

---

# FUNCTIONAL ANNOTATION OF CELL TYPE AND CONDITION SPECIFIC SCHIZOPHRENIA ASSOCIATED NON-CODING GENETIC VARIANTS IN DISEASE RELEVANT CELL TYPES

---

Dissertation der Fakultät für Biologie

der Ludwig-Maximilians-Universität München



Christine Kirsten Rummel

aus

Nürnberg (Deutschland)

München, 2023



**Diese Dissertation wurde angefertigt  
unter der Leitung von Prof. Dr. Michael J. Ziller  
in der Arbeitsgruppe Genomics of Complex Diseases  
am Max-Planck-Institut für Psychiatrie.**

**Erstgutachter:** **Prof. Dr. Wolfgang Enard**

**Zweitgutachter:** **Prof. Dr. Michael J. Ziller**

**Tag der Abgabe:** **09.01.2023**

**Tag der mündlichen Prüfung:** **10.07.2023**

## **ERKLÄRUNG**

**Ich versichere hiermit an Eides statt, dass meine Dissertation selbstständig und ohne unerlaubte Hilfsmittel angefertigt worden ist. Die vorliegende Dissertation wurde weder ganz, noch teilweise bei einer anderen Prüfungskommission vorgelegt. Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder an einer Doktorprüfung teilzunehmen.**

**Heidelberg, den 28.07.2023**

---

**Christine Kirsten Rummel**



# TABLE OF CONTENT

<b>1 ABSTRACT .....</b>	<b>13</b>
<b>2 INTRODUCTION .....</b>	<b>15</b>
2.1 THE COMPLEX PSYCHIATRIC DISEASE SCHIZOPHRENIA .....	15
2.1.1 Classifications, symptoms, and social economic effects of schizophrenia .....	15
2.1.2 Inheritance of schizophrenia .....	17
2.1.3 Linkage analysis and SNP association studies in schizophrenia research .....	19
2.1.4 Functional annotation of schizophrenia associated SNPs .....	21
2.2 HUMAN INDUCED PLURIPOTENT STEM CELLS IN SCHIZOPHRENIA RESEARCH.....	25
2.2.1 Generation of human induced pluripotent stem cells from primary cells .....	25
2.2.2 Neural differentiations of pluripotent stem cells .....	27
2.2.3 iPSC derived neural cells as an <i>in vitro</i> model system for schizophrenia .....	29
2.2.4 Schizophrenia induced transcriptional and epigenetic alterations in iPSC derived neurons.....	32
2.3 AIMS OF THIS THESIS .....	35
<b>3 MATERIALS AND METHODS .....</b>	<b>37</b>
3.1 MATERIALS.....	37
3.1.1 Cell Culture Plastics .....	37
3.1.2 Cell culture reagents.....	37
3.1.3 Cell culture media compositions .....	39
3.1.4 Chemicals .....	39
3.1.5 Enzymes, corresponding buffers, master mixes, premade solutions .....	40
3.1.6 Kits.....	40
3.1.7 Plasmid backbones .....	41
3.1.8 Bacteria work .....	42
3.1.9 Immunocytochemistry .....	42
3.2 CELL CULTURE .....	42
3.2.1 Preparation of lentiviruses .....	43
3.2.2 Primary cells derived from mouse brain tissue .....	45
3.2.3 Reprogramming induced pluripotent stem cells from primary human cells.....	46
3.2.4 Maintenance of induced pluripotent stem cells.....	50
3.2.5 Generation of doxycycline inducible dCas9-KRAB cell lines .....	51
3.2.6 Neural differentiations of the patient and healthy control derived iPSCs.....	51
3.2.7 Stimulation of neuronal networks.....	55
3.3 ELECTROPHYSIOLOGICAL CHARACTERIZATION OF THE GENERATED iNEURONS .....	56
3.3.1 Multi electrode array field recordings of mature iNeurons.....	56
3.3.2 Single cell recording of mature iNeurons .....	57
3.4 DETECTION OF PROTEINS IN iPSCS AND iPSC DERIVED NEURONS .....	58
3.4.1 Immunocytochemistry .....	58
3.4.2 Confocal imaging .....	58

3.4.3	Analysis of intensity of expressed markers.....	59
3.4.4	Western Blot with WesSimple .....	61
3.5	RNA LEVELS IN iNEURONS.....	61
3.5.1	Bulk RNA isolation .....	61
3.5.2	Quantitative real time PCR to analyze gene expression .....	62
3.5.3	Bulk 3'RNA sequencing .....	62
3.5.4	Single cell RNA sequencing.....	64
3.6	BULK OPEN CHROMATIN .....	65
3.7	TARGETED REPRESSION OF ENHANCER ACTIVITY USING dCAS9-KRAB.....	68
3.8	GENOME BROWSER REPRESENTATIONS WITH IGV .....	71
3.9	MASSIVELY PARALLEL REPORTER ASSAY.....	72
3.9.1	Selection of the enhancer fragments <i>in silico</i> .....	72
3.9.2	Low bias amplification and purification of the enhancer fragments .....	76
3.9.3	Cloning of the barcoded enhancer fragments into a cloning vector .....	78
3.9.4	Transformation of the library and transformation efficiency.....	79
3.9.5	Subcloning the reporter gene into the initial library for the intermediate library .....	80
3.9.6	Subcloning enhancer-minP-nLuc-barcode fragments into lentiviral vector .....	81
3.9.7	Association by sequencing of the barcodes with their respective enhancer.....	82
3.9.8	Generation of sequencing libraries from different cell types .....	83
3.9.9	Computational analysis of differentially active elements in the MPRA.....	85
3.9.10	Intersection of differential active MPRA regions with genomic data sets.....	87
<b>4</b>	<b>RESULTS .....</b>	<b>88</b>
4.1	GENERATION OF INDUCED PLURIPOTENT STEM CELLS FROM FIBROBLASTS.....	88
4.1.1	Generation of pluripotent stem cells from human skin fibroblasts .....	88
4.1.2	Karyotypic analysis reveals several genomic aberrations caused by the reprogramming process.....	91
4.2	CHARACTERIZATION OF THE NEURONAL IDENTITY OF iNEURONS.....	92
4.2.1	Electrophysiological active burst firing excitatory neurons .....	92
4.2.2	iNeurons positive for neurite and cortical layer markers .....	94
4.2.3	RNA expression verified protein profiles and presence of peripheral neurons.....	98
4.2.4	Single cell RNA-Seq confirmed central and peripheral nervous system identity .....	102
4.3	TRANSCRIPTIONAL ACTIVITY AND REGULATION IN HUMAN DAY-49 iNEURONS.....	111
4.3.1	Schizophrenia altered expression of 5 genes in day-49 iNeurons .....	112
4.3.2	Induction of gene expression after a sustained depolarization stimulus .....	112
4.3.3	Open chromatin upon depolarization of neurons .....	115
4.3.4	Proximity of activity regulated chromatin states and transcription levels .....	118
4.4	DIFFERENTIAL ENHANCER ACTIVITY IN NEURONAL CELL MODELS .....	119
4.4.1	Library design of the different MPRA libraries .....	120
4.4.2	Successful generation of barcoded enhancer libraries .....	122
4.4.3	Efficient strategy for enhancer barcoding and retrieval of barcodes .....	124
4.4.4	Good correlation of oligo pools after normalization of barcode abundance .....	129
4.4.5	Classification of active and inactive enhancer regions .....	131

