figure 14: PRONIA PRS on Schizophrenia Nagelkerke's R2 model fit by different P value thresholds

under each bar, on top of each bar the P-value resulting from the association of the calculated PRS and schizophrenia is shown, the Y axis shows the Nagelkerke's R^2 which corresponds to the fit for each bar. The highest bar corresponds to the most predictive PRS, which is a function of the effect size distribution as well as the power of both base and target datasets. Figures 15 and 17 also show the goodness of fit for each model but at a higher resolution for which thousands of P-value cutoffs are tested and from which the highest bar for the other plots is calculated.

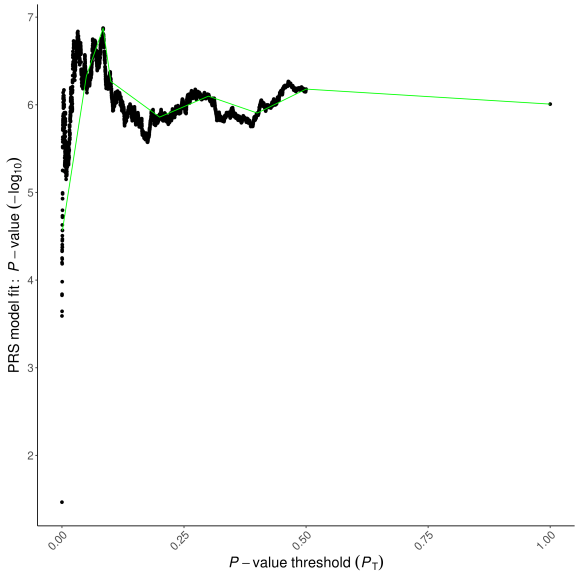

Figure 15: PRONIA PRS on Schizophrenia Nagelkerke's R2 model fit by different P value thresholds in high resolution

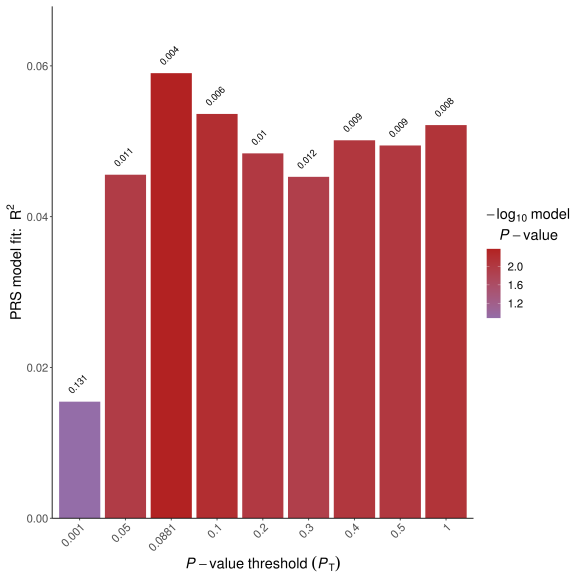

Figure 16: MIMICS-In-House PRS on Schizophrenia Nagelkerke's R2 model fit by different P value thresholds

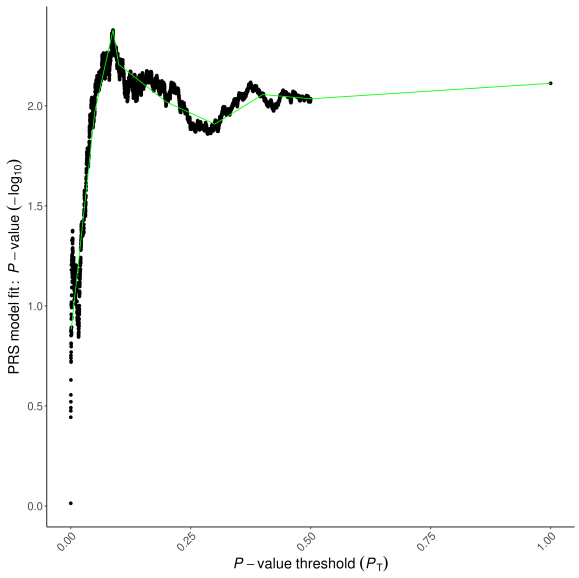

Figure 17: MIMICS-In-House PRS on schizophrenia Nagelkerke's R2 model fit by different P value thresholds in high resolution

7.3 MIMICS-In-House PRS on Schizophrenia HR

Samples from the MIMICS-InHouse cohort from which iPSCs were generated including bipolar disease cases, schizophrenia cases, and controls were assessed for copy number variations known to be associated with schizophrenia, figure 18 shows all the analyzed copy number variants and the samples for which the analysis was carried out, sample names starting with BD correspond to bipolar disorder diagnosed subject, those starting with CLZ or SCZ correspond to schizophrenia diagnosis and the rest are control samples.10 out of 49 samples had one or more overlapping copy number variants with those that have been associated with schizophrenia, 9 of them are heterozygous deletions, one of the assessed samples diagnosed with a bipolar disorder (BD6) has a heterozygous duplication resulting in 3 copies of a known causal copy number variation.

7.4 Schizophrenia associated alleles content assesment

After obtaining all the representative SNPs from all SCZ GWAS loci based on Ripke 2014 [50], the number of schizophrenia-associated alleles for these variants was assessed for each subject in the MIMICS-In-house combined cohort, with the goal of investigating whether there was a clear separation of schizophrenia-associated alleles

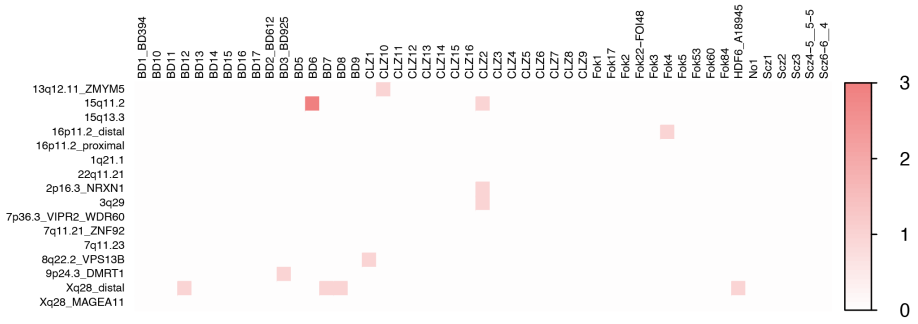

Figure 18: Schizophrenia–associated copy number variant analysis for the MIMICS-In-House samples from which iPSCs were derived

in cases versus controls. Figure 19 shows the overlap of the selected SNPs indicated by a red line on top of each crhomosome's cytoband, with the schizophrenia-GWAS SNP association. Figure 20 shows the results of this analysis on the first 4 chromosomes, with one heat map corresponding to each chromosome. The color scale corresponds to the number of schizophrenia-associated alleles which range from 0 to 2, if the position is colored in white for a given sample this means that the variant was not genotyped nor imputed, and therefore the genotype is unknown. All samples were clustered by cases and controls, and no clear trend separating both groups can be observed for any of the chromosomes. Only 4 chromosomes are shown for brevity, all other plots corresponding to the rest of the genome are included in the supplementary figures section.

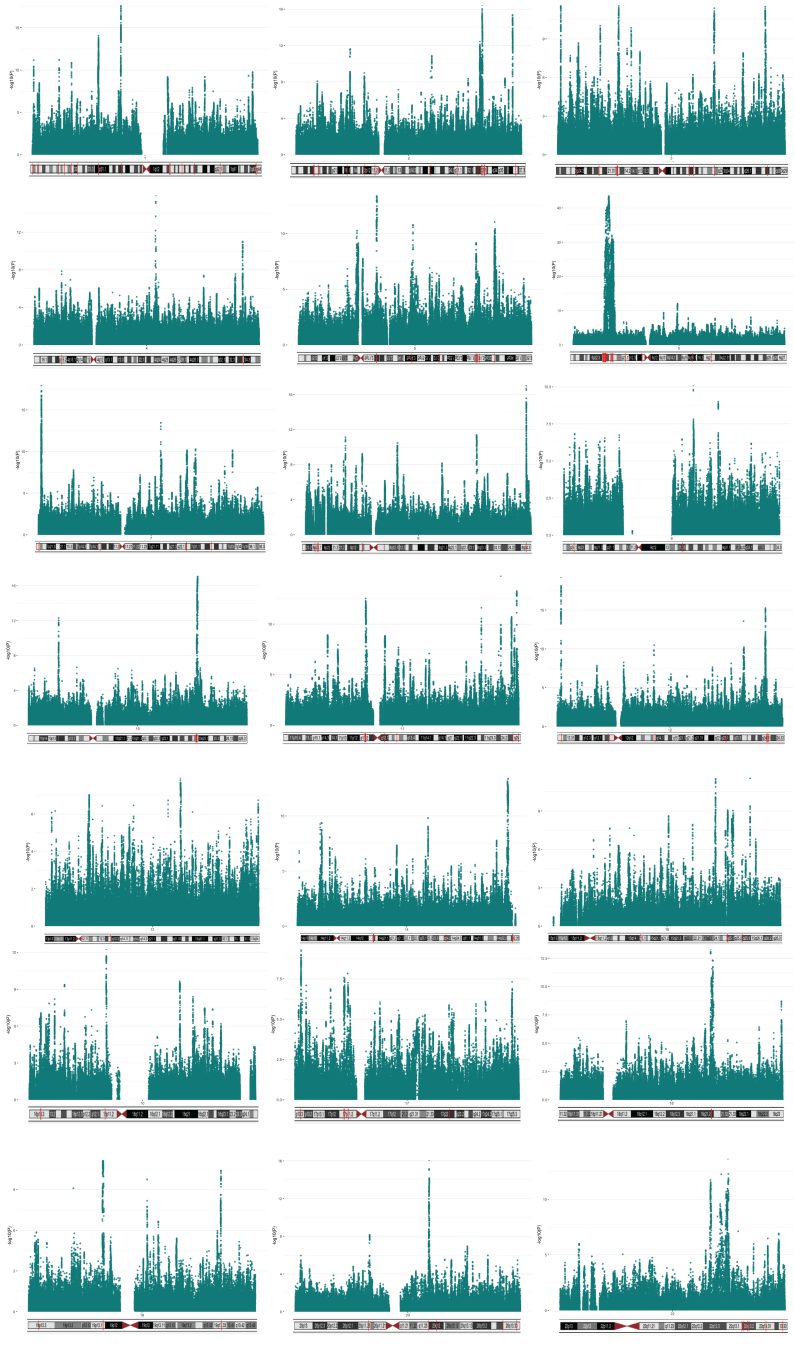

Figure 19: Schizphrenia GWAS by chromosome and the matching positions for the selected schizophrenia-associated SNPs

Figure 20: Schizophrenia–associated SNPs in chr1−chr4. The number of schizophrenia−associated alleles for each of the selected SNPs for which their SNP IDs are displayed on the X-axis of the plot, each row of the heatmap represents a sample. Each square represents the number of schizophrenia associated alleles for a specific SNP in a given sample, it ranges from 0 2 and a missing genotype is represented by a white colored square

7.5 Human post-mortem brain tissue characterization and heritability assessment

7.5.1 Single-nuclei RNAseq of human post-mortem brain tissue

Single-nuclei RNAseq was used to get the transcriptomic profiles of human post-mortem brain tissue taken from the prefrontal cortex, the resulting transcriptomes underwent dimensional reduction by Uniform Manifold Approximation and Projection (UMAP) Figure 21a shows the different clusters that represent different cell types, each cluster depicted by a different color, every dot represents a single cell. Identified cell types include Astrocytes (Astro), Cortical projection neurons, Corticothalamic Projection Neurons (CthPN), Inhibitory neurons, and oligodendrocytes.

7.5.2 Single cell ATACseq of human post mortem brain tissue

Single-cell ATACseq was also performed on post-mortem brain tissue from the prefrontal cortex, the resulting chromatin profiles from the sequenced cells are represented in different clusters resulting from the uniform manifold approximation and Projection dimensional reduction and are shown in Figure 21b, each dot corresponds to a cell and each cell's identity is determined by the cluster to which it belongs, clusters are separated by colors which correspond to cell types identified as Astrocytes, cortical sub cerebral projection neurons (ScPN), Inhibitory somatostatin expressing neurons (Inh−Sst), Oligodendrocytes (Oligo), Cortical Projection Neurons (CPN), Corticothalamic Projection Neurons (CthPN), Inhibitory Parvalbumin-positive neurons (Inh−Pvalb), Inhibitory-vasoactive intestinal polypeptide and calbindin2 expressing neurons (Inh−Vip−Calb2), Inhibitory−CALB2 and neuropeptide Y expressing neurons and dorsolateral cortical projection neurons (CPN−DL).

7.5.3 stratified LD-score regression for heritability enrichment assessment of post-mortem brain tissue

For a given polygenic trait, enrichment of a category or transcriptomic profile of a cell type is defined as the proportion of the SNP heritability in the profile divided by the proportion of SNPs in that profile. Enrichment is shown in Figure 22 a and b indicated by a positive (red) regression coefficient represented as Z-scores which are shown in the scale on the heatmaps, for every significant hit (after P value

correction) there is a star in the middle of the circle which indicates the magnitude of the enrichment (Z-score).

The heritability enrichment of these transcriptomic profiles was assessed for 19 different polygenic traits divided into 4 classifications. For psychiatric disorders, I analyzed enrichment in schizophrenia, bipolar disorder, autism spectrum disorders, and major depressive disorder (MDD), corresponding to Central Nervous System (CNS) traits I included the polygenic traits of antidepresant treatment resistance, any psychotic experience, clozapine levels, norclozapine levels, clozapine to norclozapine ratio, years of education, and neuroticism. Two different groups of controls were assessed, one control group corresponded to polygenic diseases which included Coronary Artery Disease (CAD), celiac disease, Type I and type II diabetes, for the second group I included polygenic traits including height, fasting glucose levels, if ever smoked, and low-density lipoprotein (LDL) cholesterol levels.

Figure 22 a shows the heritability enrichment results corresponding to 10 different transcriptomic profiles, corresponding to the single-nuclei RNAseq profiles of the post-mortem brain tissue samples. The resulting heritability enrichment analysis on snRNAseq data for the 19 included traits and diseases resulted in a significant enrichment in corticothalamic Layer 6 cells for schizophrenia, bipolar disorder and neuroticism, excitatory neurons of the cortical layers 2-3 were enriched for schizophrenia, bipolar disorder, neuroticism, and ever smoked. Excitatory neurons of the cortical layers 2-3 showed significant enrichment for schizophrenia and bipolar disorder, Parvalbumin expressing neurons (PVALB) also had a positive enrichment for schizophrenia, bipolar disorder and neuroticism, somatostatin-expressing neurons only showed significant enrichment for MDD and Neuroticism.