4.4.6	MPRA identified 620 allele specific active enhancer regions .....	132
4.4.7	Cell type and stimulus dependent active enhancer regions.....	134
4.5	IDENTIFICATION OF TARGET GENES FROM EMVAR REGIONS.....	139
4.5.1	Linkage of 535 emVars to target genes .....	140
4.5.2	Decreased enhancer activity decreased RNA expression at the TCF4 locus.....	140
<b>5</b>	<b>DISCUSSION.....</b>	<b>147</b>
5.1	SUCCESSFUL GENERATION OF KARYOTYPICAL NORMAL IPSC LINES.....	147
5.2	GENERATION OF NGN2 INDUCED EXCITATORY NEURONS FROM IPSCs .....	148
5.2.1	Electrophysiological active, excitatory, and mature iNeurons .....	149
5.2.2	Mature excitatory iNeurons of the peripheral and central nervous system .....	150
5.3	EFFECTS OF SCHIZOPHRENIA ON THE TRANSCRIPTOME AND EPIGENOME .....	157
5.4	DEPOLARIZATION INDUCED CHANGES IN DAY-49 iNEURONS .....	159
5.4.1	Disruption of network bursting after high KCl depolarization.....	159
5.4.2	Transcriptomic and epigenomic alterations after depolarization of iNeurons.....	160
5.5	GENE REGULATORY ACTIVITY IS INFLUENCED BY CELL TYPE, STIMULATION, AND GENOMIC VARIANTS .....	163
5.5.1	Applied strategies to improve sensitivity and specificity of the MPRA .....	163
5.5.2	Context dependent allele specific gene regulatory activity.....	167
5.6	FUNCTIONAL LINK BETWEEN EMVAR LOCI AND ALTERED TARGET GENE EXPRESSION.....	170
5.7	CONCLUSION .....	173
<b>6</b>	<b>BIBLIOGRAPHY .....</b>	<b>174</b>
<b>7</b>	<b>ACKNOWLEDGEMENT .....</b>	<b>187</b>
<b>8</b>	<b>SUPPLEMENTARY TABLES .....</b>	<b>189</b>
8.1	CONTROL FRAGMENTS FOR THE MPRA EXPERIMENTS .....	189
8.2	FRAGMENTS OF THE DEVLIB .....	194
8.3	FRAGMENTS OF THE EQTLIB .....	200
8.4	FRAGMENTS OF THE VALLIB .....	210

## LIST OF TABLES

Table 1 Description of cell culture equipment .....	43
Table 2 Cell type specific settings of the automatized cell counter .....	43
Table 3 Overview of the reprogrammed and utilized iPSC lines in this thesis .....	47
Table 4 Media composition for the differentiation of iPSCs into neural progenitor cells..	51
Table 5 Overview of the applied stimuli on mouse primary neurons.....	56
Table 6 Summary table of the barcodes used for the open chromatin libraries.....	66
Table 7 CRISPRi targeted regions and their gRNA targeted gDNA sequence .....	68
Table 8 Description of the cloning reaction for the gRNA plasmids .....	70
Table 9 Lentiviral particles for CRISPRi mediated inactivation of gene regulation .....	70
Table 10 Primer sequence to quantify gene expression after gene knockdown .....	71
Table 11 Overview of tracks to verify location of SNPs in open chromatin .....	75
Table 12 Restriction digest and dephosphorylation of the pMPRA1 plasmid .....	78
Table 13 Reaction to ligate the barcoded emPCR product into pMPRA1 .....	79
Table 14 Serial dilutions to estimate the number of transformed bacteria colonies.....	79
Table 15 Digest of the nanoLuciferase fragment and pMPRA1-enhancer-barcode.....	80
Table 16 Ligation of pMPRA1-enhancer-barcode and minimalPromoter-nanoLuc .....	81
Table 17 Restriction digest of pMPRAgenti and the MPRA coding insert .....	81
Table 18 Ligation reaction of digested MPRA coding insert and pMPRAgenti .....	81
Table 19 Overview of the cell types and stimulation conditions used for the MPRA .....	83
Table 20 Summary table of the barcodes used for MPRA library indexing .....	84
Table 21 Annotation of copy number variations in reprogrammed iPSC samples .....	91
Table 22 Overview of the bulk RNA Seq samples from day-49 iNeurons .....	99
Table 23 Significantly dysregulated genes between case and control iNeurons.....	112
Table 24 Overview of the ATACSeq samples from day-49 iNeurons .....	116
Table 25 Overview of the number of putative enhancer fragments in each MPRA.....	122
Table 26 Number of colony forming units for the three MPRA experiments .....	123
Table 27 Number of sample replicates per cell type and MPRA .....	125
Table 28 Overview of batches, samples, and read depth for samples used in the MPRA	127

# LIST OF FIGURES

Figure 1 Production of lentiviruses with HEK293T cells .....	44
Figure 2 Timeline of the reprogramming process from primary fibroblasts to iPSCs.....	48
Figure 3 Timeline of the differentiation process from iPSCs to iNeurons.....	52
Figure 4 Visualization of successfully nucleofected primary human fibroblasts .....	88
Figure 5 Morphological changes during the differentiation process .....	89
Figure 6 Successful generation of human induced pluripotent stem cells .....	90
Figure 7 Field and patch clamp recording of excitatory cortical neurons.....	93
Figure 8 Morphological changes during the differentiation of iNeurons .....	95
Figure 9 Expression of neuronal and cortical markers .....	96
Figure 10 Quantification of cortical markers express by the generated neurons .....	97
Figure 11 Quality control metrics of bulk RNA-Seq data.....	100
Figure 12 Neuronal markers expression day-49 iNeurons in bulk RNA-Seq .....	101
Figure 13 Baseline quantifications of single cells grouped in seven clusters.....	103
Figure 14 Cell line specific contribution to different annotated single cell clusters.....	104
Figure 15 Annotation of cell types from the central and peripheral nervous system .....	105
Figure 16 Several stages of neural maturation in day-49 iNeurons.....	106
Figure 17 iNeurons expressed genes related to GABA - and glutamatergic signaling.....	108
Figure 18 Expression of cortical and peripheral marker genes in day-49 iNeurons .....	110
Figure 19 Activity dependent transcriptional alterations after hKCl depolarization .....	113
Figure 20 Locations of open chromatin after hKCl stimulation .....	117
Figure 21 Differential regulation upon sustained depolarization of the neurons .....	118
Figure 22 Schematic overview of different steps for the MPRA experiment .....	119
Figure 23 Selection of 5,601 putative enhancer regions for functional analysis .....	121
Figure 24 Cloning and fragment size quality control of the MPRA libraries .....	124
Figure 25 Successful annotation of enhancers with their corresponding barcodes.....	125
Figure 26 Barcode recovery from technical replicates of the MPRA libraries .....	127
Figure 27 Barcode abundance dependent on cell type and library .....	128
Figure 28 Reproducibility between libraries and cutoffs .....	130
Figure 29 Activity threshold for iNeurons and mouse Neurons.....	131
Figure 30 Classification of allele specific enhancer regions .....	133
Figure 31 Allele specific induction of enhancer activity at the SNX19 gene locus.....	134
Figure 32 Annotation of cell type specific enhancer regions and enhancer elements....	135
Figure 33 Allele specific MPRA activity of rs9806806 at the GRIN2A gene locus .....	136
Figure 34 Condition and cell type specific emVars.....	137
Figure 35 Neuronal activity dependent enhancer activity in neurons.....	138
Figure 36 Stimulation dependent variants at the FURIN locus .....	139
Figure 37 Linking of emVars to potential target genes .....	140
Figure 38 TCF4 expression in human iNeurons and DLPFC tissue.....	141
Figure 39 Cloning strategy and verification of plasmids coding for gRNAs .....	143
Figure 40 dCas9-KRAB dependent closure of open chromatin regions at the TCF4 locus	144