Figure 22b shows the heritability enrichment results corresponding to 15 different chromatin profiles that resulted from performing single-cell ATACseq on cells obtained from post-mortem brain tissue from the prefrontal cortex. Similar to the results obtained from snRNAseq data, Corticothalamic Layer 6 cells, excitatory neurons of the cortical layers 2-3, excitatory neurons of the cortical layers 2-3 and Parvalbumin expressing neurons showed a significant enrichment in schizophrenia and bipolar disorder heritabilities. However, additional cell types were also found to have a significant enrichment for schizophrenia, these include vasoactive intestinal polypeptide expressing neurons (VIP), somatostatin expressing neurons (SST), Oligodendrocyte progenitor cells (OPC), Microglia, Astrocytes and Oligodendro-

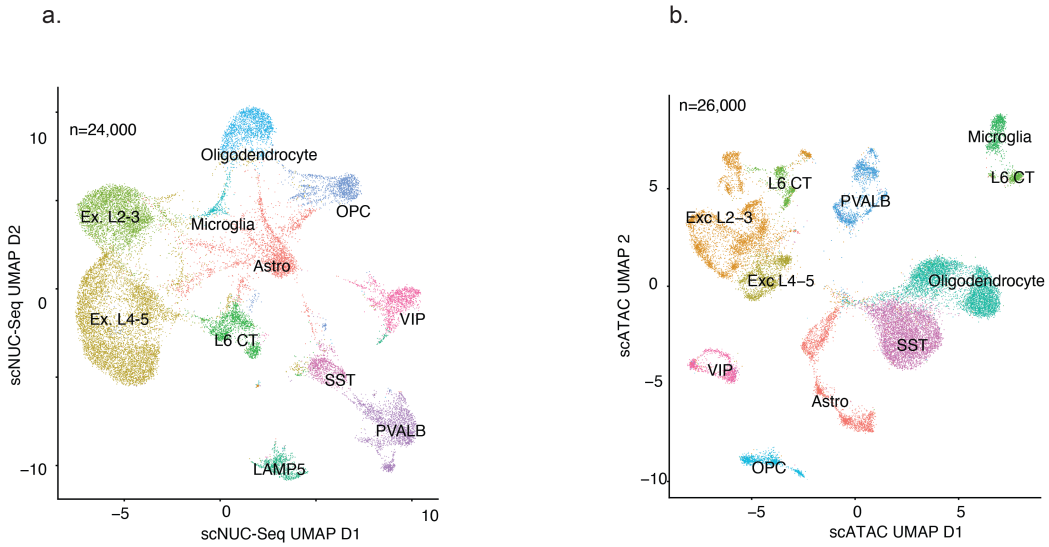

Figure 21: Uniform Manifold Approximation and Projection for a. snRNAseq and b. scATACseq on PFC post-mortem cells

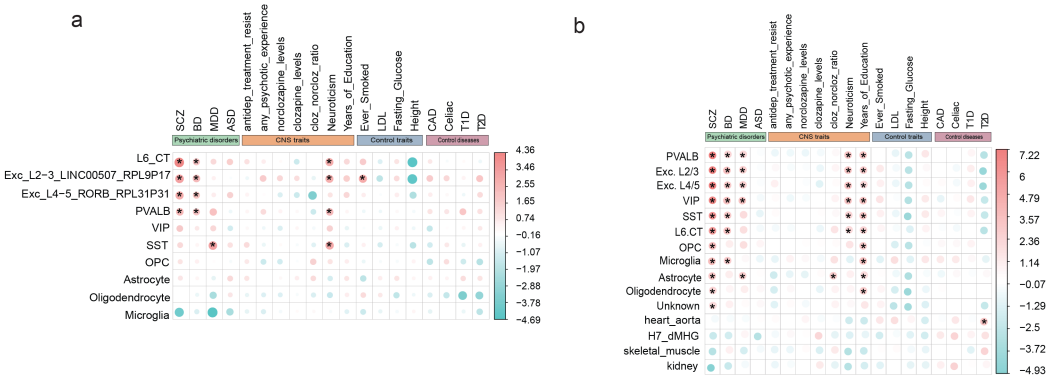

Figure 22: LD-score regression results for a. snRNAseq and b. scATACseq of cortical cells from the prefrontal cortex of post-mortem brain tissue

cytes. A cluster with unknown specific cell identity but also isolated from the prefrontal cortex showed varied enrichment profiles that only included traits and diseases from the psychiatric diseases and CNS traits groups. Remarkably, all cell clusters derived from the PFC post-mortem data showed significant enrichment for schizophrenia and years of education. None of the cells showed enrichment for any of the control traits and diseases.

Chromatin accessibility profiles corresponding to heart, kidney, skeletal muscle and the H7 dMHG cell line were used as controls, as expected none of them showed significant enrichment for any of the psychiatric diseases or CNS traits.

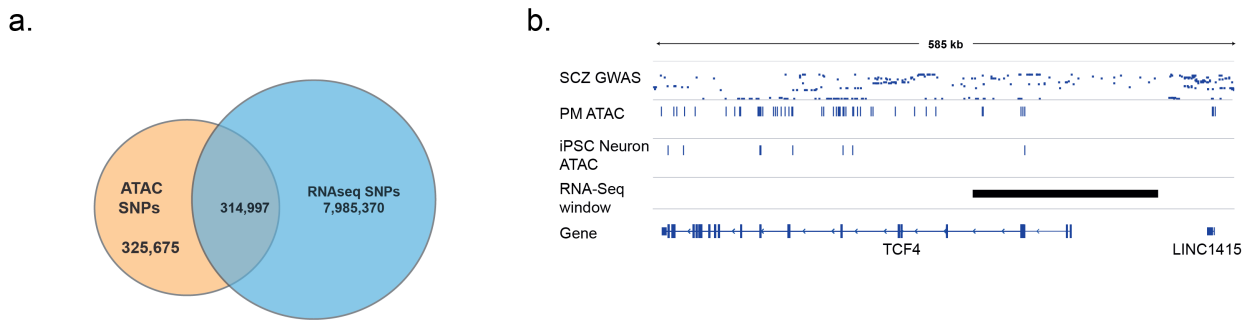

Figure 23: a. Venn diagram representing the number of SNPs covered by ATACseq and RNAseq of the post-mortem brain tissue analysed cells. b. shows the overlap between different genomic functional features like schizophrenia GWAS SNPs and ATACseq and RNAseq loci in different cells.

7.5.4 snRNAseq vs scATACseq

Figure 23 shows a comparison between RNAseq and ATACseq interrogated loci. A total of 325,675 SNPs overlapped the pool of all analyzed loci in the scATACseq experiments for all cells, The loci corresponding to the snRNAseq experiments, covered a total of 7,985,370 SNPs. Both datasets shared 314,997 SNPs (Figure 23a). The SNPs dataset used for the overlap was that of the schizophrenia GWAS from Pardiñas et al 2018. [40]. Figure 23b shows how different genomic features overlap. TCF4 gene, which has been repeatedly found to harbor schizophrenia risk loci [5] is used as an example to show where the different functional categories of scATACseq and snRNAseq as well as the schizophrenia-associated SNPs overlap with the gene.

7.6 Model Systems to capture the polygenic basis of mental illness

With the aim of finding a suitable model system to further study psychiatric disorders, I performed the stratified LD–score regression analysis on scATACseq data from organoids, mouse cortical neurons across different developmental stages, and patient-derived iPSCs.

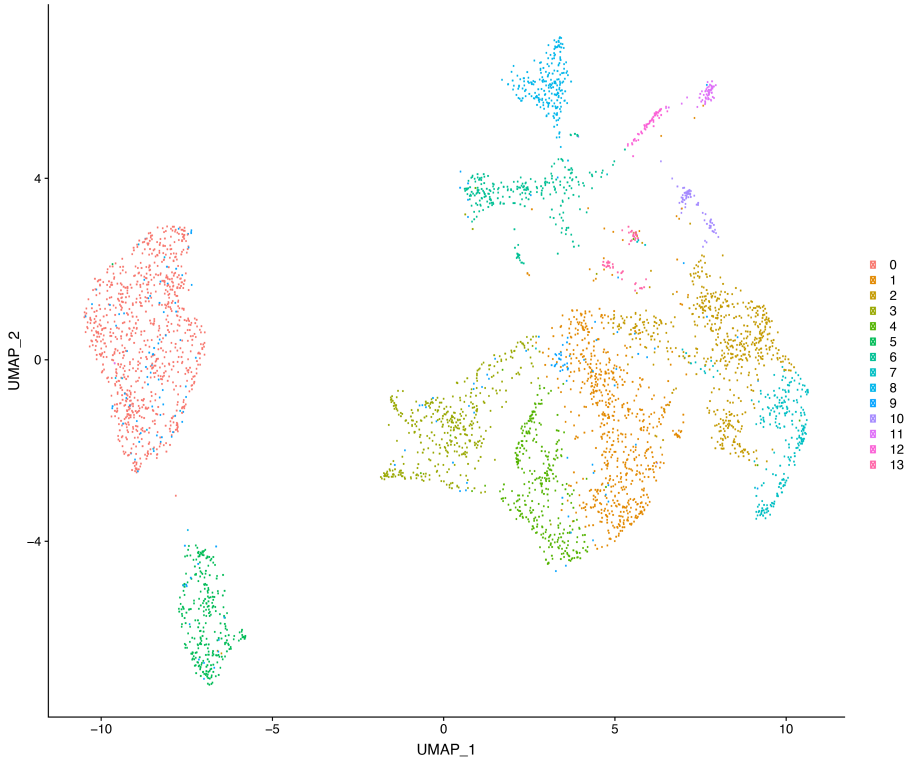

Figure 24: Uniform Manifold Approximation and Projection for Organoid's scATACseq data.

7.6.1 Single-cell ATACseq from cerebral organoids

Figure 24 shows the dimensional reduction by Uniform Manifold Approximation and Projection of the single-cell ATACseq data of cerebral organoids resulting in 13 different clusters corresponding to different neuronal maturation stages. Figure 25 shows the resulting heritability enrichment for the same 13 diseases and traits analyzed for the post-mortem brain samples. Neurons differentiated from patient-derived iPSCs are also shown in this heat map. Schizophrenia heritability was enriched in 4 neuronal cell types and 3 different organoids, bipolar disease had a similar result, with 4 neuronal cell types, and 4 different organoids which had enrichment for bipolar disorder heritability. Only three neuronal cell types and one cerebral organoid were enriched for autism spectrum disorder. MDD was also enriched in three neuronal cell types and four cerebral organoids, control CNS traits of intelligence, and educational attainment were enriched in all neurons and several cerebral organoids. Only two organoids were significantly enriched for the polygenic control trait of height, none of the organoids or cells were enriched for any control disease.

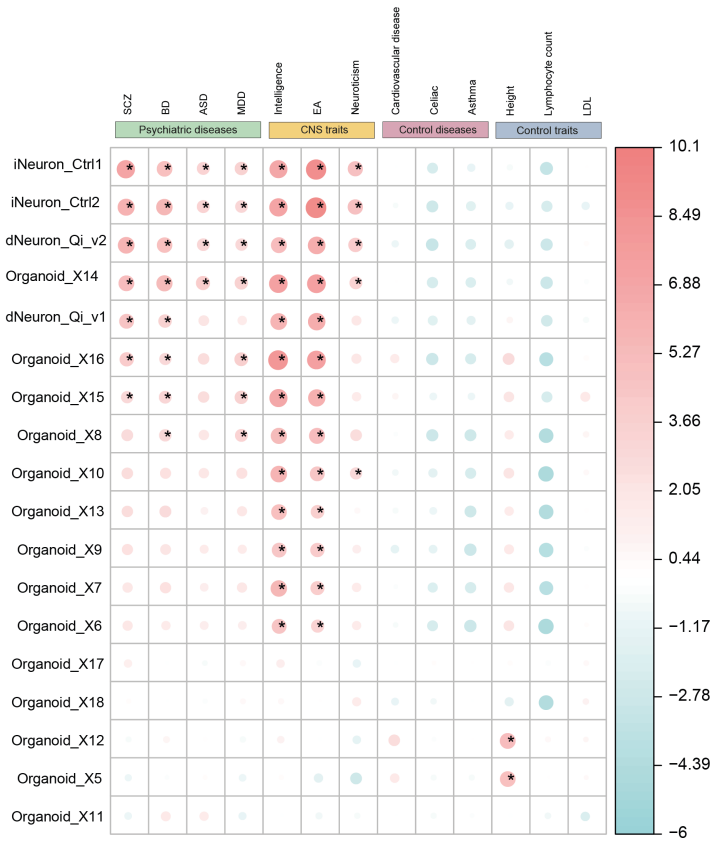

Figure 25: LD-score regression results different neuronal types and organoids

7.6.2 Single-cell ATACseq from different brain cell types differentiated from patient-derived iPSCs

A second dataset comprising different brain cell types differentiated from patient-derived iPSCs was also analyzed using stratified LD-score regression to assess them for heritability enrichment of various polygenic diseases and traits. Shown in Figure 26 are the heritability enrichment results for the chromatin profiles of the differentiated cells, in this analysis I also included the ATACseq chromatin profiles from heart, liver, hESC, CD19, and H7 dAPS cells.

Schizophrenia heritability enrichment was found in Parvalbumin-positive neurons, somatostatin-expressing neurons, excitatory neurons from the cortical layers 2-3, vasoactive intestinal polypeptide-expressing neurons, excitatory neurons from the cortical layers 4-5, neocortical layer 6 neurons, patient-derived GABAergic neurons (iGABA), patient-derived neurons (iNeurons), oligodendrocyte progenitor cells, distal neural progenitor cells, and ERG positive neurons. Bipolar disorder showed significant enrichment in excitatory neurons from the cortical layers 2-3, excitatory neurons from the cortical layers 4-5, and neocortical layer 6 neurons. MDD heritability was found to be enriched in Parvalbumin-positive neurons, vasoactive intestinal polypeptide-expressing neurons, excitatory neurons from the cortical layers 4-5, patient-derived GABAergic neurons, and distal neural progenitor cells. Parvalbumin-positive neurons, somatostatin-expressing neurons, excitatory neurons from the cortical layers 2-3, vasoactive intestinal polypeptide-expressing neurons, patient-derived neurons (iNeurons), oligodendrocyte progenitor cells, and excitatory neurons from the cortical layers 4-5 had significant enrichment for both neuroticism and educational attainment (Years of education). neocortical layer 6 neurons,

7.6.3 Single-cell ATACseq from murine prefrontal cortex cells at different developmental stages.

scATACseq was performed on mouse cortical neurons from postnatal days 1, 7, and 21 (P1, P7 & P21). Figure 27a shows the heritability enrichment results for 15 different polygenic traits and diseases. Schizophrenia heritability enrichment was found in Astrocytes and layer IV neurons at P1, in sub-cerebral projection neurons, neuropeptide Y-expressing neurons, cortical projection neurons, corticothalamic projection neurons and astrocytes at P7, for P21 enrichment was found in

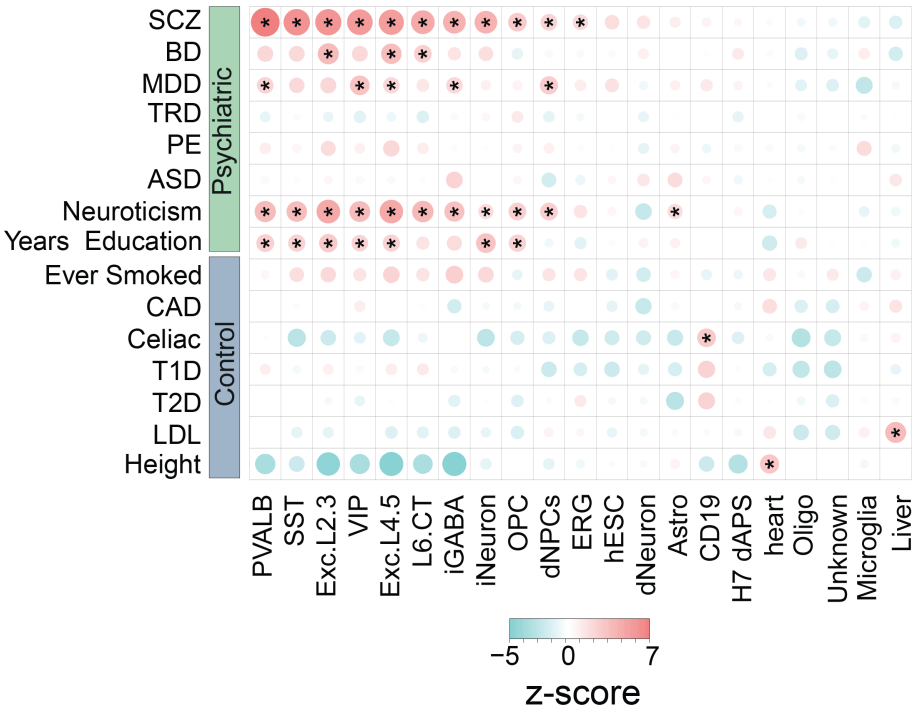

Figure 26: LD-score regression results differentiated cortical cells from patient-derived iPSCs

neuropeptide Y-expressing neurons, parvalbumin-positive neurons, somatostatin-expressing neurons, cortical projection neurons, NFOL neurons, sub-cerebral projection neurons, corticothalamic projection neurons, layer IV neurons and Astrocytes. MDD heritability enrichment was only significant for corticothalamic projection neurons at P7. Figure 27b summarizes the enrichment heritability for schizophrenia for all cell types in the three developmental stages.