# ABBREVIATIONS

Abbreviation	Definition
4-AP	4-Aminopyridine
A	Ampere
AA	Amino acid
AB	Antibody
ac	Acetylation
ACSF	Artificial cerebrospinal fluid
Alt.	Alternative
AM	Astrocyte media
AMPA	$\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
ANOVA	Analysis of variance
AP	Action potential
APA	Anterior posterior axis
APV	(2R)-amino-5-phosphonovaleric acid
AraC	Cytosine arabinoside
ARCh	Activity regulated chromatin region
ARG	Activity regulated gene
AS-RNA	Antisense RNA
ATAC-Seq	Assay for transposase-accessible chromatin sequencing
BDNF	Brain derived neurotrophic factor
bHLH	Basic helix-loop-helix
bp	Base pair(s)
BPM	BrainPhys media
BSA	Bovine serum albumin
CA1-3	Cornu ammonis 1-3
caQTL	Chromatin accessibility quantitative trait loci
cDNA	Complementary DNA
CFUs	Colony forming units
ChIP	Chromatin immunoprecipitation
Chr.	Chromosome
CI	Confidence interval
cM	Centimorgan
CNS	Central nervous system
CNTF	Ciliary neurotrophic factor
CNVs	Copy number variations
COS	Childhood onset schizophrenia
cpm	Counts per million
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPRi /a	Inactivating /activating CRISPR
Ctrl	Control
Da	Dalton
DAPI	4',6-diamidino-2-phenylindole
dCas9	Endonuclease dead CRISPR-associated enzyme 9
DEGs	Differential expressed genes
DHS	DNase hypersensitivity site
DIV	Day <i>in vitro</i>
DLPFC	Dorsolateral prefrontal cortex
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotides
DPBS	Dulbecco's phosphate buffered saline (noCa2+, no Mg2+)
(DL)PFC	(Dorsolateral) prefrontal cortex

DRG	Dorsal root ganglia
DSM	Diagnostic and Statistical Manual of Mental Disorders
DVA	Dorsal-ventral axis
EB	Embryoid body
EDTA	Ethylenediaminetetraacetate
(p)EFs	(Putative) enhancer fragments
EGTA	Ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid
emVar(s)	Expression modulating variant(s)
(m)EPSCs	(Miniature) excitatory post synaptic currents
eQTL	Expression quantitative trait loci
(p)ERs	(Putative) enhancer regions
ESCs	Embryonic stem cells
Fastq	Qualitative FASTA format
FBM	Fibroblast media
FC	Fold change
FCS	Fetal calf serum
(e)FDR	(Estimated) false discovery rate
FPKM	Fragments per kilobase million
FPR	False positive rate
FW primer	Forward primer
G	Giga
GABA	Gamma amino butyric acid
gDNA	Genomic DNA
GDNF	Glial cell-derived neurotrophic factor
(E)GFP	(Enhanced) green fluorescent protein
GLM	Generalized linear model
GO	Gene ontology
GRE	Gene regulatory element
gRNA	Guide RNA
GWAS	Genome-wide association study
h	Hour(s)
H3K27ac	Histone 3 acetylation on lysine at position 27
H3K4me3	Histone 3 trimethylation at lysine at position 4
HC	Hippocampus
HDAC	Histone deacetylase
HDR	Homologous directed repair
HEK293T	Human primary embryonal kidney cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF	High fidelity
Hg19/38	Human genome version 19 / 38
HGNC	Human genome organization gene nomenclature committee
Hi-C	High resolution chromatin interaction
HIV	Human immunodeficiency virus
hKCl	High potassium chloride
HM	HEK293T media
HS	High sensitivity
Hz	Hertz
IGV	Integrative genomics viewer
IM	Infection media
iNeurons	Ngn2 overexpression induced Neurons
iPSCs	Induced pluripotent stem cells
K	Kilo-
Kb	Kilo bases
KRAB	Krüppel associated box
l	Liter
LB media	Lysogeny broth media

LD	Linkage disequilibrium
LMB	Leptomycin B
Low TE	Tris-EDTA (0.1 mM EDTA)
m	Milli-
M	Molar (mol/l)
M	Mega-
me	Methylation
$\mu$	Micro-
MC	Media change
MEA	Multielectrode assay
MEFs	Mouse embryonic fibroblasts
MHC	Major histocompatibility complex
min	Minute(s)
minP	Minimal promoter
MIR / miRNA	Micro RNA
MPRA	Massively parallel reporter array
mRNA	Messenger RNA
MT	Mitochondrial
n	Nano-
NBM	Neurobasal media
NBQX	2,3-dioxo-6-nitro-7-sulfamoyl-benzo[f]quinoxaline
NCCs	Neural crest cells
NEB	New England Biolabs GmbH
NES	Nuclear export signal
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NIH	National institute of health (USA)
NLS	Nuclear localization signal
nLuc	Nano luciferase
NMDA	N-methyl-d-aspartate
No.	Number
NPCs	Neuronal progenitor cells
NPM	Neuronal progenitor media
ns	Not significant
NSCs	Neural stem cells
nt	Nucleotide(s)
NT	Neurotransmitter
OCRs	Open chromatin regions
OD	Optical density
OE	Over expression
OMIM	Online mendelian inheritance in man
OR	Odds ratio
oRG	Outer radial glia
ORI	Origin of replication
p	Pico-
PBS	Dulbecco's phosphate buffered saline (Ca2+/Mg2+)
PC(A)	Principal component (analysis)
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PFC	Prefrontal cortex
PM	Patterning media
PNK	Polynucleotide kinase
PNS	Peripheral neural system
PRG	Primary regulated gene
PTHS	Pitt-Hopkins-syndrome (OMIM: 610954)
qPCR	Quantitative polymerase chain reaction

RA	Retinoic acid
RC	RevitaCell
rcf	Relative centrifugal force
Ref.	Reference
RFP	Red fluorescent protein
RG	Radial glia
RIN	RNA integrity number
RISC	RNA-interfering silencing complex
(sc)RNA(-Seq)	(single cell) ribonucleic acid (sequencing)
ROI	Region of interest
RPKM	Reads per kilobase million
rpm	Rounds per minute
rSAP	Recombinant shrimp alkaline phosphatase
RSB	Buffer for ATAC Seq
RSRS	Arginine-Serine-Arginine-Serine motif
RT	Room temperature
RV primer	Reverse primer
s	Second(s)
SCZ	Schizophrenia
SD	Standard deviation
SEM	Standard error of the mean
Seq	Sequencing
shRNA	Short hairpin RNA
SM	Splitting media
SMR	Standard mortality ratio
SNPs	Single nucleotide polymorphisms
SOC	Super optimal broth
STARR-Seq	Self-transcribing active regulatory region sequencing
stimERs	Stimulation dependent active enhancer regions
stimVars	Stimulation dependent active emVars
TBS(-T)	Tris buffered saline (with 0.05 % Tween20)
TE	Tris-HCl and EDTA
TET	Ten-eleven translocation
TET-ON system	Reverse tetracycline-controlled trans activator and TET responsive promoter
TF(BS)	Transcription factor (binding site)
TH	Tyrosine hydroxylase
TPR	True positive rate
TRS	Treatment resistant schizophrenia
TSS	Transcription start site
TTX	Tetrodotoxin
TWAS	Transcriptome-wide association study
U	Units
UCSC	University of California in Santa Cruz
UK	United Kingdom
UMAP	Uniform manifold approximation and projection
UMI	Unique molecular identifier
USA	United States of America
UTR	Untranslated region
V	Volt
Var. Pos.	Variant position
vRG	Ventral radial glia
vst	Variance stabilization
WPRE	Woodchuck hepatitis virus posttranscriptional regulatory element
WT	Wild type

---

## 1 Abstract

Psychiatric diseases are a major public health burden worldwide, yet little is known about their underlying molecular mechanisms. Among those highly polygenic diseases, schizophrenia is characterized by a heritability above 80 %. However, the ability to translate statistically associated variants from genome – wide association studies into insights on disease mechanisms has been severely hampered by the facts that 1) the vast majority of the associated single nucleotide polymorphisms (SNPs) reside in non-coding regions of the genome, 2) the molecular causal genetic variants cannot be annotated among the statistically associated variants in linkage disequilibrium, 3) the variants act cell type and condition specific, and, 4) the effect of those non-coding functional variants and 5) the corresponding target genes of the variants are unknown.