7.6.4 Single-cell RNAseq from human and mouse callosal neurons of the developing neocortex

Zuccaro et al [73] generated a longitudinal transcriptional map of excitatory projection neuron (PN) and inhibitory interneuron (IN) subtypes of the cerebral cortex, across a timeline of mouse embryonic and postnatal development, as well as fetal human cortex and human cortical organoids. To evaluate the association of different neuronal signatures with neuropsychiatric disease risk loci, I performed stratified LD-score regression on their chromatin profiles obtained by single-cell ATACseq.

Figures 28a and 28b show no enrichment for any of the characterized cortical neurons from mouse, nor the embryonic or post−natal neurons nor the interneurons

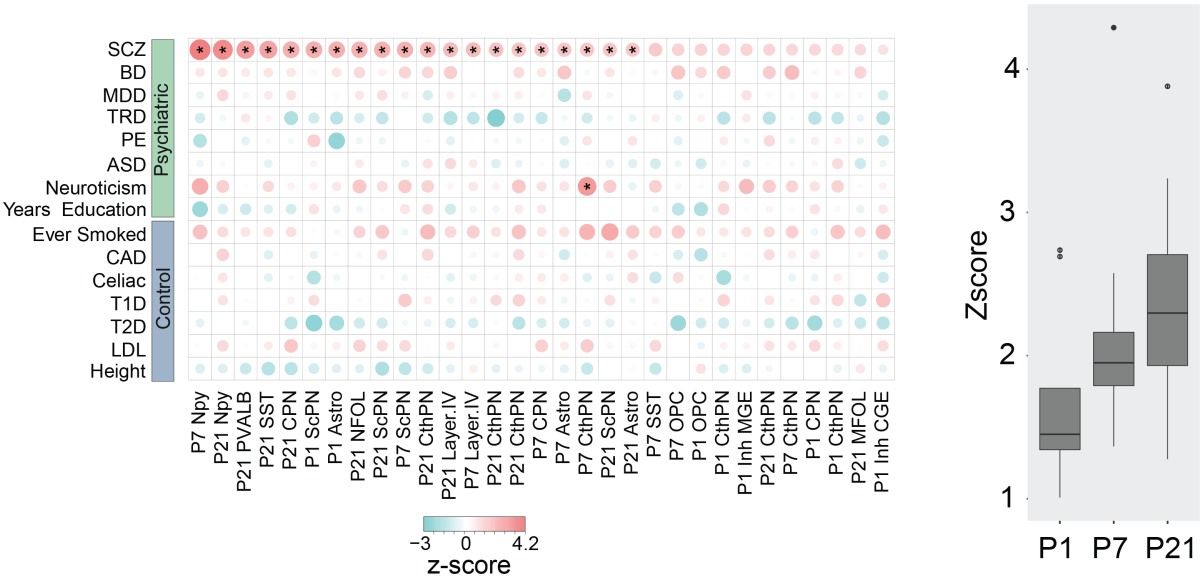

Figure 27: LD-score regression results from single-cell ATACseq from mouse neurons on different developmental stages

at post-natal day 30. For the cortical neuron signature genes on fetal tissue, cortical projection neurons showed a significant heritability enrichment for schizophrenia (Figure 28c). Similarly, cells identified as cortical projection neurons in human cortical organoids also showed significant enrichment for schizophrenia, at 3 and 6 months corticothalamic projection neurons showed significant enrichment for schizophrenia heritability. Cerebral organoids at 3 months also showed significant enrichment for schizophrenia and bipolar disorder heritability in interneuron markers positive cells and in sub−cerebral projection neurons (Figure 28d).

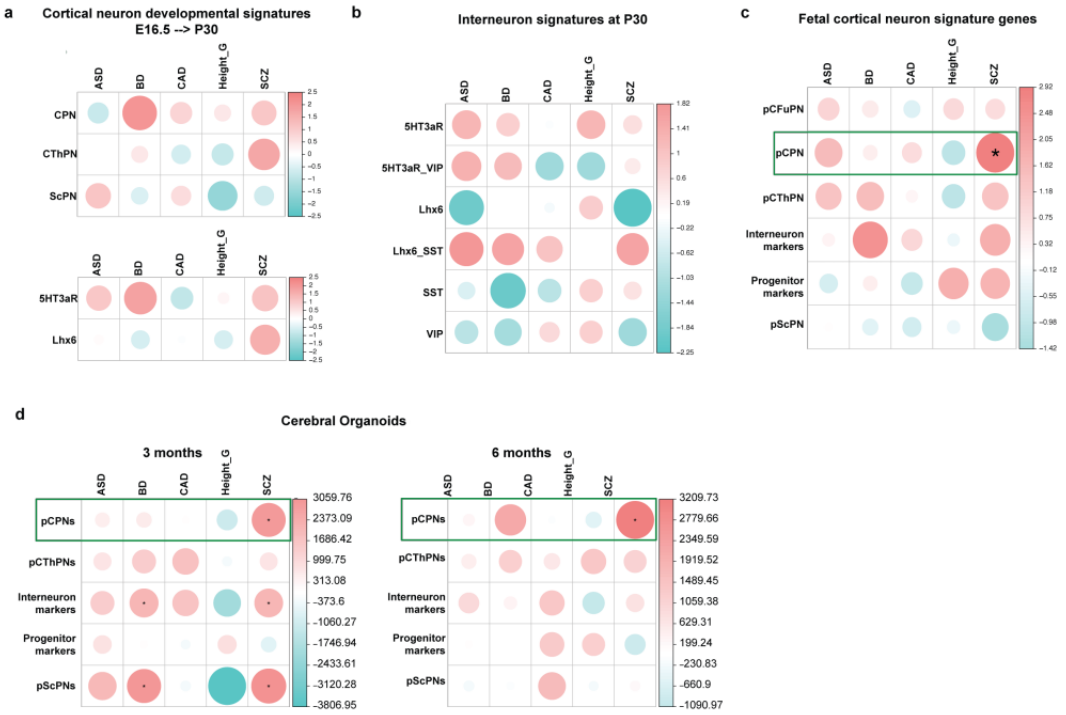

Figure 28: Cortical Neurons and fetal developmental LD-score regression results

7.6.5 Partitioned heritability analysis of scATACseq chromatin profiles from iNeurons after a sustained depolarization stimulus

High potassium chloride (hKCl) depolarizing treatment was added to iNeurons and the resulting chromatin accessibility profile was obtained with ATAC−Seq. This profile was then analyzed using stratified LD−score regression to find out whether the chromatin accessibility changes induced by hKCL treatment were significantly enriched for psychiatric disorders heritability. Figure 29 shows the comparison of

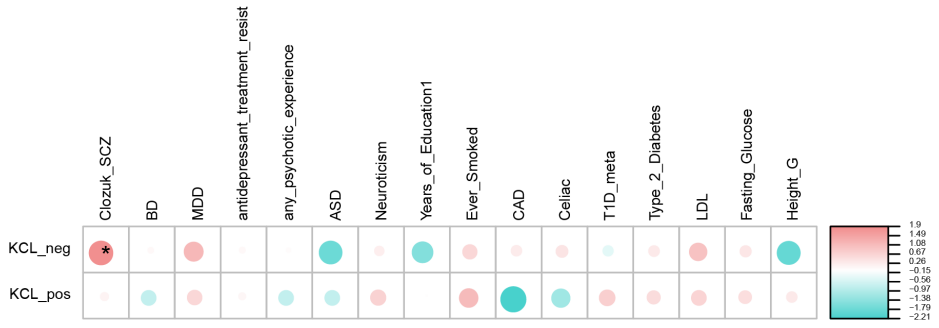

Figure 29: LD-score regression for high potassium chloride depolarizing treatment in iNeurons

heritability enrichment of the chromatin accessibility profile of iNeurons with and without treatment. Cells without treatment resulted in a positive enrichment for schizophrenia whereas the ones with hKCl treatment showed no enrichment for any of the analyzed polygenic traits.

7.6.6 Summary

In summary, these results show highlight the power of genomic imputation to a better characterization of case controls cohorts, at the same time by seeing how heterogeneous the distribution of disease associated allele content can be, it supports the need of developing better polygenic risk score calculation methods to better utilize results from genome wide association studies.

The heritability enrichment results for all the analyzed post-mortem brain tissue cells as well as the ones coming from all tested model systems supports previous evidence that shows the involvement of various brain cell types in the development of schizophrenia. In parallel, this adds evidence for the benefit of using chromatin accessibility profiles when analyzing cell-type specific signatures relevant for disease risk. Moreover, this further demonstrates the validity of iPSCs as in vitro model systems that not only can capture the polygenic architecture of psychiatric disorders but can also be used as a tool to perform high throughput functional experiments, accelerating the research that will ultimately lead us to better diagnosis and treatments.

8 DISCUSSION

The results presented in this thesis highlight the relevance and validity of induced pluripotent stem cells as an in vitro model system to study psychiatric disorders like schizophrenia. Neuronal cell types differentiated from patient-derived iPSCs showed to capture the polygenic basis of psychiatric disorders in a way other model systems cannot; iPSCs have the advantage of being able to carry the genomic liability of subjects diagnosed with a psychiatric disorder, making it possible to study the effect of specific genomic profiles with varying disease-associated variants, taking advantage of modern functional genomics techniques as single-cell RNA sequencing and single-cell ATAC sequencing, which can be performed in a high throughput fashion. This opens up the possibility of studying the cell type-specific mechanisms through which psychiatric diseases act which will be fundamental for the identification of drugable targets as well as the development of new therapies, furthermore understanding which associated variants deferentially affect specific cell types is key to developing better methods for early identification of subjects at risk of developing psychiatric disorders.

8.1 Genome-wide studies in psychiatric disorders

Genotyping imputation plays a crucial role in genetic association studies by enhancing statistical power and assisting in the interpretation of findings [11]. In this thesis, I performed the genomic imputation of the PRONIA and MIMICs-In-House cohorts, which were composed of healthy controls and subjects diagnosed with psychiatric disorders like schizophrenia and bipolar disorder or with a diagnosis of recent-onset psychosis. Leveraging the power of reference panels like the 1000 genomes project which provides the haplotypes for subjects of different ethnicities from around the world [4], I successfully and accurately inferred the genotypes at untyped markers for two cohorts, thereby expanding the coverage of known genomic SNPs for all samples. Figure 30 shows the population structure of the 10 populations used as a reference panel to impute the described cohorts, as comprehensive as this panel looks, not all subjects can be imputed with the same accuracy, there are many ethnicities that are not included in reference panels and which harbor allele frequencies specific to them, therefore some genomic regions will be harder to impute. For the same reason disease-associated variants resulting from genome-

wide association studies carried out in a specific population should ideally be further studied in individuals with a similar genetic background [34].

The influence of genetic risk factors on psychiatric diseases has been addressed by a number of studies over the past several years [40][66][50] Many genome-wide association studies and meta-analyses have been carried out finding hundreds of loci implicated in the liability for psychiatric diseases, making it clear these are polygenic and complex [40][50]. Nevertheless, valuable insights have been gained which have influenced the direction of functional studies at the same time they have improved the power of polygenic risk scores, useful to evaluate shared etiology between different psychiatric disorders, stratify patients with the same diagnosis but with different underlying genetic makeup and as a predictive tool to identify subjects at risk [10].

The genomic imputation of both cohorts allowed me to calculate better and more informative schizophrenia polygenic risk scores for all samples. However polygenic risk scores for psychiatric diseases are still far from being a reliable tool in a clinical setting. Given that the polygenic risk of psychiatric diseases like schizophrenia is distributed across many genes and genomic regulatory elements, all of which correspond to different cell types, considering all of them at once might not be as informative as partitioning the risk according to the portion of the genome that is relevant for a cell-type. The results from the Schizophrenia-associated alleles content assessment, reported in section 7.4 (Figures 19 and 20) show the heterogeneity of schizophrenia-associated alleles across the MIMICS-In-House cohort. Figure 20 reports the schizophrenia-associated number of alleles as defined by the GWAS performed by [50] samples are grouped by controls on the top half of the heatmap and cases in the lower half, there is no clear trend that separates both groups by schizophrenia associated allele content, showing how simply counting the number of associated variants without assigning them a weight or effect size is not informative about disease status. The practical application of better polygenic risk scores can be valuable in subjects where there is a greater preexisting likelihood of disease, such as during the initial phases of illnesses, aiding in diagnostic processes or guiding treatment decisions [34].

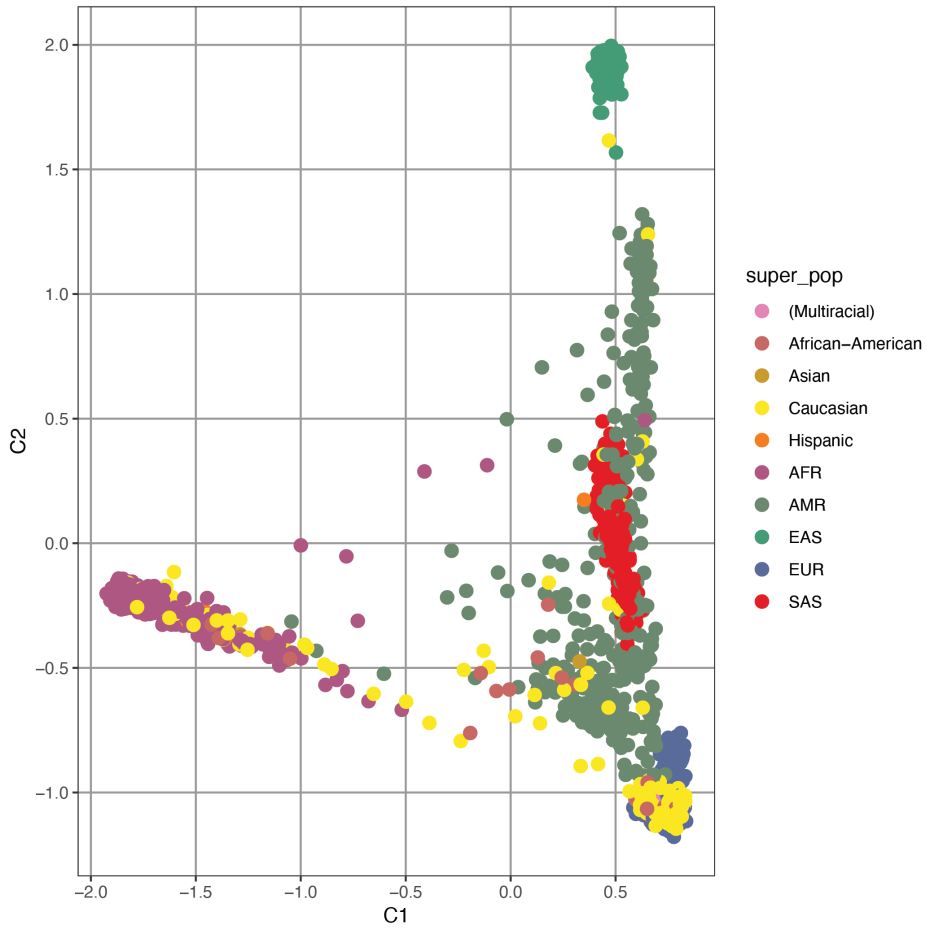

Figure 30: Principal component analysis of all samples in the 1000 genomes project. The superopulations included are African-American, Asian, Caucassian, Hispanic, SAS = Southasians, EUR = Europeans, EAS = East Asians, AMR = Admixed americans, and AFR = Africans.