In order to address these questions in a disease relevant human *in vitro* cellular model, several patient and healthy control primary human cells were successfully reprogrammed into induced pluripotent stem cells during the time of this thesis. Immunocytochemistry, bulk and single cell ribonucleic acid sequencing (RNA-Seq) of Neurogenin-2 (*Ngn2*) induced neurons (iNeurons) confirmed excitatory neuronal identity of the peripheral and central nervous system. To pinpoint the effect of schizophrenia associated genetic variants on gene regulation and expression, assay for transposase-accessible chromatin sequencing (ATAC-Seq), single cell and bulk RNA-Seq, massively parallel reporter assay (MPRA), and clustered regularly interspaced short palindromic repeats interference (CRISPRi) analyses were performed on induced pluripotent stem cells (iPSCs) derived iNeurons and primary mouse neurons. High potassium induced depolarization of iNeurons altered the chromatin state of 7,635 regions and the expression levels of 3,055 genes independently of their disease status. These generated data sets of the iNeurons and other publicly available data sets were used to preselect enhancer regions containing schizophrenia associated variants to test for their activity with MPRA in different cell types and states. From the tested regions, 13 % (620) single nucleotide variants were identified to impact the activity of the tested enhancer fragment. Intersection of those allele specific enhancer regions with postmortem brain derived expression quantitative trait loci (eQTLs) and chromatin conformation data sets as well as CRISPRi interference at the *TCF4* gene locus linked several genes per variant and several genetic variants per gene.

---

The annotation of those disease associated functional variants and their linked genes, therefore, offers a comprehensive molecular insight into development and stimulation dependent molecular processes modulated by schizophrenia associated genetic variation. In the future, this strategy can be applied to improve the targeted development of new treatments to attenuate or even abolish progression of other psychiatric and in general complex diseases.

## 2 Introduction

In the following introduction, I provide the information to comprehend the work performed in this thesis to functionally annotate the genetic basis of the psychiatry disease schizophrenia. In the beginning, I introduce the classifications, hallmarks, and impact on society of schizophrenia. Next, I focus on the genetic inheritance of schizophrenia and approaches to investigate the genetic etiology of this devastating disease. Over one hundred genomic loci are associated with schizophrenia and most of the associated variants are located in inter- or intragenic regions which are enriched for gene regulatory elements. Therefore, I present different statistical and lab-based techniques to annotate causal variants and their effect on gene regulation and expression. Gene regulatory activity of genetic regions is indirectly inferred by gene expression and can be directly analyzed by the accessibility of the chromatin. A previously developed massively parallel reporter assay decodes the possible impact of associated genetic variants on gene regulatory activity. Furthermore, I explain in depth this reporter assay and its advantages and limitations.

The second part of the introduction describes in detail the model system used in this thesis to understand the genetic component of schizophrenia in disease relevant cell types. I explain the process and challenges of generating human induced pluripotent stem cells (iPSCs) from primary human cells. In the next section, I present and discuss several approaches to differentiate iPSCs into disease relevant excitatory neurons. Afterwards, I elaborate on the impact and advantage of human iPSCs as a model system to investigate schizophrenia. After introducing the concept of deactivated Cas9 mediated gene regulation, I conclude the introduction with exemplary results from iPSCs studies on the etiology of schizophrenia.

### 2.1 The complex psychiatric disease schizophrenia

#### 2.1.1 Classifications, symptoms, and social economic effects of schizophrenia

Over a hundred years ago, Emil Kraepelin (1856-1926) was the first to report schizophrenia as a separate disease with unknown etiology (Bronner, 1961). Kraepelin divided the people with psychosis without any other cause e.g. infections, into two groups: those suffering from dementia praecox or manic depressive illnesses (Bronner, 1961), which were renamed afterwards to schizophrenia and bipolar disease, respectively. Kraepelin

emphasized the progressive and deteriorating processes taking place in people suffering from schizophrenia. In response to this, Eugen Bleuler (1857-1939) stated that schizophrenia is caused by a disorder of thought and feeling rather than a deterioration of the brain. Bleuler published his book “Dementia praecox oder Gruppe der Schizophrenien” about the categorization of schizophrenia in 1911 (Bleuler, 2011). In this book, Bleuler renamed “dementia praecox”, the disease category developed by Kraepelin, to schizophrenia due to his observation that the symptoms of this disease were a consequence of an associational defect rather than dementia (Shershow et al., 1978). Bleuler established the idea of the 4 A's as hallmarks of schizophrenia: associations, affect, ambivalence, and autistic isolation.

For most of the twentieth century, schizophrenia was seen as a disease affecting the mind rather than the brain. This notion was only changed by the neo-Kraepelinian movement to view schizophrenia, and psychiatric diseases in general, as a treatable medical problem (Shershow et al., 1978). With the implementation of the diagnostic and statistical manual of mental disorders version 3 (DSM-III) in 1980 in the US, delusions and hallucinations highlighted by Kurt Schneider as “first rank symptoms of schizophrenia” became the main markers for the diagnosis of schizophrenia (Schneider, 1959). Consequently, the psychotic aspect of the disease is still the major diagnostic criterion in the current DSM-V. Despite huge efforts to identify objective criteria rather than subjective ratings, schizophrenia is still diagnosed by subjective self-reports and medical questionnaires. Up to date, there are no reliable biomarkers in the blood and measurements such as volume of brain regions or brain connectivity are not predictive (Lai et al., 2016). One reason for the lack of biomarkers, similar to the lack of predictive value of genetic counselling, can be the heterogeneous patient population and a need for subgroup stratification (Lai et al., 2016; Perkovic et al., 2017).

Nowadays, schizophrenia [OMIM:181500] is described as a complex syndrome which is comprised of three symptoms: negative (less than normal), positive (psychotic behaviors not detected in the normal population), and cognitive. People suffering from schizophrenia can, but do not have to, present with abnormal behavior such as self-neglect, mumbling or laughing to themselves, and aimless wandering. Patients can also suffer from hallucinations (the perception of sounds, visions, or touch which are not perceivable by others) and delusions (fixed false beliefs or suspicions which are believed despite evidence). Those

symptoms also include incoherent or irrelevant speech. Moreover, patients can perceive less emotions (apathy and anhedonia). Some patients are displaying an inconsistent body or facial expression compared to their self-reported emotion. Only the persistence of at least two symptoms for at least one month (unless the symptoms are treated) and a total dysfunctional time of more than six months permits the diagnosis of schizophrenia. Additionally, the level of social and or occupational function must be decreased in comparison to before the onset of schizophrenia or general expectations.

Patients do not only suffer from the immediate symptoms of the disease but also from social stigma and discrimination. Suffering from schizophrenia is associated with a considerable decrease in educational and occupational performance. Only 30.3 % of people suffering from schizophrenia were employed (half-time, full-time, unpaid, or supported work) in Germany in 2007; with even lower rates in France (11.5 %) and the United Kingdom (12.9 %) (Marwaha et al., 2007). People suffering from schizophrenia are 2-3 times more likely to die 10-15 years earlier than the general population (White et al., 2009). One of the reasons for the shortened life expectancy are additional physical illnesses, such as cardiovascular, metabolic, and infectious diseases (Tanskanen et al., 2018). However, the biggest driver of the decreased life expectancy is the high suicide rate (standard mortality ratio, SMR, is 6.6 in 2014 in Finland) in people suffering from schizophrenia compared to normal healthy people (Sher & Kahn, 2019; Tanskanen et al., 2018). Reasons for these suicides are not only the psychotic events but also the awareness of the impact of the disease on the quality of life (Sher & Kahn, 2019). In addition to a schizophrenia diagnosis, psychosocial and demographic factors such as disease start at a young age, being male, being unsupported by loved ones, and being intelligent increase the risk of suicide (Sher & Kahn, 2019). All these disabling components of schizophrenia also have a high social economic impact as they generate a total cost (directly and indirectly) of 0.02 to 1.65 % of the gross domestic product (Chong et al., 2016).

### **2.1.2 Inheritance of schizophrenia**

Schizophrenia affects more than 21 million people worldwide according to the World Health Organization. The median incidence of schizophrenia is 15.2/100,000 with a lifetime prevalence of 1 %. More men (60 %) than women (40 %) suffer from this devastating disease (McGrath et al., 2008). In line with the higher prevalence in males, the age of onset

for schizophrenia is earlier in men than in women. Age of onset is during late adolescence and young adulthood at the age of 15–24. However, there is a second peak of onset for women after the menopause (Messias et al., 2007).

The genetic heritability based on twin studies for schizophrenia ranges between 79-87 % (Cardno et al., 1999; Farmer et al., 1987; Hilker et al., 2018; Sullivan et al., 2003). This heritability is considerably lower than and is unlikely to reach the heritability for mendelian diseases (McGuffin et al., 1994). The relative risk of a child with both biological parents affected is 46.9. The relative risk remains increased if only one person of the direct family is affected (Lichtenstein et al., 2009; Mortensen et al., 1999). The proband wise concordance rate of schizophrenia is estimated at 40.8 (95% CI, 26.9 to 54.7) in monozygotic and 5.3 (95 %, CI 0.0 to 11.2) in dizygotic twins (Cardno et al., 1999). Intriguingly, monochorionic twins which shared one placenta have an average concordance rate of 60 %, whereas the concordance rate for dichorionic monozygotic twins drops to 10.7 % (Davis et al., 1995). In line with this, the concordance rate in the offspring of a healthy monozygotic or healthy dizygotic twin differs by over 10 % (17.4 instead of 2.1 %) (Gottesman & Bertelsen, 1989). Unambiguous evidence of a genetic cause for schizophrenia is published for the hemi deletion on chromosome 22q11 which leads to the so-called velocardiofacial or DiGeorge syndrome. This hemi deletion is accompanied in 30 % of the cases with symptoms undistinguishable from schizophrenia and is present in 0.3 % of the patients suffering from schizophrenia (Avramopoulos, 2018; Meyer-Lindenberg, 2010). Several genes are affected by this large copy number variation (CNV), for instance *proline dehydrogenase (PRODH)* and *catechol-O-methyltransferase (COMT)*.

Nevertheless, additional factors such as environmental influences are associated with an increased risk for schizophrenia as well: smoking, alcohol addiction, child abuse (physical and or psychological), mind altering drugs, living in a city, birth in March, and immigration (Leask, 2004; Mortensen et al., 1999). A meta analytic summary of common environmental influences on the liability to schizophrenia is estimated at 11 % (95 % CI, 3-19 %) (Sullivan et al., 2003). In conclusion, schizophrenia disease outbreak is impacted by environmental factors and heavily predetermined by inherited genetics.

### 2.1.3 Linkage analysis and SNP association studies in schizophrenia research

Early on, schizophrenia was recognized as a genetically heritable disease (Kety, 1987; Kety et al., 1994). Hence, over the years, several approaches were followed to investigate the underlying genetic cause of schizophrenia.

The first genome-wide DNA-based approach aimed at linking co-segregated genomic abnormalities of affected families with the phenotypic trait schizophrenia (Ng et al., 2009). De novo and rare CNVs are less common in healthy control (5 %) compared to patients suffering from adult (15 %) or child (20 %) onset schizophrenia (Haraldsson et al., 2011). 2.5 % of the patients and 0.9 % of healthy controls carry a CNV at a schizophrenia associated risk loci (Rees et al., 2014). Similarly, concordant and discordant monozygotic twins display different CNV profiles and epigenetic marks. This explains partly the discordance rate of schizophrenia between monozygotic twins (Haraldsson et al., 2011). CNVs associated with schizophrenia are deletions (e.g. as mentioned beforehand of several 100 kb in the case of the 22q11 deletion syndrome) and duplications of e.g. 16p11.2 or 16p13.1 (Avramopoulos, 2018). Rare recurrent CNVs affect increase disease risk around tenfold while the odds ratio (OR) of common single nucleotide variants ranges around 1.1 (Haraldsson et al., 2011; Meyer-Lindenberg, 2010). So far, the results of follow-up studies or meta-analysis of linkage studies failed to replicate (Badner & Gershon, 2002; DeLisi et al., 2002) or did not pass genome-wide significance level (Ng et al., 2009). The failed replication is partly explained by linkage of uncommon familial genetic alterations resulting from the inclusion of family studies and, thus, population stratification (Harrison & Weinberger, 2005). Additionally, the linkage analysis disregards the non-Mendelian transmission and the polygenic architecture of schizophrenia.

As a consequence, and with the emergence of next generation sequencing strategies, the focus of the genetic analysis of schizophrenia shifted towards an approach that investigates single nucleotide variation between individuals. Single nucleotide variants are located throughout the genome in particular in less conserved regions: intronic, non-coding, and intergenic regions (Hindorff et al., 2009; Huo et al., 2019; Pardiñas et al., 2018). Therefore, these variants affect the regulation of gene expression rather than the function of the genes themselves. This genetic variation is also the basis for non-disease phenotypes like skin, hair, and eye color. Common single nucleotide polymorphisms (SNPs) are taken into account in genome-wide association studies (GWAS) and occur with at least 1 % minor

allele frequency in the general population (Castle, 2011). With the possibility to investigate the occurrence of SNPs between several thousand patients and healthy controls at the same time, GWAS are widely used to analyze complex disease.

GWAS are based on microarrays that genotype 200,000 – 1,000,000 SNPs per individual. Utilizing the public available human genome databases (international HapMap project, 1000 genomes project), it is possible to impute SNPs in close proximity to the genotyped SNPs under the assumption of linkage disequilibrium (LD) between the genotyped and the adjacent imputed SNP (Coleman et al., 2016). LD describes the non-randomly inheritance of two proximal genomic regions that are jointly inherited during meiosis. LD is influenced by the rate of genetic recombination, drift, and the population structure.

The genotyped and imputed SNPs are assessed for their allelic association with schizophrenia. The annotation as a major risk allele or top-hit candidate with the highest p-value in an LD-block or associated loci does not imply a causal function of this variant (Tewhey et al., 2016). The association strength changes between GWASs and depends heavily on ancestral background (Harrison & Weinberger, 2005; S. C. Page et al., 2018). In general, (highly) associated SNPs or risk alleles can be either causative or in LD with a variant that is the actual pathogenic variant. The pathogenic variant resembles the variant which is associated with the disease trait and has a direct functional impact on disease development. Compared to mendelian diseases, the presence of a schizophrenia risk allele or pathogenic variant does not imply a 100 % predisposition to schizophrenia. The calculated effect size of a risk allele is depending, amongst others, on the magnitude of the genetic association, the contribution to the total trait, and the statistical power of the analysis itself (Fullerton & Nurnberger, 2019).

More than 150 loci with several thousand associated SNPs are associated with schizophrenia. Those large scale GWASs are based on more than 60,000 healthy controls and 40,000 individuals suffering from schizophrenia (Z. Li et al., 2017; Pardiñas et al., 2018; Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2014; Trubetskoy et al., 2022). Thousands of additional genetic variants show a strong subthreshold association ( $p - \text{value} > 5 \times 10^{-8}$ ) with schizophrenia and predictions indicate that subthreshold variants can become significant with increasing sample sizes. Furthermore, subthreshold variants were shown previously to be causal of the investigated trait (X. Wang et al., 2016). 70 % of the so far associated SNPs from the publicly available

GWAS data sets for schizophrenia (Z. Li et al., 2017; Pardiñas et al., 2018; Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2014) are located in non-coding regions. Moreover, 40 % of the haplotype blocks with disease associated SNPs do not contain exons (Hindorff et al., 2009; Pardiñas et al., 2018). Hence, follow-up studies are needed to annotate functional variants in associated loci to predict molecular function of the associated SNPs: allelic impact on the activity of gene regulatory elements and expression levels of genes.