8.2 Neuronal cell subtypes vulnerable to schizophrenia liability

As discussed in the previous section, disease-associated variants are likely to act in a highly condition and cell type specific fashion, therefore identifying the cells that are vulnerable to the to psychiatric disease associated polygenic risk is crucial. In order to do so, I performed partitioned heritability analysis on single nuceli RNA sequencing and single cell ATAC sequencing data from human post-mortem brain tissue of the prefrontal cortex. I analyzed the functional profiles resulting from both sequencing techniques using the stratified LD-score regression method by Finucane et al. 2015 [17]. This resulted in a list of cell types that were significantly enriched for psychiatric disorders.

In the case of schizophrenia, LD-scroe regression on single nuclei RNAseq data was able to detect a significant positive enrichment for schizophrenia heritability in corticothalamic layer 6 cells, neurons of the cortical layers 2−3, neurons of the cortical layers 4−5, and Parvalbumin expressing neurons (PVALB)(Fiigure 22a).

Significant enrichment heritability analysis also performed on adult post-mortem brain tissue but using single-cell ATACseq also detected positive significant enrichment in corticothalamic layer 6 cells, neurons of the cortical layers 2−3, neurons of the cortical layers 4−5, and Parvalbumin expressing neurons (PVALB). Additionally significant enrichment for schizophrenia heritability was also found in vasoactive intestinal polypeptide expressing neurons (VIP), somatostatin expressing neurons (SST), Oligodendrocyte progenitor cells (OPC), microglia, astrocytes and oligodendrocytes.

For bipolar disorder, the stratified LD-score analysis on snRNAseq data was able to detect significant enrichment in corticothalamic layer 6 cells, neurons of the cortical layers 2−3, neurons of the cortical layers 4−5, and Parvalbumin expressing neurons (PVALB). These same cells were also enriched according to scATACseq data, similar to schizophrenia, scATACseq was able to also identify more cell enriched for heritability, these inlcued vasoactive intestinal polypeptide expressing neurons (VIP), somatostatin expressing neurons (SST) and Microglia.

Many of these cells were also enriched for heritability for other traits or diseases related to the central nervous system, supporting the notion that there is a shared

genetic liability among psychiatric disorders. All of this aligns with previous research where these cell types had been found to be relevant to susceptibility to variants associated with the above traits [45].

8.3 RNAseq vs ATACseq

Single-cell RNAseq and ATACseq have been proven useful to understand which genomic regulatory elements and which genes are active in time point for a specific cell, in this thesis I aimed to test both and find out which technique was best suited to study the how the genetic risk of schizophrenia and bipolar disorder affect specific cell types. In the post-mortem brain tissue studies, the partitioned heritability analysys of single-cell ATACseq data was able to identify a bigger proportion of cell types that harbor heritability enrichment, this findings are consistent with previous findings of many schizophrenia risk loci being located in non−coding regions of the genome [40]. Figure 23a shows the number of SNPs used to calculate heritability enrichment with the ATACseq and RNAseq genomic tracks of all analysed cells pooled together, the Venn diagram illustrates how scATACseq covers a lot less SNPs than snRNAseq, which actually includes most of the SNPs also covered by scATACseq, indicating that maybe ATACseq data converges higher overlap specificity with the schizophrenia risk loci and therefore being able to yield better results when calculating heritability enrichment. However, looking at the combined profiles generated by both techniques could also shed light on the regulatory mechanisms through which gene transcription is governed.

8.4 Model systems for studying psychiatric disorders

Post-mortem brain tissue studies have yielded valuable biological insights from many psychiatric disorders, however schizophrenia is a neuro-develppmental disorder that already affects brain cells in early stages of life [22], post-mortem brain tissue is not always available from young individuals, adult post-mortem brain tissue inevitably misses important insights of the mechanisms that act earlier in life.

With the goal of finding suitable in vitro model systems that were able to capture the polygenic basis of psychiatric disorders observed in post-mortem brain tissue cells, I explored the heritability enrichment of different brain cell types from various sources.

Differentiated neuronal cells from iPSCs showed a significant enrichment for schizophrenia and bipolar disorder like their ex-vivo counterparts, confirming their validity as a great in vitro model system to study psychiatric disorders. Furthermore, to investigate whether activity-dependent chromatin accessibility profiles had an impact in the enrichment for polygenic risk associated with diseases, a comparison study was carried out by adding hKCL to iPSCs derived neurons. The resulting partitioned heritability analysis on chromatin accessibility profiles in the neurons before and after depolarization by high potassium chloride (hKCL) identified significant enrichment of polygenic risk for schizophrenia within the set of open chromatin regions that displayed increased accessibility following stimulation, emphasizing the importance of considering the functional implications of disease-associated genetic variations not only under normal conditions but also highlight the role of activity-dependent processes in contributing to subtle modifications in neuronal plasticity properties associated with schizophrenia.

These results show how induced pluripotent stem cells grant access to functioning human brain cells, enabling the exploration of molecular mechanisms of disease and variations among individuals. Their compatibility with diverse technical methods like the single-cell RNA and ATAC sequencing used in this thesis, make them an exceptional in vitro model system in many respects[12]. Differentiated brain cells derived from iPSCs are especially valuable for investigating the genetic risk of psychiatric diseases, as well as the effects of pharmacological interventions, on fundamental processes taking place within and between neural cells.

iPSCs offer distinct advantages in comprehending essential cellular processes and basic network activity. However, they may be less suitable for studying complete brain systems and, importantly, how intricate brain functions evolve throughout the dynamic landscape of neurodevelopment [12] To overcome these limiations, cerebral organoids have recently gained popularity as a higher organization human neural system that allows for the study of more complex circuitry and cell interactions.

In this thesis, the resulting heritability enrichment from two independent organoid scATACseq experiemtns are reported. One of these results sets corrresponds to the organoid scRNAseq data by Zuccaro et al. [73] where enrichment for schizophrenia was found on cells identified as Corticothalamic projection neurons and sub-cerebral projection neurons. The second set of results, corresponding to the those reported in Section 4.6.1 show schizophrenia heritability enrichment for 3 different cell clusters,

unfortunately both studies cannot be compared because in the second study no cell identity could be assigned, proving that when working with organoids, assigning cell identity by ATACseq profiles can be very challenging; the presence of disordered heterogeneity and low reproducibility continue to pose a notable struggle, emphasizing the importance of meticulous experimental design aimed at minimizing their influence. [12] As the protocols for the generation of organoids improve, I expect it to become the next best suitable in vitro model system to study psychiatric disorders.

Lastly, partitioned heritability was assessed in cortical cells from mouse from E16.5 to P30, no heritability enrichment was found for any of the obtained cell types at any of the time points. In the same study by Zuccaro et al. [73] transcriptomic profiles from human fetal cortical neurons was obtained and also analysed using LD-score regression method, identifying significant heritability enrcihment for cortical projection neurons, suggesting that signature transcriptomic profiles are highly cell type specific and might also be species specific. However analysing the scATACseq data from mouse cortical neurons at different post natal time points from a second study by Yuan et al [71] revealed significant heritability for schizophrenia in astrocytes and layer IV neurons at P1, in sub-cerebral projection neurons, neuropeptide Y-expressing neurons, cortical projection neurons, corticotha- lamic projection neurons and astrocytes at P7, for P21 enrichment was found in neuropeptide Y-expressing neurons, parvalbumin-positive neurons, somatostatin expressing neurons, cortical projection neurons, NFOL neurons, subcerebral projec- tion neurons, corticothalamic projection neurons, layer IV neurons and astrocytes. These differences could arise from the functional approach used to determine the cell type specific signatures, while Zuccaro et al. [73] used RNAseq, Yuan et al. [71] used ATACseq, which as observed earlier in the analysis of the postmortem data, seems more suitable for assessing the heritability enrichment of psychiatric disorders.

8.5 Future directions

These findings could potentially boost the improvement of methods for calculating polygenic risk scores to maximize the utilization of genome-wide association study results by modifying the weights of the SNPs contributing to the score according to what is relevant for a specific cell type, currently most methods just perfom a sum of the effect sizes of all selected SNPs.

The heritability enrichment results obtained from the analysis of post-mortem

brain tissue cells and model systems support existing evidence regarding the involvement of various brain cell types in schizophrenia and bipolar disease development.

Additionally, this research reinforces the value of using chromatin accessibility profiles for analyzing cell-type specific signatures relevant to disease risk. Furthermore, it further validates the utility of induced pluripotent stem cells as in vitro model systems that not only capture the polygenic architecture of psychiatric disorders but also serve as a tool for conducting high-throughput functional experiments. These advancements accelerate research efforts aimed at enhancing diagnosis and treatment strategies in the field.

9 PUBLICATION STATEMENT

Part of the work presented in this PhD thesis is part of a co-authored manuscript titled “Cell type and condition specific functional annotation of schizophrenia associated non coding genetic variants” which has been submitted to a peer-reviewed journal.

A second part of this thesis was also part of a manuscript titled ”Human-specific enrichment of schizophrenia risk-genes in callosal neurons of the developing neocortex” which has been submitted to BioRxiv.

A ELECTRONIC APPENDIX

Code and figures are provided in electronic form:

https://www.dropbox.com/sh/n401qsjx4gqiyv9/AAAKiJjEuv1BAKq3F9MfgBwwa?dl=0

References

[1] Abascal, F., Acosta, R., Addleman, N. J., Adrian, J., Afzal, V., Aken, B., Akiyama, J. A., Jammal, O. A., Amrhein, H., Anderson, S. M., Andrews, G. R., Antoshechkin, I., Ardlie, K. G., Armstrong, J., Astley, M., Banerjee, B., Barkal, A. A., Barnes, I. H., Barozzi, I., Barrell, D., Barson, G., Bates, D., Baymuradov, U. K., Bazile, C., Beer, M. A., Beik, S., Bender, M. A., Bennett, R., Bouvrette, L. P. B., Bernstein, B. E., Berry, A., Bhaskar, A., Bignell, A., Blue, S. M., Bodine, D. M., Boix, C., Boley, N., Borrman, T., Borsari, B., Boyle, A. P., Brandsmeier, L. A., Breschi, A., Bresnick, E. H., Brooks, J. A., Buckley, M., Burge, C. B., Byron, R., Cahill, E., Cai, L., Cao, L., Carty, M., Castanon, R. G., Castillo, A., Chaib, H., Chan, E. T., Chee, D. R., Chee, S., Chen, H., Chen, H., Chen, J. Y., Chen, S., Cherry, J. M., Chhetri, S. B., Choudhary, J. S., Chrast, J., Chung, D., Clarke, D., Cody, N. A., Coppola, C. J., Coursen, J., D'Ippolito, A. M., Dalton, S., Danyko, C., Davidson, C., Davila-Velderrain, J., Davis, C. A., Dekker, J., Deran, A., DeSalvo, G., Despacio-Reyes, G., Dewey, C. N., Dickel, D. E., Diegel, M., Diekhans, M., Dileep, V., Ding, B., Djebali, S., Dobin, A., Dominguez, D., Donaldson, S., Drenkow, J., Dreszer, T. R., Drier, Y., Duff, M. O., Dunn, D., Eastman, C., Ecker, J. R., Edwards, M. D., El-Ali, N., Elhajjajy, S. I., Elkins, K., Emili, A., Epstein, C. B., Evans, R. C., Ezkurdia, I., Fan, K., Farnham, P. J., Farrell, N. P., Feingold, E. A., Ferreira, A. M., Fisher-Aylor, K., Fitzgerald, S., Flicek, P., Foo, C. S., Fortier, K., Frankish, A., Freese, P., Fu, S., Fu, X. D., Fu, Y., Fukuda-Yuzawa, Y., Fulciniti, M., Funnell, A. P., Gabdank, I., Galeev, T., Gao, M., Giron, C. G., Garvin, T. H., Gelboin-Burkhart, C. A., Georgolopoulos, G., Gerstein, M. B., Giardine, B. M., Gifford, D. K., Gilbert, D. M., Gilchrist, D. A., Gillespie, S., Gingeras, T. R., Gong, P., Gonzalez, A., Gonzalez, J. M., Good, P., Goren, A., Gorkin, D. U., Graveley, B. R., Gray, M., Greenblatt, J. F., Griffiths, E., Groudine, M. T., Grubert, F., Gu, M., Guigó, R., Guo, H., Guo, Y., Guo, Y., Gursoy, G., Gutierrez-Arcelus, M., Halow, J., Hardison, R. C., Hardy, M., Hariharan, M., Harmanci, A., Harrington, A., Harrow, J. L., Hashimoto, T. B., Hasz, R. D., Hatan, M., Haugen, E., Hayes, J. E., He, P., He, Y., Heidari, N., Hendrickson, D., Heuston, E. F., Hilton, J. A., Hitz, B. C., Hochman, A., Holgren, C., Hou, L., Hou, S., Hsiao, Y. H. E., Hsu, S., Huang, H., Hubbard, T. J., Huey, J., Hughes, T. R., Hunt, T., Ibarrientos, S., Issner, R., Iwata, M., Izuogu, O., Jaakkola, T., Jameel, N., Jansen, C., Jiang, L., Jiang, P., Johnson, A., Johnson, R., Jungreis, I., Kadaba, M., Kasowski, M., Kasparian, M., Kato, M., Kaul,

R., Kawli, T., Kay, M., Keen, J. C., Keles, S., Keller, C. A., Kelley, D., Kellis, M., Kheradpour, P., Kim, D. S., Kirilusha, A., Klein, R. J., Knoechel, B., Kuan, S., Kulik, M. J., Kumar, S., Kundaje, A., Kutyavin, T., Lagarde, J., Lajoie, B. R., Lambert, N. J., Lazar, J., Lee, A. Y., Lee, D., Lee, E., Lee, J. W., Lee, K., Leslie, C. S., Levy, S., Li, B., Li, H., Li, N., Li, X., Li, Y. I., Li, Y., Li, Y., Li, Y., Lian, J., Libbrecht, M. W., Lin, S., Lin, Y., Liu, D., Liu, J., Liu, P., Liu, T., Liu, X. S., Liu, Y., Liu, Y., Long, M., Lou, S., Loveland, J., Lu, A., Lu, Y., Lécuyer, E., Ma, L., Mackiewicz, M., Mannion, B. J., Mannstadt, M., Manthravadi, D., Marinov, G. K., Martin, F. J., Mattei, E., McCue, K., McEown, M., McVicker, G., Meadows, S. K., Meissner, A., Mendenhall, E. M., Messer, C. L., Meuleman, W., Meyer, C., Miller, S., Milton, M. G., Mishra, T., Moore, D. E., Moore, H. M., Moore, J. E., Moore, S. H., Moran, J., Mortazavi, A., Mudge, J. M., Munshi, N., Murad, R., Myers, R. M., Nandakumar, V., Nandi, P., Narasimha, A. M., Narayanan, A. K., Naughton, H., Navarro, F. C., Navas, P., Nazarovs, J., Nelson, J., Neph, S., Neri, F. J., Nery, J. R., Nesmith, A. R., Newberry, J. S., Newberry, K. M., Ngo, V., Nguyen, R., Nguyen, T. B., Nguyen, T., Nishida, A., Noble, W. S., Novak, C. S., Novoa, E. M., Nuñez, B., O'Donnell, C. W., Olson, S., Onate, K. C., Otterman, E., Ozadam, H., Pagan, M., Palden, T., Pan, X., Park, Y., Partridge, E. C., Paten, B., Pauli-Behn, F., Pazin, M. J., Pei, B., Pennacchio, L. A., Perez, A. R., Perry, E. H., Pervouchine, D. D., Phalke, N. N., Pham, Q., Phanstiel, D. H., Plajzer-Frick, I., Pratt, G. A., Pratt, H. E., Preissl, S., Pritchard, J. K., Pritykin, Y., Purcaro, M. J., Qin, Q., Quinones-Valdez, G., Rabano, I., Radovani, E., Raj, A., Rajagopal, N., Ram, O., Ramirez, L., Ramirez, R. N., Rausch, D., Raychaudhuri, S., Raymond, J., Razavi, R., Reddy, T. E., Reimonn, T. M., Ren, B., Reymond, A., Reynolds, A., Rhie, S. K., Rinn, J., Rivera, M., Rivera-Mulia, J. C., Roberts, B. S., Rodriguez, J. M., Rozowsky, J., Ryan, R., Rynes, E., Salins, D. N., Sandstrom, R., Sasaki, T., Sathe, S., Savic, D., Scavelli, A., Scheiman, J., Schlaffner, C., Schloss, J. A., Schmitges, F. W., See, L. H., Sethi, A., Setty, M., Shafer, A., Shan, S., Sharon, E., Shen, Q., Shen, Y., Sherwood, R. I., Shi, M., Shin, S., Shoresh, N., Siebenthall, K., Sisu, C., Slifer, T., Sloan, C. A., Smith, A., Snetkova, V., Snyder, M. P., Spacek, D. V., Srinivasan, S., Srivas, R., Stamatoyannopoulos, G., Stamatoyannopoulos, J. A., Stanton, R., Steffan, D., Stehling-Sun, S., Strattan, J. S., Su, A., Sundararaman, B., Suner, M. M., Syed, T., Szynkarek, M., Tanaka, F. Y., Tenen, D., Teng, M., Thomas, J. A., Toffey, D., Tress, M. L., Trout, D. E., Trynka, G., Tsuji, J., Upchurch, S. A., Ursu, O., Uszczynska-Ratajczak, B., Uziel, M. C., Valencia, A., Biber,

B. V., van der Velde, A. G., Van Nostrand, E. L., Vaydylevich, Y., Vazquez, J., Victorsen, A., Vielmetter, J., Vierstra, J., Visel, A., Vlasova, A., Vockley, C. M., Volpi, S., Vong, S., Wang, H., Wang, M., Wang, Q., Wang, R., Wang, T., Wang, W., Wang, X., Wang, Y., Watson, N. K., Wei, X., Wei, Z., Weisser, H., Weissman, S. M., Welch, R., Welikson, R. E., Weng, Z., Westra, H. J., Whitaker, J. W., White, C., White, K. P., Wildberg, A., Williams, B. A., Wine, D., Witt, H. N., Wold, B., Wolf, M., Wright, J., Xiao, R., Xiao, X., Xu, J., Yan, K. K., Yan, Y., Yang, H., Yang, X., Yang, Y. W., Yardimci, G. G., Yee, B. A., Yeo, G. W., Young, T., Yu, T., Yue, F., Zaleski, C., Zang, C., Zeng, H., Zeng, W., Zerbino, D. R., Zhai, J., Zhan, L., Zhan, Y., Zhang, B., Zhang, J., Zhang, J., Zhang, K., Zhang, L., Zhang, P., Zhang, Q., Zhang, X. O., Zhang, Y., Zhang, Z., Zhao, Y., Zheng, Y., Zhong, G., Zhou, X. Q., Zhu, Y. and Zimmerman, J. [2020]. Expanded encyclopaedias of DNA elements in the human and mouse genomes, *Nature 2020 583:7818* **583**(7818): 699–710.

[2] *A map of human genome variation from population-scale sequencing* [2010]. *Nature 2010 467:7319* **467**(7319): 1061–1073.

[3] Andrey Ziyatdinov, Joshua Backman, J. M. S. M. G. T. J. Y. Z. D. L. J. S. R. P. X. B. S. B. L. H. R. L. A. L. E. M. M. J. E. J. W. S. J. O. J. R. T. T. G. A. A. B. J. M. a. c. e. a. s. o. R. G. C. o. R. P. A. N. K. S. and are current employees and/or stockholders of AstraZeneca Mark Reppell is current employee and/or stockholder of AbbVie., S. P. [2022]. Genotyping, sequencing and analysis of 140,000 adults from the Mexico City Prospective Study, *bioRxiv* p. 2022.06.26.495014.

[4] Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., Mardis, E. R., Marth, G. T., McVean, G. A., Nickerson, D. A., Schmidt, J. P., Sherry, S. T., Wang, J., Wilson, R. K., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., Kovar, C., Lee, S., Muzny, D., Reid, J. G., Zhu, Y., Chang, Y., Feng, Q., Fang, X., Guo, X., Jian, M., Jiang, H., Jin, X., Lan, T., Li, G., Li, J., Li, Y., Liu, S., Liu, X., Lu, Y., Ma, X., Tang, M., Wang, B., Wang, G., Wu, H., Wu, R., Xu, X., Yin, Y., Zhang, D., Zhang, W., Zhao, J., Zhao, M., Zheng, X., Gupta, N., Gharani, N., Toji, L. H., Gerry, N. P., Resch, A. M., Barker, J., Clarke, L., Gil, L., Hunt, S. E., Kelman,

G., Kulesha, E., Leinonen, R., McLaren, W. M., Radhakrishnan, R., Roa, A., Smirnov, D., Smith, R. E., Streeter, I., Thormann, A., Toneva, I., Vaughan, B., Zheng-Bradley, X., Grocock, R., Humphray, S., James, T., Kingsbury, Z., Sudbrak, R., Albrecht, M. W., Amstislavskiy, V. S., Borodina, T. A., Lienhard, M., Mertes, F., Sultan, M., Timmermann, B., Yaspo, M. L., Fulton, L., Ananiev, V., Belaia, Z., Beloslyudtsev, D., Bouk, N., Chen, C., Church, D., Cohen, R., Cook, C., Garner, J., Hefferon, T., Kimelman, M., Liu, C., Lopez, J., Meric, P., O'Sullivan, C., Ostapchuk, Y., Phan, L., Ponomarov, S., Schneider, V., Shekhtman, E., Sirotkin, K., Slotta, D., Zhang, H., Balasubramaniam, S., Burton, J., Danecek, P., Keane, T. M., Kolb-Kokocinski, A., McCarthy, S., Stalker, J., Quail, M., Davies, C. J., Gollub, J., Webster, T., Wong, B., Zhan, Y., Campbell, C. L., Kong, Y., Marcketta, A., Yu, F., Antunes, L., Bainbridge, M., Sabo, A., Huang, Z., Coin, L. J., Fang, L., Li, Q., Li, Z., Lin, H., Liu, B., Luo, R., Shao, H., Xie, Y., Ye, C., Yu, C., Zhang, F., Zheng, H., Zhu, H., Alkan, C., Dal, E., Kahveci, F., Garrison, E. P., Kural, D., Lee, W. P., Leong, W. F., Stromberg, M., Ward, A. N., Wu, J., Zhang, M., Daly, M. J., DePristo, M. A., Handsaker, R. E., Banks, E., Bhatia, G., Del Angel, G., Genovese, G., Li, H., Kashin, S., McCarroll, S. A., Nemesh, J. C., Poplin, R. E., Yoon, S. C., Lihm, J., Makarov, V., Gottipati, S., Keinan, A., Rodriguez-Flores, J. L., Rausch, T., Fritz, M. H., Stütz, A. M., Beal, K., Datta, A., Herrero, J., Ritchie, G. R., Zerbino, D., Sabeti, P. C., Shlyakhter, I., Schaffner, S. F., Vitti, J., Cooper, D. N., Ball, E. V., Stenson, P. D., Barnes, B., Bauer, M., Cheetham, R. K., Cox, A., Eberle, M., Kahn, S., Murray, L., Peden, J., Shaw, R., Kenny, E. E., Batzer, M. A., Konkel, M. K., Walker, J. A., MacArthur, D. G., Lek, M., Herwig, R., Ding, L., Koboldt, D. C., Larson, D., Ye, K., Gravel, S., Swaroop, A., Chew, E., Lappalainen, T., Erlich, Y., Gymrek, M., Willems, T. F., Simpson, J. T., Shriver, M. D., Rosenfeld, J. A., Bustamante, C. D., Montgomery, S. B., De La Vega, F. M., Byrnes, J. K., Carroll, A. W., DeGorter, M. K., Lacroute, P., Maples, B. K., Martin, A. R., Moreno-Estrada, A., Shringarpure, S. S., Zakharia, F., Halperin, E., Baran, Y., Cerveira, E., Hwang, J., Malhotra, A., Plewczynski, D., Radew, K., Romanovitch, M., Zhang, C., Hyland, F. C., Craig, D. W., Christoforides, A., Homer, N., Izatt, T., Kurdoglu, A. A., Sinari, S. A., Squire, K., Xiao, C., Sebat, J., Antaki, D., Gujral, M., Noor, A., Ye, K., Burchard, E. G., Hernandez, R. D., Gignoux, C. R., Haussler, D., Katzman, S. J., Kent, W. J., Howie, B., Ruiz-Linares, A., Dermitzakis, E. T., Devine, S. E., Kang, H. M., Kidd, J. M., Blackwell, T., Caron, S., Chen, W., Emery, S., Fritsche, L., Fuchsberger, C., Jun, G., Li, B., Lyons, R., Scheller, C., Sidore,

C., Song, S., Sliwerska, E., Taliun, D., Tan, A., Welch, R., Wing, M. K., Zhan, X., Awadalla, P., Hodgkinson, A., Li, Y., Shi, X., Quitadamo, A., Lunter, G., Marchini, J. L., Myers, S., Churchhouse, C., Delaneau, O., Gupta-Hinch, A., Kretzschmar, W., Iqbal, Z., Mathieson, I., Menelaou, A., Rimmer, A., Xifara, D. K., Oleksyk, T. K., Fu, Y., Liu, X., Xiong, M., Jorde, L., Witherspoon, D., Xing, J., Browning, B. L., Browning, S. R., Hormozdiari, F., Sudmant, P. H., Khurana, E., Tyler-Smith, C., Albers, C. A., Ayub, Q., Chen, Y., Colonna, V., Jostins, L., Walter, K., Xue, Y., Gerstein, M. B., Abyzov, A., Balasubramanian, S., Chen, J., Clarke, D., Fu, Y., Harmanci, A. O., Jin, M., Lee, D., Liu, J., Mu, X. J., Zhang, J., Zhang, Y., Hartl, C., Shakir, K., Degenhardt, J., Meiers, S., Raeder, B., Casale, F. P., Stegle, O., Lameijer, E. W., Hall, I., Bafna, V., Michaelson, J., Gardner, E. J., Mills, R. E., Dayama, G., Chen, K., Fan, X., Chong, Z., Chen, T., Chaisson, M. J., Huddleston, J., Malig, M., Nelson, B. J., Parrish, N. F., Blackburne, B., Lindsay, S. J., Ning, Z., Zhang, Y., Lam, H., Sisu, C., Challis, D., Evani, U. S., Lu, J., Nagaswamy, U., Yu, J., Li, W., Habegger, L., Yu, H., Cunningham, F., Dunham, I., Lage, K., Jespersen, J. B., Horn, H., Kim, D., Desalle, R., Narechania, A., Sayres, M. A., Mendez, F. L., Poznik, G. D., Underhill, P. A., Mittelman, D., Banerjee, R., Cerezo, M., Fitzgerald, T. W., Louzada, S., Massaia, A., Yang, F., Kalra, D., Hale, W., Dan, X., Barnes, K. C., Beiswanger, C., Cai, H., Cao, H., Henn, B., Jones, D., Kaye, J. S., Kent, A., Kerasidou, A., Mathias, R., Ossorio, P. N., Parker, M., Rotimi, C. N., Royal, C. D., Sandoval, K., Su, Y., Tian, Z., Tishkoff, S., Via, M., Wang, Y., Yang, H., Yang, L., Zhu, J., Bodmer, W., Bedoya, G., Cai, Z., Gao, Y., Chu, J., Peltonen, L., Garcia-Montero, A., Orfao, A., Dutil, J., Martinez-Cruzado, J. C., Mathias, R. A., Hennis, A., Watson, H., McKenzie, C., Qadri, F., LaRocque, R., Deng, X., Asogun, D., Folarin, O., Happi, C., Omoniwa, O., Stremlau, M., Tariyal, R., Jallow, M., Joof, F. S., Corrah, T., Rockett, K., Kwiatkowski, D., Kooner, J., Hien, T. T., Dunstan, S. J., Thuy-Hang, N., Fonnie, R., Garry, R., Kanneh, L., Moses, L., Schieffelin, J., Grant, D. S., Gallo, C., Poletti, G., Saleheen, D., Rasheed, A., Brooks, L. D., Felsenfeld, A. L., McEwen, J. E., Vaydylevich, Y., Duncanson, A., Dunn, M. and Schloss, J. A. [2015]. A global reference for human genetic variation, *Nature 2015 526:7571* **526**(7571): 68–74.

[5] Badowska, D. M., Brzózka, M. M., Kannaiyan, N., Thomas, C., Dibaj, P., Chowdhury, A., Steffens, H., Turck, C. W., Falkai, P., Schmitt, A., Papiol, S., Scheuss, V., Willig, K. I., Martins-de Souza, D., Rhee, J. S., Malzahn, D.

and Rossner, M. J. [2020]. Modulation of cognition and neuronal plasticity in gain- and loss-of-function mouse models of the schizophrenia risk gene Tcf4, *Translational Psychiatry 2020 10:1* **10**(1): 1–12.
URL: *https://www.nature.com/articles/s41398-020-01026-7*

[6] Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R., Farnham, P. J., Hirst, M., Lander, E. S., Mikkelsen, T. S. and Thomson, J. A. [2010]. The NIH Roadmap Epigenomics Mapping Consortium, *Nature biotechnology* **28**(10): 1045–1048.

[7] Brennand, K. J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., Li, Y., Mu, Y., Chen, G., Yu, D., McCarthy, S., Sebat, J. and Gage, F. H. [2011]. Modelling schizophrenia using human induced pluripotent stem cells, *Nature 2011 473:7346* **473**(7346): 221–225.

[8] Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. [2013]. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position, *Nature methods* **10**(12): 1213–1218.

[9] Cannon, T. D., Kaprio, J., Lönnqvist, J., Huttunen, M. and Koskenvuo, M. [1998]. The genetic epidemiology of schizophrenia in a Finnish twin cohort. A population-based modeling study, *Archives of general psychiatry* **55**(1): 67–74.