#### **2.1.4 Functional annotation of schizophrenia associated SNPs**

GWASs link several thousand SNPs with the complex disease schizophrenia. However, these statistical associations (p-values) neither indicate nor correlate with the functionality of the SNP (Schaid et al., 2018). In order to improve the knowledge of the genetic component of schizophrenia, it is necessary to pinpoint the causal variants and their functional effect(s). Over two third of the schizophrenia associated SNPs are located in non-coding gene regulatory regions and are enriched in neuronal specific open chromatin regions (Girdhar et al., 2018). 132 SNPs are predicted to directly disrupt transcription factor binding sites (TFBS) of 21 transcription factors (TFs) according to ChIP studies for open chromatin and TFs. Some TFBS disrupting SNPs even obstruct the binding sites of several other TFs (Huo et al., 2019). Therefore, a considerable number of approaches have been developed to annotate the allelic impact of SNPs on gene regulation (Kreimer et al., 2017). Other routes by which associated variants could impact levels of gene expression are, but are not limited to, alternative splicing, the stability of mRNA, and protein localization. The regulatory effect of SNPs provides information to predict impacted e.g., signaling or metabolic pathways to eventually discover new treatment opportunities for this devastating disease.

One of the computational based methods to select and prioritize genetic variants for further investigations is to statistically fine – map associated loci. Fine – mapping algorithms determine the genetic functional variant(s) considering LD and haplotype blocks assuming at least one causal variant exists in each loci (Farh et al., 2015; Schaid et al., 2018). Apart from the fact that this method relies purely on correlation, another disadvantage from this statistical approach is, that for most of the loci the predictions are ambiguous and predict several SNPs to be causal (Farh et al., 2015).

An alternative way to annotate functional SNPs is to correlate the presence of the major and minor alleles with gene expression levels. The magnitude of the allelic effect at the expression quantitative loci (eQTL) can significantly differ between tissues and a substantial fraction of eQTL are tissue specific (Farh et al., 2015; Nica et al., 2011). Additionally, eQTL are affected by environmental influences as described after the stimulation of naïve macrophages with IFNy and Salmonella. 60 % of those stimulation specific eQTL are located in regions with stimulation dependent chromatin accessibility (Alasoo et al., 2018). eQTL from a different population can be combined with individual genotype data to individually predict gene expression levels to generate transcriptome-wide association study (TWAS) data sets (Hall et al., 2020). Schizophrenia TWAS associated genes are significantly increased in expression levels during mid-fetal brain development. This further supports the developmental hypothesis of schizophrenia and neurons as the affected cell type by schizophrenia (Gusev et al., 2018; Huo et al., 2019). Furthermore, TWAS genes indicate synaptic abnormalities in schizophrenia (Hall et al., 2020).

However, linkage, association, or correlation cannot replace biological evidence for risk variants and, in general, also lack molecular validation in disease relevant cell types and states (Y. Liu et al., 2017; G. P. Page et al., 2003). Hence, functional validation of the allelic SNP effect on gene regulation is necessary. Gene regulatory loci can be annotated according to the accessibility of epigenetic marks. Open or closed chromatin is referred to as euchromatin or heterochromatin, respectively. Euchromatin can be annotated with DNase I hypersensitive sequencing (DHS-Seq) or with the assay for transposase – accessible chromatin using sequencing (ATAC-Seq). Chromatin immunoprecipitation (ChIP) indicates the presence of genomic DNA (gDNA) binding proteins such as transcription factors or post-translational histone modifications such as methylation (me) or acetylation (ac). The histone modifications H3K4me1/3, and H3K9ac and H3K27ac are indicative for accessible promoters and enhancers, respectively (Klemm et al., 2019). Chromatin accessibility quantitative trait loci (caQTL) are regions which are impacted in their accessibility in the presence of the alternative allele. caQTL analysis, however, requires several hundred samples of the same cell type or tissue to achieve sufficient power to investigate SNPs with minor allele frequencies below 5 %.

In order to circumvent this population-based underrepresentation of common variants, putative enhancer regions (pERs) of interest are individually or in parallel assessed for their

capability to regulate gene expression. For this purpose, putative enhancer fragments (pEFs) containing the reference or alternative allele are synthetically synthesized. In general, reporter assays analyze the ability to recruit transcription factors and RNA polymerase II to the gene regulatory element (GRE) of interest. Therefore, this method only detects those genetic variants that exert their disease modulatory effects by altering the activity of GREs. Epigenetic alterations such as (hydroxy)methylation or carboxylation of CpG sites are not modelled in the current high throughput massively parallel reporter assays (MPRAs) even though these modifications impact TF binding strength (Yang et al., 2019; Ziller et al., 2018). Hence, current MPRAs are only capable to annotate a fraction of the disease modulating SNPs which modulate gene expression in *cis* (Yu et al., 2016). Reporter assays compared to caQTL studies lack the context of genomic DNA (Cheng et al., 2014) and investigate synthesized oligos with a fragment length of around 160 bp even though mammalian enhancers can span around 1 kilobase (kb) (Patwardhan et al., 2012). In order to increase the length of the investigated ERs, longer enhancer fragments can be captured either from fragmented gDNA, cloned from bacteria artificial chromosome libraries, synthesized with a lower throughput, or elongated by multiplex pairwise assembly of small synthesized fragments (Klein et al., 2020; Muerdter et al., 2018; van Arensbergen et al., 2019; Vanhille et al., 2015). However, longer fragment sizes (354 or 678 nt) correlate well (0.78) with the activity of their shorter (192 nt) counterparts (Klein et al., 2020; Muerdter et al., 2018; Vanhille et al., 2015). Furthermore, as a result of the fragmentation independent of the length of the construct, these reporter assays cannot investigate altered TF binding independent of the primary nucleotide sequence of the TFBS motif (Kheradpour et al., 2013; Lu et al., 2021; Siggers & Gordân, 2014).

Traditional approaches of reporter assays locate the GRE 5' of a minimal promoter (minP). The minP drives the expression of a quantifiable reporter gene such as fluorescent proteins (e.g., GFP and derivatives) or enzymes (luciferases and derivatives). The activity range of the minP system is suitable to detect enhancing but not repressing effects (Doni Jayavelu et al., 2020; Tewhey et al., 2016). If the pEF is located in the 3' end of the reporter gene it induces the expression of itself alongside with the reporter and is, therefore, referred to as self-transcribing active regulatory sequencing (STARR-Seq). The advantage of STARR-Seq is that the GRE can interact in a 3D space with the minimal promoter partially mimicking enhancer-promoter loops (Melnikov et al., 2012, 2014; Ulirsch et al., 2016). Nevertheless,

utilizing a minimal promoter, the classic assays cannot decipher the promoter sequence or type specific activity of enhancer elements (Martinez-Ara et al., 2022; Zabidi et al., 2015).

pEFs of interest are either transfected, nucleofected, or transduced into cell lines, primary cells, or tissues of interest. Lentiviral integration of the construct increases the reproducibility and decreases the dynamic range compared to episomal based reporter assays (Inoue et al., 2017; Klein et al., 2020). The fold change between allelic enhancer activity for the majority of the tested enhancer fragments is low (Lu et al., 2021; Melnikov et al., 2014; Ulirsch et al., 2016). Therefore, the MPRA is required to be sufficiently sensitive to reliably detect such small effects. In order to increase the power and sensitivity of the analysis, each enhancer fragment is tagged with several barcodes. In general, barcodes are located at the 3' of the reporter gene and 5' of the polyA tail (Klein et al., 2020). The analysis is performed on purified (m)RNA and utilizes the abundance of unique barcodes as a measurement of the activity of the pEFs with next generation sequencing (NGS). Barcoding of the enhancer reduces sequencing costs and increases the signal to noise ratio (Lu et al., 2021; Tewhey et al., 2016).