[10] Choi, S. W., Mak, T. S. H. and O'Reilly, P. F. [2020]. Tutorial: a guide to performing polygenic risk score analyses, *Nature protocols* **15**(9): 2759–2772.
URL: *https://pubmed.ncbi.nlm.nih.gov/32709988/*

[11] Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P. R., Iacono, W. G., Swaroop, A., Scott, L. J., Cucca, F., Kronenberg, F., Boehnke, M., Abecasis, G. R. and Fuchsberger, C. [2016]. Next-generation genotype imputation service and methods, *Nature Genetics 2016 48:10* **48**(10): 1284–1287.

[12] De, A., Angeles, L., Fernando, M. B., Hall, N. A. L., Brennand, K. J., Harrison, P. J., Maher, B. J., Weinberger, D. R. and Tunbridge, E. M. [2021]. Review

Induced Pluripotent Stem Cells in Psychiatry: An Overview and Critical Perspective, *Biological Psychiatry* **90**: 362–372.

[13] de Jonge, J. C., Vinkers, C. H., Hulshoff Pol, H. E. and Marsman, A. [2017]. GABAergic mechanisms in schizophrenia: Linking postmortem and In vivo studies, *Frontiers in Psychiatry* **8**(AUG): 118.

[14] Dong, X., Liu, C. and Dozmorov, M. [2021]. Review of multi-omics data resources and integrative analysis for human brain disorders, *Briefings in Functional Genomics* **20**(4): 223–234.

[15] Euesden, J., Lewis, C. M. and O'Reilly, P. F. [2015]. PRSice: Polygenic Risk Score software, *Bioinformatics (Oxford, England)* **31**(9): 1466–1468.
URL: *https://pubmed.ncbi.nlm.nih.gov/25550326/*

[16] Fiers, M. W., Minnoye, L., Aibar, S., González-Blas, C. B., Atak, Z. K. and Aerts, S. [2018]. Mapping gene regulatory networks from single-cell omics data, *Briefings in Functional Genomics* **17**(4): 246–254.

[17] Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P. R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F. R., Purcell, S., Stahl, E., Lindstrom, S., Perry, J. R., Okada, Y., Raychaudhuri, S., Daly, M. J., Patterson, N., Neale, B. M. and Price, A. L. [2015]. Partitioning heritability by functional annotation using genome-wide association summary statistics, *Nature Genetics* **47**(11): 1228–1235.

[18] Finucane, H. K., Reshef, Y. A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P. R., Lareau, C., Shoresh, N., Genovese, G., Saunders, A., Macosko, E., Pollack, S., Perry, J. R., Buenrostro, J. D., Bernstein, B. E., Raychaudhuri, S., McCarroll, S., Neale, B. M. and Price, A. L. [2018]. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types, *Nature Genetics* **50**(4): 621–629.

[19] *Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders* [2019]. *Cell* **179**(7): 1469–1482.e11.

[20] Hayes, B. [2013]. Overview of Statistical Methods for Genome-Wide Association Studies (GWAS), *Methods in molecular biology (Clifton, N.J.)* **1019**: 149–169.

[21] Hilker, R., Helenius, D., Fagerlund, B., Skytthe, A., Christensen, K., Werge, T. M., Nordentoft, M. and Glenthøj, B. [2018]. Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register, *Biological psychiatry* **83**(6): 492–498.

[22] Hoffmann, A., Ziller, M. and Spengler, D. [2019]. Progress in iPSC-Based Modeling of Psychiatric Disorders, *International Journal of Molecular Sciences 2019, Vol. 20, Page 4896* **20**(19): 4896.

[23] Howie, B. N., Donnelly, P. and Marchini, J. [2009]. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies, *PLOS Genetics* **5**(6): e1000529.

[24] Hwang, Y., Kim, J., Shin, J. Y., Kim, J. I., Seo, J. S., Webster, M. J., Lee, D. and Kim, S. [2013]. Gene expression profiling by mRNA sequencing reveals increased expression of immune/inflammation-related genes in the hippocampus of individuals with schizophrenia, *Translational Psychiatry* **3**(10): e321.
URL: */pmc/articles/PMC3818014/ /pmc/articles/PMC3818014/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3818014/*

[25] Illumina [2023]. Infinium Global Screening Array-24 Kit, `https://www.illumina.com/products/by-type/microarray-kits/infinium-global-screening.html`. [Online; accessed 8-June-2023].

[26] Kahn, R. S., Sommer, I. E., Murray, R. M., Meyer-Lindenberg, A., Weinberger, D. R., Cannon, T. D., O'Donovan, M., Correll, C. U., Kane, J. M., Van Os, J. and Insel, T. R. [2015]. Schizophrenia, *Nature Reviews Disease Primers 2015 1:1* **1**(1): 1–23.

[27] Kavanagh, D. H., Dwyer, S., O'Donovan, M. C. and Owen, M. J. [2013]. The ENCODE project: implications for psychiatric genetics, *Molecular Psychiatry 2013 18:5* **18**(5): 540–542.

[28] Kohen, R., Dobra, A., Tracy, J. H. and Haugen, E. [2014]. Transcriptome profiling of human hippocampus dentate gyrus granule cells in mental illness, *Translational psychiatry* **4**(3).
URL: *https://pubmed.ncbi.nlm.nih.gov/24594777/*

[29] Kucharska-Mazur, J., Ratajczak, M. Z. and Samochowiec, J. [2019]. Stem Cells in Psychiatry, *Advances in experimental medicine and biology* **1201**: 159–174.

[30] *LD score regression distinguishes confounding from polygenicity in genome-wide association studies* [2015]. *Nature Genetics* **47**(3): 291–295.

[31] Lee, P. H., Feng, Y. C. A. and Smoller, J. W. [2021]. Pleiotropy and Cross-Disorder Genetics Among Psychiatric Disorders, *Biological psychiatry* **89**(1): 20–31.

[32] Legge, S. E., Santoro, M. L., Periyasamy, S., Okewole, A., Arsalan, A. and Kowalec, K. [2021]. Genetic architecture of schizophrenia: a review of major advancements, *Psychological Medicine* **51**(13): 2168–2177.

[33] Levchenko, A. and Plotnikova, M. [2023]. Genomic regulatory sequences in the pathogenesis of bipolar disorder, *Frontiers in psychiatry* **14**.

[34] Lewis, C. M. and Vassos, E. [2020]. Polygenic risk scores: From research tools to clinical instruments, *Genome Medicine* **12**(1): 1–11.

[35] McIntyre, R. S., Berk, M., Brietzke, E., Goldstein, B. I., López-Jaramillo, C., Kessing, L. V., Malhi, G. S., Nierenberg, A. A., Rosenblat, J. D., Majeed, A., Vieta, E., Vinberg, M., Young, A. H. and Mansur, R. B. [2020]. Bipolar disorders, *Lancet (London, England)* **396**(10265): 1841–1856.

[36] Mu, Q., Chen, Y. and Wang, J. [2019]. Deciphering Brain Complexity Using Single-cell Sequencing, *Genomics, Proteomics Bioinformatics* **17**(4): 344.

[37] Nagalakshmi, U., Waern, K. and Snyder, M. [2010]. RNA-Seq: a method for comprehensive transcriptome analysis, *Current protocols in molecular biology* **Chapter 4**(SUPPL. 89).

[38] Nakazawa, K., Zsiros, V., Jiang, Z., Nakao, K., Kolata, S., Zhang, S. and Belforte, J. E. [2012]. GABAergic interneuron origin of schizophrenia pathophysiology, *Neuropharmacology* **62**(3): 1574–1583.

[39] Nestler, E. J. and Hyman, S. E. [2010]. Animal models of neuropsychiatric disorders, *Nature neuroscience* **13**(10): 1161–1169.

[40] Pardiñas, 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., Bækvad-Hansen, M., Mortensen, P. B., Pedersen, C. B., Børglum, 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., Harold, D., Sims, R., Gerrish, A., Chapman, J., Abraham, R., Hollingworth, P., Pahwa, J., Denning, N., Thomas, C., Taylor, S., Powell, J., Proitsi, P., Lupton, M., Lovestone, S., Passmore, P., Craig, D., McGuinness, B., Johnston, J., Todd, S., Maier, W., Jessen, F., Heun, R., Schurmann, B., Ramirez, A., Becker, T., Herold, C., Lacour, A., Drichel, D., Nothen, M., Goate, A., Cruchaga, C., Nowotny, P., Morris, J. C., Mayo, K., O'Donovan, M. C., Owen, M. J., Williams, J., Achilla, E., Barr, C. L., Böttger, T. W., Cohen, D., Curran, S., Dempster, E., Dima, D., Sabes-Figuera, R., Flanagan, R. J., Frangou, S., Frank, J., Gasse, C., Gaughran, F., Giegling, I., Hannon, E., Hartmann, A. M., Heißerer, B., Helthuis, M., Horsdal, H. T., Ingimarsson, O., Jollie, K., Kennedy, J. L., Köhler, O., Konte, B., Lang, M., Lewis, C., MacCaba, J., Malhotra, A. K., McCrone, P., Meier, S. M., Mill, J., Nöthen, M. M., Pedersen, C. B., Rietschel, M., Rujescu, D., Schwalber, A., Sørensen, H. J., Spencer, B., Støvring, H., Strohmaier, J., Sullivan, P., Vassos, E., Verbelen, M., Collier, D. A., Kirov, G. and Walters, J. T. [2018]. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection, *Nature Genetics 2018 50:3* **50**(3): 381–389.

[41] *Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases* [2014]. *American Journal of Human Genetics* **95**(5): 535–552.

[42] Portin, P. and Alanen, Y. O. [1997]. A critical review of genetic studies of schizophrenia. I. Epidemiological and brain studies, *Acta psychiatrica Scandinavica* **95**(1): 1–5.

[43] Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., Ruderfer, D. M., McQuillin, A., Morris, D. W., Oĝdushlaine, C. T., Corvin, A., Holmans, P. A., Oĝdonovan, M. C., MacGregor, S., Gurling, H., Blackwood, D. H., Craddock, N. J., Gill, M., Hultman, C. M., Kirov, G. K., Lichtenstein, P., Muir, W. J., Owen, M. J., Pato, C. N., Scolnick, E. M., St Clair, D., Williams, N. M., Georgieva, L., Nikolov, I., Norton, N., Williams, H., Toncheva, D., Milanova, V., Thelander, E. F., O'Dushlaine, C. T., Kenny, E., Quinn, E. M., Choudhury, K., Datta, S., Pimm, J., Thirumalai, S., Puri, V., Krasucki, R., Lawrence, J., Quested, D., Bass, N., Crombie, C., Fraser, G., Leh Kuan, S., Walker, N., McGhee, K. A., Pickard, B., Malloy, P., MacLean, A. W., Van Beck, M., Pato, M. T., Medeiros, H., Middleton, F., Carvalho, C., Morley, C., Fanous, A., Conti, D., Knowles, J. A., Paz Ferreira, C., MacEdo, A., Helena Azevedo, M., Kirby, A. N., Ferreira, M. A., Daly, M. J., Chambert, K., Kuruvilla, F., Gabriel, S. B., Ardlie, K., Moran, J. L. and Sklar, P. [2009]. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder, *Nature 2009 460:7256* **460**(7256): 748–752.

[44] Raabe, F. J., Galinski, S., Papiol, S., Falkai, P. G., Schmitt, A. and Rossner, M. J. [2018]. Studying and modulating schizophrenia-associated dysfunctions of oligodendrocytes with patient-specific cell systems, *npj Schizophrenia* **4**(1).

[45] Raabe, F. J., Slapakova, L., Rossner, M. J., Cantuti-Castelvetri, L., Simons, M., Falkai, P. G. and Schmitt, A. [2019]. Oligodendrocytes as A New Therapeutic Target in Schizophrenia: From Histopathological Findings to Neuron-Oligodendrocyte Interaction, *Cells* **8**(12).
URL: */pmc/articles/PMC6952785/ /pmc/articles/PMC6952785/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6952785/*

[46] Räsänen, N., Tiihonen, J., Koskuvi, M., Lehtonen, Š. and Koistinaho, J. [2022]. The iPSC perspective on schizophrenia, *Trends in Neurosciences* **45**(1): 8–26.

[47] Reed, E., Nunez, S., Kulp, D., Qian, J., Reilly, M. P. and Foulkes, A. S. [2015]. A guide to genome-wide association analysis and post-analytic interrogation, *Statistics in medicine* **34**(28): 3769–3792.

[48] Rees, E., Walters, J. T., Georgieva, L., Isles, A. R., Chambert, K. D., Richards, A. L., Mahoney-Davies, G., Legge, S. E., Moran, J. L., McCarroll, S. A.,

O'Donovan, M. C., Owen, M. J. and Kirov, G. [2014]. Analysis of copy number variations at 15 schizophrenia-associated loci, *The British Journal of Psychiatry* **204**(2): 108.
URL: */pmc/articles/PMC3909838/ /pmc/articles/PMC3909838/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3909838/*

[49] Richards, A. L., Cardno, A., Harold, G., Craddock, N. J., Di Florio, A., Jones, L., Gordon-Smith, K., Jones, I., Sellers, R., Walters, J. T., Holmans, P. A., Owen, M. J. and O'Donovan, M. C. [2022]. Genetic Liabilities Differentiating Bipolar Disorder, Schizophrenia, and Major Depressive Disorder, and Phenotypic Heterogeneity in Bipolar Disorder, *JAMA psychiatry* **79**(10): 1032–1039.

[50] Ripke, S., Neale, B. M., Corvin, A., Walters, J. T., Farh, K. H., Holmans, P. A., Lee, P., Bulik-Sullivan, B., Collier, D. A., Huang, H., Pers, T. H., Agartz, I., Agerbo, E., Albus, M., Alexander, M., Amin, F., Bacanu, S. A., Begemann, M., Belliveau, R. A., Bene, J., Bergen, S. E., Bevilacqua, E., Bigdeli, T. B., Black, D. W., Bruggeman, R., Buccola, N. G., Buckner, R. L., Byerley, W., Cahn, W., Cai, G., Campion, D., Cantor, R. M., Carr, V. J., Carrera, N., Catts, S. V., Chambert, K. D., Chan, R. C., Chen, R. Y., Chen, E. Y., Cheng, W., Cheung, E. F., Chong, S. A., Cloninger, C. R., Cohen, D., Cohen, N., Cormican, P., Craddock, N., Crowley, J. J., Curtis, D., Davidson, M., Davis, K. L., Degenhardt, F., Del Favero, J., Demontis, D., Dikeos, D., Dinan, T., Djurovic, S., Donohoe, G., Drapeau, E., Duan, J., Dudbridge, F., Durmishi, N., Eichhammer, P., Eriksson, J., Escott-Price, V., Essioux, L., Fanous, A. H., Farrell, M. S., Frank, J., Franke, L., Freedman, R., Freimer, N. B., Friedl, M., Friedman, J. I., Fromer, M., Genovese, G., Georgieva, L., Giegling, I., Giusti-Rodríguez, P., Godard, S., Goldstein, J. I., Golimbet, V., Gopal, S., Gratten, J., De Haan, L., Hammer, C., Hamshere, M. L., Hansen, M., Hansen, T., Haroutunian, V., Hartmann, A. M., Henskens, F. A., Herms, S., Hirschhorn, J. N., Hoffmann, P., Hofman, A., Hollegaard, M. V., Hougaard, D. M., Ikeda, M., Joa, I., Julià, A., Kahn, R. S., Kalaydjieva, L., Karachanak-Yankova, S., Karjalainen, J., Kavanagh, D., Keller, M. C., Kennedy, J. L., Khrunin, A., Kim, Y., Klovins, J., Knowles, J. A., Konte, B., Kucinskas, V., Kucinskiene, Z. A., Kuzelova-Ptackova, H., Kähler, A. K., Laurent, C., Keong, J. L. C., Lee, S. H., Legge, S. E., Lerer, B., Li, M., Li, T., Liang, K. Y., Lieberman, J., Limborska, S., Loughland, C. M., Lubinski, J., Lönnqvist, J., Macek, M., Magnusson, P. K., Maher, B. S., Maier, W., Mallet, J., Marsal, S., Mattheisen, M., Mattingsdal,