In order to compare SNP effects on GRE activity, both the reference and the alternative allele are analyzed, and their activities are compared to detect expression modulating variants (emVars). EmVars are highly cell type and condition specific with only one fifth of the emVars shared between K562 and HepG2 cells (van Arensbergen et al., 2019). Similarly, 9 emVars out of 1,049 schizophrenia associated SNPs are shared between K562 (148 emVars) and SKY-5Y cells (53 emVars). In total, 14 % or 5 % of the tested SNPs are emVars in K562 or SK-5Y cells accounting for an average of 2.6 emVars per investigated loci, respectively (Myint et al., 2020). Likewise, only 52 % of active ERs are shared between HepG2 cells and an EBV-transformed lymphoblastoid B-cell line (Tewhey et al., 2016). Active ERs are enriched for colocalization with open chromatin marks: DHS-Seq, TF- and open ChIP-Seq (Tewhey et al., 2016). The functional impact of SNPs and single base pair CNVs is modest with 22 % of the investigated SNPs impacting gene regulating activity by > 1.2-fold and only 3 % by > 2-fold in an *in vivo* MPRA screen for liver enhancers in mouse liver (Patwardhan et al., 2012). These alterations in the activity of several enhancers moderately affect the expression of several genes. These modest alterations can then dysregulate of pathways, metabolic homeostasis, and signaling cascades impacting in their combination the functionality of certain cell types. Therefore, the impact of schizophrenia

associated variants on gene regulation needs to be investigated in disease relevant (human) cell types.

## 2.2 Human induced pluripotent stem cells in schizophrenia research

In order to investigate the genetic cause of schizophrenia in disease relevant *in vitro* cellular models, human induced pluripotent stem cells (iPSCs) derived from primary human cells can be utilized. Those iPSCs are differentiated into disease relevant cells such as neuronal progenitor cells and neurons.

### 2.2.1 Generation of human induced pluripotent stem cells from primary cells

Over a decade ago, Takashi et al. published the groundbreaking report that primary cells from mouse (Takahashi & Yamanaka, 2006) as well as humans (Takahashi et al., 2007) can be transformed into a state resembling the characteristics of embryonic stem cells (Wernig et al., 2007): induced pluripotent stem cells (iPSCs) (Takahashi & Yamanaka, 2006). The reprogramming is accomplished by the artificial overexpression of transcription factors (Yamanaka factors) that induce the pluripotent state: *POU5F1* (also known as *OCT3/4*), *SOX2*, *MYC*, and *KLF4* (Takahashi et al., 2007). During the reprogramming process, the epigenetic landscape reshapes, the telomerase activity is restored, and the cells revert to the pluripotent state while maintaining their genetic background (Bock et al., 2011; Marión et al., 2009; Wernig et al., 2007). iPSCs, like embryonic stem cells, have the potential to generate cells from all three different germ layers: endo-, meso-, and ectoderm. After reprogramming, this capability can be verified *in vivo* by the teratoma assay in immunocompromised mice or *in vitro* by embryoid body (EB) formation (Baghbaderani et al., 2016; Bock et al., 2011; Sheridan et al., 2012).

However, the overexpression of the Yamanaka factors increases the oxidative stress (Banito et al., 2009), while increased c-MYC and KLF4 induce TP53 overexpression. TP53 is involved in cell cycle arrest in G1 phase and induction of senescence and cell apoptosis (Banito et al., 2009; Beyfuss & Hood, 2018; D. Liu & Xu, 2011; Spike & Wahl, 2011). Senescence does not only arrest cells in their cell cycle, but also dramatically changes their transcriptomic levels (Goldstein, 1990; Zorin et al., 2017). The combination of high oxidative stress and high levels of TP53 minimizes the number of cells surviving the

reprogramming process (Beyfuss & Hood, 2018; D. Liu & Xu, 2011). In line with this, the presence of DNA damage such as short telomeres, DNA repair deficiencies, or exogenously inflicted DNA damage aborts the reprogramming process by the activation of DNA damage response pathways and TP53-dependent apoptosis (Marión et al., 2009; Spike & Wahl, 2011). To circumvent TP53 induced apoptosis, transient decrease of TP53 activity can be accomplished by shRNA mediated RNA decay and the TP53 inhibitor Pifithrin- $\alpha$  (PFT $\alpha$ ) (Komarova et al., 2003). This reduction in TP53 levels during reprogramming mimics the low TP53 expression in ESCs and iPSCs (Spike & Wahl, 2011; Takahashi & Yamanaka, 2006) and increases reprogramming efficiency (10-fold in human fibroblasts) (Banito et al., 2009; Ebrahimi, 2015; Marión et al., 2009; Rasmussen et al., 2014). However, knockdown of p53 increases DNA damage, chromosomal end to end fusions (6-fold), and DNA double strand breaks or fragmentations (6 fold) compared to wild type (WT) mouse embryonic fibroblasts (MEFs) (Marión et al., 2009). Chromosomal translocations can be visualized by Giemsa banding, however, the detection limit of translocations and CNVs is 6 Mbp. Analysis of the genome integrity of the reprogrammed cell lines is necessary, due to the fact that the iPSCs are used to analyze the genetic component of schizophrenia. SNP based microarrays to analyze the genome integrity offer the advantage that the detection limit of CNVs is decreased to 0.2 Mbp. Nevertheless, both methods fail to detect single coding mutations which occur during the reprogramming on average in five positions (Gore et al., 2011). Non-integrating instead of lenti- or retroviral reprogramming approaches can further reduce the number of changes in the genomic background of the iPSCs (X. Kang et al., 2015).

The efficiency of the reprogramming process can be improved by changing the delivery method of the coding sequences (lentiviral, mRNA transfection, episomal), and the media composition (Ebrahimi, 2015; Mali et al., 2010; Miller et al., 2013; Z. Zhang et al., 2014). Epigenetic rearrangements are facilitated by inhibiting the methyltransferase DOT1L (inhibitor: EPZ-4777) (C. Chen et al., 2016) as well as histone deacetylases (HDACs; sodium butyrate, NaB) (Mali et al., 2010; Z. Zhang et al., 2014) during the reprogramming. The addition of ascorbic acid to the reprogramming media mitigates the impact of cell senescence and promotes gene expression changes to facilitate transition from pre to full iPSC stage (De Los Angeles et al., 2015; Esteban et al., 2010). Recent efforts in the reprogramming field have further optimized the procedures for several primary human cell types, for instance dermal or embryonic fibroblasts, peripheral blood cells (CD34+ cells, T-

and B- lymphocytes), keratinocytes from the root of hair, and pancreatic islet beta cells (Beers et al., 2015; Chou et al., 2015; Diecke et al., 2014; Raab et al., 2014; Schlaeger et al., 2015; Stadtfeld et al., 2008; Takahashi et al., 2007). These improvements offer the possibility to generate human iPSCs in parallel from several individuals in less than two months to investigate human complex diseases in a human cellular model system.