M., McCarley, R. W., McDonald, C., McIntosh, A. M., Meier, S., Meijer, C. J., Melegh, B., Melle, I., Mesholam-Gately, R. I., Metspalu, A., Michie, P. T., Milani, L., Milanova, V., Mokrab, Y., Morris, D. W., Mors, O., Murphy, K. C., Murray, R. M., Myin-Germeys, I., Müller-Myhsok, B., Nelis, M., Nenadic, I., Nertney, D. A., Nestadt, G., Nicodemus, K. K., Nikitina-Zake, L., Nisenbaum, L., Nordin, A., O'Callaghan, E., O'Dushlaine, C., O'Neill, F. A., Oh, S. Y., Olincy, A., Olsen, L., Van Os, J., Pantelis, C., Papadimitriou, G. N., Papiol, S., Parkhomenko, E., Pato, M. T., Paunio, T., Pejovic-Milovancevic, M., Perkins, D. O., Pietiläinen, O., Pimm, J., Pocklington, A. J., Powell, J., Price, A., Pulver, A. E., Purcell, S. M., Quested, D., Rasmussen, H. B., Reichenberg, A., Reimers, M. A., Richards, A. L., Roffman, J. L., Roussos, P., Ruderfer, D. M., Salomaa, V., Sanders, A. R., Schall, U., Schubert, C. R., Schulze, T. G., Schwab, S. G., Scolnick, E. M., Scott, R. J., Seidman, L. J., Shi, J., Sigurdsson, E., Silagadze, T., Silverman, J. M., Sim, K., Slominsky, P., Smoller, J. W., So, H. C., Spencer, C. C., Stahl, E. A., Stefansson, H., Steinberg, S., Stogmann, E., Straub, R. E., Strengman, E., Strohmaier, J., Stroup, T. S., Subramaniam, M., Suvisaari, J., Svrakic, D. M., Szatkiewicz, J. P., Söderman, E., Thirumalai, S., Toncheva, D., Tosato, S., Veijola, J., Waddington, J., Walsh, D., Wang, D., Wang, Q., Webb, B. T., Weiser, M., Wildenauer, D. B., Williams, N. M., Williams, S., Witt, S. H., Wolen, A. R., Wong, E. H., Wormley, B. K., Xi, H. S., Zai, C. C., Zheng, X., Zimprich, F., Wray, N. R., Stefansson, K., Visscher, P. M., Adolfsson, R., Andreassen, O. A., Blackwood, D. H., Bramon, E., Buxbaum, J. D., Børglum, A. D., Cichon, S., Darvasi, A., Domenici, E., Ehrenreich, H., Esko, T., Gejman, P. V., Gill, M., Gurling, H., Hultman, C. M., Iwata, N., Jablensky, A. V., Jönsson, E. G., Kendler, K. S., Kirov, G., Knight, J., Lencz, T., Levinson, D. F., Li, Q. S., Liu, J., Malhotra, A. K., McCarroll, S. A., McQuillin, A., Moran, J. L., Mortensen, P. B., Mowry, B. J., Nöthen, M. M., Ophoff, R. A., Owen, M. J., Palotie, A., Pato, C. N., Petryshen, T. L., Posthuma, D., Rietschel, M., Riley, B. P., Rujescu, D., Sham, P. C., Sklar, P., St Clair, D., Weinberger, D. R., Wendland, J. R., Werge, T., Daly, M. J., Sullivan, P. F. and O'Donovan, M. C. [2014]. Biological insights from 108 schizophrenia-associated genetic loci, *Nature 2014 511:7510* **511**(7510): 421–427.
URL: *https://www.nature.com/articles/nature13595*

[51] Robette, N., Génin, E. and Clerget-Darpoux, F. [2022]. Heritability: What's the point? What is it not for? A human genetics perspective, *Genetica* **150**(3-

4): 199–208.

[52] Rosen, M., Betz, L. T., Schultze-Lutter, F., Chisholm, K., Haidl, T. K., Kambeitz-Ilankovic, L., Bertolino, A., Borgwardt, S., Brambilla, P., Lencer, R., Meisenzahl, E., Ruhrmann, S., Salokangas, R. K., Upthegrove, R., Wood, S. J., Koutsouleris, N. and Kambeitz, J. [2021]. Towards clinical application of prediction models for transition to psychosis: A systematic review and external validation study in the PRONIA sample, *Neuroscience and biobehavioral reviews* **125**: 478–492.
URL: *https://pubmed.ncbi.nlm.nih.gov/33636198/*

[53] Shashikant, T. and Ettensohn, C. A. [2019]. Genome-wide analysis of chromatin accessibility using ATAC-seq, *Methods in cell biology* **151**: 219–235.
URL: *https://pubmed.ncbi.nlm.nih.gov/30948010/*

[54] Sinclair, D., Fillman, S. G., Webster, M. J. and Weickert, C. S. [2013]. Dysregulation of glucocorticoid receptor co-factors FKBP5, BAG1 and PTGES3 in prefrontal cortex in psychotic illness, *Scientific reports* **3**.
URL: *https://pubmed.ncbi.nlm.nih.gov/24345775/*

[55] Skene, N. G., Bryois, J., Bakken, T. E., Breen, G., Crowley, J. J., Gaspar, H. A., Giusti-Rodriguez, P., Hodge, R. D., Miller, J. A., Muñoz-Manchado, A. B., O'Donovan, M. C., Owen, M. J., Pardiñas, A. F., Ryge, J., Walters, J. T., Linnarsson, S., Lein, E. S., Sullivan, P. F. and Hjerling-Leffler, J. [2018]. Genetic identification of brain cell types underlying schizophrenia, *Nature Genetics 2018 50:6* **50**(6): 825–833.
URL: *https://www.nature.com/articles/s41588-018-0129-5*

[56] Smoller, J. W., Andreassen, O. A., Edenberg, H. J., Faraone, S. V., Glatt, S. J. and Kendler, K. S. [2019]. Psychiatric Genetics and the Structure of Psychopathology HHS Public Access, *Mol Psychiatry* **24**(3): 409–420.

[57] Smoller, J. W., Andreassen, O. A., Edenberg, H. J., Faraone, S. V., Glatt, S. J. and Kendler, K. S. [2020]. Psychiatric Genetics and the Structure of Psychopathology, **24**(3): 409–420.

[58] Song, L. and Crawford, G. E. [2010]. Dnase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian

cells, *Cold Spring Harbor Protocols* **2010**(2): pdb–prot5384.

[59] Srivastava, A., Dada, O., Qian, J., Al-Chalabi, N., Fatemi, A. B., Gerretsen, P., Graff, A. and De Luca, V. [2021]. Epigenetics of Schizophrenia, *Psychiatry research* **305**.

[60] Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J. R., Gaspar, H. A., de Leeuw, C. A., Steinberg, S., Pavlides, J. M., 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., Bækvad-Hansen, M., Barchas, J. D., Bass, N., Bauer, M., Belliveau, R., Bergen, S. E., Pedersen, C. B., Bøen, 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., Elvsåshagen, T., Escott-Price, V., Fan, C. C., Fischer, S. B., Flickinger, M., Foroud, T. M., Forty, L., Frank, J., Fraser, C., Freimer, N. B., Frisén, 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., Juréus, 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., Mühleisen, T. W., Mullins, N., Nguyen, H., Nievergelt, C. M., Adolfsson, A. N., Nwulia, E. A., O'Donovan, C., Loohuis, L. M., Ori, A. P., Oruc, L., Ösby, 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., Sánchez-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., Søholm 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., Zöllner, S., Adolfsson, R., Agartz, I., Alda, M., Backlund, L., Baune, B. T., Bellivier, F., Berrettini, W. H., Biernacka, J. M., Blackwood, D. H., Boehnke, M., Børglum, 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., Landén, 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., Müller-Myhsok, B., Myers, R. M., Neale, B. M., Nimgaonkar, V., Nordentoft, M., Nöthen, 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., Ribasés, 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. and Sklar, P. [2019]. Genome-wide association study identifies 30 loci associated with bipolar disorder, *Nature Genetics 2019 51:5* **51**(5): 793–803.

[61] Stein, D. J., Phillips, K. A., Bolton, D., Fulford, K. W. M., Sadler, J. Z. and Kendler, K. S. [n.d.]. What is a mental/psychiatric disorder? From DSM-IV to DSM-V.

[62] Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G. and Meyre, D. [2019]. Benefits and limitations of genome-wide association studies, *Nature reviews. Genetics* **20**(8): 467–484.

[63] Tandon, R., Gaebel, W., Barch, D. M., Bustillo, J., Gur, R. E., Heckers, S., Malaspina, D., Owen, M. J., Schultz, S., Tsuang, M., Van Os, J. and Carpenter, W. [2013]. Definition and description of schizophrenia in the DSM-5, *Schizophrenia research* **150**(1): 3–10.

[64] Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T. K., Diegel, M., Dunn, D., Ebersol, A. K., Frum, T., Giste, E., Johnson, A. K., Johnson, E. M., Kutyavin, T., Lajoie, B., Lee, B. K., Lee, K., London, D., Lotakis, D., Neph, S., Neri, F., Nguyen, E. D., Qu, H., Reynolds, A. P., Roach, V., Safi, A., Sanchez, M. E., Sanyal, A., Shafer, A., Simon, J. M., Song, L., Vong, S., Weaver, M., Yan, Y., Zhang, Z., Zhang, Z., Lenhard, B., Tewari, M., Dorschner, M. O., Hansen, R. S., Navas, P. A., Stamatoyannopoulos, G., Iyer, V. R., Lieb, J. D., Sunyaev, S. R., Akey, J. M., Sabo, P. J., Kaul, R., Furey, T. S., Dekker, J., Crawford, G. E. and Stamatoyannopoulos, J. A. [2012]. The accessible chromatin landscape of the human genome, *Nature 2012 489:7414* **489**(7414): 75–82.

[65] Trubetskoy, V., Pardiñas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T. B., Bryois, J., Chen, C. Y., Dennison, C. A., Hall, L. S., Lam, M., Watanabe, K., Frei, O., Ge, T., Harwood, J. C., Koopmans, F., Magnusson, S., Richards, A. L., Sidorenko, J., Wu, Y., Zeng, J., Grove, J., Kim, M., Li, Z., Voloudakis, G., Zhang, W., Adams, M., Agartz, I., Atkinson, E. G., Agerbo, E., Al Eissa, M., Albus, M., Alexander, M., Alizadeh, B. Z., Alptekin, K., Als, T. D., Amin, F., Arolt, V., Arrojo, M., Athanasiu, L., Azevedo, M. H., Bacanu, S. A., Bass, N. J., Begemann, M., Belliveau, R. A., Bene, J., Benyamin, B., Bergen, S. E., Blasi, G., Bobes, J., Bonassi, S., Braun, A., Bressan, R. A., Bromet, E. J., Bruggeman, R., Buckley, P. F., Buckner, R. L., Bybjerg-Grauholm, J., Cahn, W., Cairns, M. J., Calkins, M. E., Carr, V. J., Castle, D., Catts, S. V., Chambert, K. D., Chan, R. C., Chaumette, B., Cheng, W., Cheung, E. F., Chong, S. A., Cohen, D., Consoli, A., Cordeiro, Q., Costas, J., Curtis, C., Davidson, M., Davis, K. L., de Haan, L., Degenhardt, F., DeLisi, L. E., Demontis, D., Dickerson, F., Dikeos, D., Dinan, T., Djurovic, S., Duan, J., Ducci, G., Dudbridge, F., Eriksson, J. G., Fañanás, L., Faraone, S. V., Fiorentino, A., Forstner, A., Frank, J., Freimer, N. B., Fromer, M., Frustaci, A., Gadelha, A., Genovese, G., Gershon, E. S., Giannitelli, M., Giegling, I., Giusti-Rodríguez, P., Godard, S., Goldstein, J. I., González Peñas, J., González-Pinto, A., Gopal, S., Gratten, J., Green, M. F., Greenwood, T. A., Guillin, O., Gülöksüz, S., Gur, R. E., Gur, R. C., Gutiérrez, B., Hahn, E., Hakonarson, H., Haroutunian, V., Hartmann, A. M., Harvey, C., Hayward, C., Henskens, F. A., Herms, S., Hoffmann, P., Howrigan, D. P., Ikeda, M., Iyegbe, C., Joa, I.,

Julià, A., Kähler, A. K., Kam-Thong, T., Kamatani, Y., Karachanak-Yankova, S., Kebir, O., Keller, M. C., Kelly, B. J., Khrunin, A., Kim, S. W., Klovins, J., Kondratiev, N., Konte, B., Kraft, J., Kubo, M., Kučinskas, V., Kučinskiene, Z. A., Kusumawardhani, A., Kuzelova-Ptackova, H., Landi, S., Lazzeroni, L. C., Lee, P. H., Legge, S. E., Lehrer, D. S., Lencer, R., Lerer, B., Li, M., Lieberman, J., Light, G. A., Limborska, S., Liu, C. M., Lönnqvist, J., Loughland, C. M., Lubinski, J., Luykx, J. J., Lynham, A., Macek, M., Mackinnon, A., Magnusson, P. K., Maher, B. S., Maier, W., Malaspina, D., Mallet, J., Marder, S. R., Marsal, S., Martin, A. R., Martorell, L., Mattheisen, M., McCarley, R. W., McDonald, C., McGrath, J. J., Medeiros, H., Meier, S., Melegh, B., Melle, I., Mesholam-Gately, R. I., Metspalu, A., Michie, P. T., Milani, L., Milanova, V., Mitjans, M., Molden, E., Molina, E., Molto, M. D., Mondelli, V., Moreno, C., Morley, C. P., Muntané, G., Murphy, K. C., Myin-Germeys, I., Nenadić, I., Nestadt, G., Nikitina-Zake, L., Noto, C., Nuechterlein, K. H., O'Brien, N. L., O'Neill, F. A., Oh, S. Y., Olincy, A., Ota, V. K., Pantelis, C., Papadimitriou, G. N., Parellada, M., Paunio, T., Pellegrino, R., Periyasamy, S., Perkins, D. O., Pfuhlmann, B., Pietiläinen, O., Pimm, J., Porteous, D., Powell, J. F., Quattrone, D., Quested, D., Radant, A. D., Rampino, A., Rapaport, M. H., Rautanen, A., Reichenberg, A., Roe, C., Roffman, J. L., Roth, J., Rothermundt, M., Rutten, B. P., Saker-Delye, S., Salomaa, V., Sanjuan, J., Santoro, M. L., Savitz, A., Schall, U., Scott, R. J., Seidman, L. J., Sharp, S. I., Shi, J., Siever, L. J., Sigurdsson, E., Sim, K., Skarabis, N., Slominsky, P., So, H. C., Sobell, J. L., Söderman, E., Stain, H. J., Steen, N. E., Steixner-Kumar, A. A., Stögmann, E., Stone, W. S., Straub, R. E., Streit, F., Strengman, E., Stroup, T. S., Subramaniam, M., Sugar, C. A., Suvisaari, J., Svrakic, D. M., Swerdlow, N. R., Szatkiewicz, J. P., Ta, T. M. T., Takahashi, A., Terao, C., Thibaut, F., Toncheva, D., Tooney, P. A., Torretta, S., Tosato, S., Tura, G. B., Turetsky, B. I., Üçok, A., Vaaler, A., van Amelsvoort, T., van Winkel, R., Veijola, J., Waddington, J., Walter, H., Waterreus, A., Webb, B. T., Weiser, M., Williams, N. M., Witt, S. H., Wormley, B. K., Wu, J. Q., Xu, Z., Yolken, R., Zai, C. C., Zhou, W., Zhu, F., Zimprich, F., Atbaşoğlu, E. C., Ayub, M., Benner, C., Bertolino, A., Black, D. W., Bray, N. J., Breen, G., Buccola, N. G., Byerley, W. F., Chen, W. J., Cloninger, C. R., Crespo-Facorro, B., Donohoe, G., Freedman, R., Galletly, C., Gandal, M. J., Gennarelli, M., Hougaard, D. M., Hwu, H. G., Jablensky, A. V., McCarroll, S. A., Moran, J. L., Mors, O., Mortensen, P. B., Müller-Myhsok, B., Neil, A. L., Nordentoft, M., Pato, M. T., Petryshen, T. L., Pirinen, M., Pulver, A. E., Schulze, T. G., Silverman, J. M., Smoller,

J. W., Stahl, E. A., Tsuang, D. W., Vilella, E., Wang, S. H., Xu, S., Dai, N., Wenwen, Q., Wildenauer, D. B., Agiananda, F., Amir, N., Antoni, R., Arsianti, T., Asmarahadi, A., Diatri, H., Djatmiko, P., Irmansyah, I., Khalimah, S., Kusumadewi, I., Kusumaningrum, P., Lukman, P. R., Nasrun, M. W., Safyuni, N. S., Prasetyawan, P., Semen, G., Siste, K., Tobing, H., Widiasih, N., Wiguna, T., Wulandari, D., Evalina, N., Hananto, A. J., Ismoyo, J. H., Marini, T. M., Henuhili, S., Reza, M., Yusnadewi, S., Abyzov, A., Akbarian, S., Ashley-Koch, A., van Bakel, H., Breen, M., Brown, M., Bryois, J., Carlyle, B., Charney, A., Coetzee, G., Crawford, G., Dracheva, S., Emani, P., Farnham, P., Galeev, T., Gandal, M., Gerstein, M., Giase, G., Girdhar, K., Goes, F., Grennan, K., Gu, M., Guerra, B., Gursoy, G., Hoffman, G., Hyde, T., Jaffe, A., Jiang, S., Jiang, Y., Kefi, A., Kim, Y., Kitchen, R., Knowles, J. A., Lay, F., Lee, D., Li, M., Liu, C., Liu, S., Mattei, E., Navarro, F., Pan, X., Peters, M. A., Pinto, D., Pochareddy, S., Polioudakis, D., Purcaro, M., Purcell, S., Pratt, H., Reddy, T., Rhie, S., Roussos, P., Rozowsky, J., Sanders, S., Sestan, N., Sethi, A., Shi, X., Shieh, A., Swarup, V., Szekely, A., Wang, D., Warrell, J., Weissman, S., Weng, Z., White, K., Wiseman, J., Witt, H., Won, H., Wood, S., Wu, F., Xu, X., Yao, L., Zandi, P., Arranz, M. J., Bakker, S., Bender, S., Bramon, E., Collier, D. A., Crepo-Facorro, B., Hall, J., Iyegbe, C., Kahn, R., Lawrie, S., Lewis, C., Lin, K., Linszen, D. H., Mata, I., McIntosh, A., Murray, R. M., van Os, J., Rujescu, D., Walshe, M., Weisbrod, M., Achsel, T., Andres-Alonso, M., Bagni, C., Bayés, À., Biederer, T., Brose, N., Brown, T. C., Chua, J. J. E., Coba, M. P., Cornelisse, L. N., de Jong, A. P., de Juan-Sanz, J., Dieterich, D. C., Feng, G., Goldschmidt, H. L., Gundelfinger, E. D., Hoogenraad, C., Huganir, R. L., Hyman, S. E., Imig, C., Jahn, R., Jung, H., Kaeser, P. S., Kim, E., Koopmans, F., Kreutz, M. R., Lipstein, N., MacGillavry, H. D., Malenka, R., McPherson, P. S., O'Connor, V., Pielot, R., Ryan, T. A., Sahasrabudhe, D., Sala, C., Sheng, M., Smalla, K. H., Smit, A. B., Südhof, T. C., Thomas, P. D., Toonen, R. F., van Weering, J. R., Verhage, M., Verpelli, C., Adolfsson, R., Arango, C., Baune, B. T., Belangero, S. I., Børglum, A. D., Braff, D., Bramon, E., Buxbaum, J. D., Campion, D., Cervilla, J. A., Cichon, S., Collier, D. A., Corvin, A., Curtis, D., Forti, M. D., Domenici, E., Ehrenreich, H., Escott-Price, V., Esko, T., Fanous, A. H., Gareeva, A., Gawlik, M., Gejman, P. V., Gill, M., Glatt, S. J., Golimbet, V., Hong, K. S., Hultman, C. M., Hyman, S. E., Iwata, N., Jönsson, E. G., Kahn, R. S., Kennedy, J. L., Khusnutdinova, E., Kirov, G., Knowles, J. A., Krebs, M. O., Laurent-Levinson, C., Lee, J., Lencz, T., Levinson, D. F., Li, Q. S., Liu, J., Malhotra, A. K., Malhotra, D.,

McIntosh, A., McQuillin, A., Menezes, P. R., Morgan, V. A., Morris, D. W., Mowry, B. J., Murray, R. M., Nimgaonkar, V., Nöthen, M. M., Ophoff, R. A., Paciga, S. A., Palotie, A., Pato, C. N., Qin, S., Rietschel, M., Riley, B. P., Rivera, M., Rujescu, D., Saka, M. C., Sanders, A. R., Schwab, S. G., Serretti, A., Sham, P. C., Shi, Y., St Clair, D., Stefánsson, H., Stefansson, K., Tsuang, M. T., van Os, J., Vawter, M. P., Weinberger, D. R., Werge, T., Wildenauer, D. B., Yu, X., Yue, W., Holmans, P. A., Pocklington, A. J., Roussos, P., Vassos, E., Verhage, M., Visscher, P. M., Yang, J., Posthuma, D., Andreassen, O. A., Kendler, K. S., Owen, M. J., Wray, N. R., Daly, M. J., Huang, H., Neale, B. M., Sullivan, P. F., Ripke, S., Walters, J. T., O'Donovan, M. C., de Haan, L., van Amelsvoort, T., van Winkel, R., Gareeva, A., Shi, Y., St Clair, D. and van Os, J. [2022]. Mapping genomic loci implicates genes and synaptic biology in schizophrenia, *Nature* **604**(7906): 502–508.

[66] Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T. and Posthuma, D. [2021]. Genome-wide association studies, *Nature Reviews Methods Primers* **1**(1).

[67] Wahbeh, M. H. and Avramopoulos, D. [2021]. Gene-Environment Interactions in Schizophrenia: A Literature Review.

[68] Winship, I. R., Dursun, S. M., Baker, G. B., Balista, P. A., Kandratavicius, L., Maia-De-Oliveira, J. P., Hallak, J., Howland, J. G., Howland, J. and Winship, I. [2019]. In Review Series Article, *The Canadian Journal of Psychiatry / La Revue Canadienne de Psychiatrie* **64**(1): 5–17.

[69] Wu, C., Bendriem, R. M., Garamszegi, S. P., Song, L. and Lee, C. T. [2017a]. RNA sequencing in post-mortem human brains of neuropsychiatric disorders, *Psychiatry and Clinical Neurosciences* **71**(10): 663–672.

[70] Wu, C., Bendriem, R. M., Garamszegi, S. P., Song, L. and Lee, C. T. [2017b]. RNA sequencing in post-mortem human brains of neuropsychiatric disorders, *Psychiatry and clinical neurosciences* **71**(10): 663–672.

[71] Yuan, W., Ma, S., Brown, J. R., Kim, K., Murek, V., Trastulla, L., Meissner, A., Lodato, S., Shetty, A. S., Levin, J. Z., Buenrostro, J. D., Ziller, M. J. and Arlotta, P. [2022]. Temporally divergent regulatory mechanisms govern neuronal diversification and maturation in the mouse and marmoset neocortex,

Nature Neuroscience 2022 25:8 **25**(8): 1049–1058.

[72] Zhang, Y., Pak, C., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., Marro, S., Patzke, C., Acuna, C., Covy, J., Xu, W., Yang, N., Danko, T., Chen, L., Wernig, M. and Sü Dhof, T. C. [n.d.]. Rapid Single-Step Induction of Functional Neurons from Human Pluripotent Stem Cells.
URL: *http://dx.doi.org/10.1016/j.neuron.2013.05.029*

[73] Zuccaro, E., Murek, V., Kim, K., Chen, H.-H., Mancinelli, S., Oyler-Castrillo, P., Jiménez-Barrón, L. T., Gerhardinger, C., Brown, J. R., Byrnes, A., Neale, B. M., Levin, J. Z., Ziller, M. J., Lodato, S. and Arlotta, P. [2021]. Human-specific enrichment of schizophrenia risk-genes in callosal neurons of the developing neocortex, *bioRxiv* p. 2021.09.10.459747.
URL: *http://biorxiv.org/content/early/2021/09/11/2021.09.10.459747.abstract*

B ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to all those who have contributed to the realization of this Ph.D. thesis through their support in various capacities.

First and foremost, I extend my sincere appreciation to my committee members: Prof. Dr. Nikolaos Koutsouleris, Prof. Dr. Moritz Rossner, and especially Prof. Dr. Michael Ziller for allowing me to pursue this project in his lab and for putting together the best research group one can work in.

I am forever thankful to Liesa, Christine, Lucia, and Vanessa for being not only great colleagues but also incredible friends, your support and invaluable contributions were key to the successful completion of this thesis. Furthermore, I would like to acknowledge the exceptional work of Dr. Till Andlauer, Dr, Florian Raabe, Dr. Miriam Gagliardi and Dr. Ruhel Ahmad, whose contributions laid the foundation for this project. I extend my special appreciation to the donors whose generous contributions made this project possible. Without your support, significant advancements in understanding human diseases would have been impossible.

I would like to express my sincere thanks to the coordinators of the International Max Planck School for Translational Psychiatry and the Munich Medical Research School. Their exceptional dedication and efforts went beyond the call of duty to ensure the successful culmination of this work despite any hardships. Special recognition goes to Dr. Isabel Osswald, Dr. Janhvi Jaiswal, and Dr. Antje Hentrich for their guidance and support throughout this journey.

Lastly, I would like to thank all the friends and family who have been there for me all these years, this journey would have been a lot harder and a lot less fun without you. Thank you Rosita and Manfred for being the family I didn't know I had on this side of the Atlantic and for supporting me on this quest. To Rosa Elena, Jason, Tania, Signe, Mez, and Karen, thank you for always rooting for me and lifting my spirits. To Ale, Gonza, Estefi, Lu, and Cami, I literally don't know how people complete PhDs without such supportive friends who also happen to live around the corner. A thousand thanks to Ikal & Tana for making me smile every single day and giving me the strength to keep going. Thank you Grace for believing in me, for going above and beyond to support me in every possible way, and for being my rock throughout every single day of this journey.

C AFFIDAVIT

Affidavit

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN | Promotionsbüro Medizinische Fakultät | MMRS

Affidavit

Jiménez Barrón Laura Teresa

Surname, first name

Kraepelinstraße 10

Street

80804 Munich, Germany

Zip code, town, country

I hereby declare, that the submitted thesis entitled:

Identification of in vitro model systems capable of capturing the polygenic basis of mental illness

……

is my own work. I have only used the sources indicated and have not made unauthorized use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree.

Munich, June 9th

place, date

Laura Teresa Jiménez Barrón

Signature doctoral candidate

D CONFIRMATION OF CONGRUENCY

Confirmation of congruency

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Promotionsbüro Medizinische Fakultät

MMRS

Confirmation of congruency between printed and electronic version of the doctoral thesis

Jiménez Barrón Laura Teresa

Surname, first name

Kraepelinstraße 10

Street

80804 Munich, Germany

Zip code, town, country

I hereby declare, that the submitted thesis entitled:

Identification of in vitro model systems capable of capturing the polygenic basis of mental illness

……

is congruent with the printed version both in content and format.

Munich, June 9th

place, date

Laura Teresa Jiménez Barrón

Signature doctoral candidate

E LIST OF PUBLICATIONS

In chronological order:

Jiménez-Barrón LT, O'Rawe JA, Wu Y, Yoon M, Fang H, Iossifov I, Lyon GJ. Genome-wide variant analysis of simplex autism families with an integrative clinical-bioinformatics pipeline. Cold Spring Harb Mol Case Stud. 2015 Oct;1(1):a000422. doi: 10.1101/mcs.a000422. PMID: 27148569; PMCID: PMC4850892.

O'Rawe JA, Wu Y, Dörfel MJ, Rope AF, Au PY, Parboosingh JS, Moon S, Kousi M, Kosma K, Smith CS, Tzetis M, Schuette JL, Hufnagel RB, Prada CE, Martinez F, Orellana C, Crain J, Caro-Llopis A, Oltra S, Monfort S, Jiménez-Barrón LT, Swensen J, Ellingwood S, Smith R, Fang H, Ospina S, Stegmann S, Den Hollander N, Mittelman D, Highnam G, Robison R, Yang E, Faivre L, Roubertie A, Rivière JB, Monaghan KG, Wang K, Davis EE, Katsanis N, Kalscheuer VM, Wang EH, Metcalfe K, Kleefstra T, Innes AM, Kitsiou-Tzeli S, Rosello M, Keegan CE, Lyon GJ. TAF1 Variants Are Associated with Dysmorphic Features, Intellectual Disability, and Neurological Manifestations. Am J Hum Genet. 2015 Dec 3;97(6):922-32. doi: 10.1016/j.ajhg.2015.11.005. PMID: 26637982; PMCID: PMC4678794.

Fang H, Bergmann EA, Arora K, Vacic V, Zody MC, Iossifov I, O'Rawe JA, Wu Y, Jimenez Barron LT, Rosenbaum J, Ronemus M, Lee YH, Wang Z, Dikoglu E, Jobanputra V, Lyon GJ, Wigler M, Schatz MC, Narzisi G. Indel variant analysis of short-read sequencing data with Scalpel. Nat Protoc. 2016 Dec;11(12):2529-2548. doi: 10.1038/nprot.2016.150. Epub 2016 Nov 17. PMID: 27854363; PMCID: PMC5507611.

Zuccaro, E., Murek, V., Kim, K., Chen, H.-H., Mancinelli, S., Oyler-Castrillo, P., Jiménez-Barrón, L. T., Gerhardinger, C., Brown, J. R., Byrnes, A., Neale, B. M., Levin, J. Z., Ziller, M. J., Lodato, S., & Arlotta, P. (2021). Human-specific enrichment of schizophrenia risk-genes in callosal neurons of the developing neocortex. BioRxiv, 2021.09.10.459747. https://doi.org/10.1101/2021.09.10.459747

Trastulla L, Moser S, Jiménez-Barrón LT, Andlauer TFM, von Scheidt M; Schizophrenia Working Group of the Psychiatric Genomics Consortium; Budde M, Heilbronner U, Papiol S, Teumer A, Homuth G, Falkai P, Völzke H, Dörr M, Schulze TG, Gagneur J, Iorio F, Müller-Myhsok B, Schunkert H, Ziller MJ. Distinct genetic liability profiles define clinically relevant patient strata across common diseases. medRxiv [Preprint]. 2023 May 11:2023.05.10.23289788. doi: 10.1101/2023.05.10.23289788. PMID: 37214898; PMCID: PMC10197798.