### 2.2.2 Neural differentiations of pluripotent stem cells

iPSCs derived from patients suffering from schizophrenia inherit the same genetic background as the primary cells taken from the donors. Therefore, iPSCs offer the unique possibility to analyze the genetic impact on disease development in disease relevant human cell types. This circumvents the extrapolation from human peripheral tissue such as blood to disease relevant cell types such as excitatory neurons. Several protocols are published that successfully generate different neuron types from iPSCs: forebrain cortical glutamatergic, basal forebrain cholinergic, midbrain dopaminergic, striatal medium spiny, GABAergic, and hindbrain serotonergic (Y.-N. Liu et al., 2017). These differentiation protocols mimic the differential temporal and spatial expression of morphogens during embryonic brain development. Morphogen concentration and presence changes alongside the anterior-posterior axis (APA) and dorsal-ventral axis (DVA) (Y.-N. Liu et al., 2017). The main morphogens in patterning the APA are fibroblast growth factors, retinoic acid and inhibition or activation of the Wingless and Int-1 (Wnt) pathway. In particular, high activity of the Wnt-pathway determines the differentiation into hindbrain and spinal cord neurons whereas decreasing activity of the Wnt pathway leads to the differentiation of stem cells into midbrain (medium dose) and forebrain (low dose) neurons (Kirkeby et al., 2012). Remarkably, high activity in the Wnt signaling pathway is also important for caudalization of the neurons (Xi et al., 2012). The inhibition of the Wnt pathway in combination with an inhibition of the bone morphogenetic protein (BMP) pathway conveys differentiation into the dorsal fate; whereas activation of the sonic hedgehog (SHH) pathway promotes the ventral fate (Kiecker & Lumsden, 2012). Addition of patterning factors during the differentiation process increases the purity in the neuronal cultures. 2D differentiation in the presence of dual SMAD inhibition deactivates BMP and TGF $\beta$  signaling with Noggin or LDN193189 and SB431542 or A83-01, respectively. iPSCs differentiated with these small molecule inhibitors generate neuronal progenitor cells (NPCs) that are able to form PAX6

positive rosettes in 2D culture systems (Chambers et al., 2009). Further improvements to the dual SMAD protocol to enrich for glutamatergic excitatory cortical neurons include additional Wnt inhibition (Mariani et al., 2012; Maroof et al., 2013) or the addition of RA (Shi et al., 2012).

Early protocols use embryoid bodies or spheres, rosette formation, or neuronal progenitor cells to differentiate iPSCs into neurons. This approach, though, can generate heterogeneous cultures that contain glutamatergic and GABAergic neurons as well as glia like cells (Brennand et al., 2015). To enrich for a homogeneous cell population, direct conversion of cells (iPSCs or fibroblasts) into neurons utilizes over expression (OE) of neuronal transcription factors. The induction of excitatory forebrain neurons is performed by OE of mouse *Ngn2* under the control of a tetracycline responsive promoter (Y. Zhang et al., 2013). Dual SMAD and Wnt inhibitors pattern the generated neurons to cortical layer excitatory forebrain neurons (Nehme et al., 2018). The addition of patterning factors for the cortical excitatory fate in the beginning of the differentiation is especially important since *NEUROG2* / *Ngn2* is expressed in several other neuronal areas of the (mouse) nervous system during development. In the spinal cord, *Ngn2* interacts with *ISL1* and *LHX3* to generate motor neurons (Velasco et al., 2017). *Ngn2* expression is also observed during the differentiation of neural crest cells (NCCs) into preferentially sensory rather than sympathetic ganglia. In the dorsal root ganglion (DRG), *Ngn2*<sup>+</sup>/*Wnt1*<sup>+</sup> cells generate peripheral glia and *Ngn2*<sup>+</sup>/*ISL1*<sup>+</sup> neurons further differentiate into nociceptive -, mechanoceptive -, and proprioceptive sensory neurons (Zirlinger et al., 2002). In peripheral neurons, *Ngn2* acts in combination with *POU4F1* to generate sensory neurons. Moreover, *POU4F1* and *Ngn2* overexpression can be used to differentiate iPSC into mechano- and cold-sensitive peripheral neurons of the dorsal root ganglion (Hulme et al., 2020; Nickolls et al., 2020). DRG neurons, with other peripheral neurons, are reported to be *PRPH* positive. *PRPH* protein expression upon *NGN2* induced unpatterned differentiation is reported to be present in 10 % of the generated iNeurons (M. Chen et al., 2020). Very high levels of *PRPH* are reported in bulk and in the majority of clusters (5 of 6) of similarly differentiated neurons in a scRNA-Seq experiments (M. Chen et al., 2020; Lin et al., 2020). *Ngn2* also induces dopaminergic midbrain neuronal fate if co - expressed with *Nurr1* (*NR4A2*) (Andersson et al., 2007).

The homogeneity of the culture can be increased by depleting dividing cells with cytosine arabinoside (AraC) and, hence, enrich the culture for postmitotic neurons (Meijer et al., 2019). In order to facilitate the maturation of the differentiated neurons, neurons are cocultured with murine astrocytes (Burke et al., 2020; Muratore et al., 2014; Nehme et al., 2018; Tang et al., 2013). For the astrocyte survival, the culture media is supplemented with 1 % FCS (Meijer et al., 2019). Additionally, expression of Ngn2 can be terminated during the differentiation protocol (as early as 1-5 days after the start of the differentiation), following the early time course of Ngn2 activity in early immature neurons during development (Lin et al., 2020; Meijer et al., 2019). The first published differentiation protocols, however, sustain Ngn2 overexpression until cells are harvested (Nehme et al., 2018; Y. Zhang et al., 2013).

The efficiency and purity of the differentiation paradigm can be estimated with marker genes and proteins. Dorsal differentiations generate predominantly excitatory cortical neurons, which express among others *CTIP2*, *SATB2*, *BRN2*, *CUX1*, *CUX2*, *FOXG1*, and *VGLUT1/2*. Spinal cord and hindbrain neurons are positive for *HOXB4*, *LIM1/2*, and *LIM3*. Nevertheless, marker genes are in most cases indicative of several (neuronal) cell types and only the combination of markers can clearly identify the investigated cell type. Likewise, the well-known cortical layer identity markers *BCL11B*, *POU3F2*, *POU3F1*, *CUX1*, *CUX2*, and *SATB2* are additionally present in neurons in different areas of the central and peripheral nervous system: the spinal cord, peripheral ganglia, diencephalon, mesencephalon, and hindbrain (Hawrylycz et al., 2012; Zeisel et al., 2018). Non-neuronal cells also express specific marker proteins which can be investigated to ensure the purity of the neuronal culture: astrocytes are positive for *S100B* and *glial fibrillary acidic protein (GFAP)*, microglia for *IBA1* and *CD11B*, and *oligodendrocytes for myelin basic protein (MBP)*, *NG2*, *myelin oligodendrocyte glycoprotein (MOG)* and *SOX10* (M. Chen et al., 2020; Hawrylycz et al., 2012; Loo et al., 2019; Ostermann et al., 2019; Zeisel et al., 2018).

### 2.2.3 iPSC derived neural cells as an *in vitro* model system for schizophrenia

There are several model systems to investigate the molecular basis of neurological diseases such as schizophrenia with their respective advantages and disadvantages. This thesis focuses on the use of induced pluripotent stem cells and derived neural cultures to investigate schizophrenia. Similar to other *in vitro* model systems, iPSCs neither experience

nor respond to complex environmental stimuli with complex behavioral changes. Additionally, interactions of genetic variations with the environment are neither detected nor modelled in *in vitro* model systems (Børglum et al., 2014). This is especially important to consider for schizophrenia due to the fact that it is known from twin studies that environmental factors play a role on the likelihood to develop schizophrenia. Nevertheless, modelling environmental and behavioral aspects of the disease in *in vivo* (mouse) studies completely disregards the genetic component of schizophrenia. Moreover, the main mouse models for schizophrenia possess the disadvantage to mainly resemble the positive symptoms of schizophrenia (Wilson & Terry, 2010). In addition, schizophrenia like phenotypes are induced in healthy mice either by drugs, lesions, genetic manipulations such as gene knock outs, or social stress and can, therefore, hardly mimic the complex genetic component of schizophrenia development (Jones et al., 2011).

Hence, the most undeniable reason to favor human cells or tissue is that it is possible to directly decipher the impact of the diseased human genome on the development of schizophrenia. This offers the unique possibility to relate the genotype and clinical aspects of the individual donor to the corresponding *in vitro* endophenotype (Ahmad et al., 2018). The use of human fetal tissue for functional studies harbors undeniable ethical and availability issues (Boer, 1994; de Wert et al., 2002). Besides, there is no record about the cognitive health of the fetus in contrast with postmortem or primary tissue. Primary human brain tissue is nearly impossible to obtain due to the fact that neurons are postmitotic and biopsies damage the brain tissue. The availability of postmortem brain tissue is limited and it is difficult to segregate the effects of the disease itself from the effects of disease progression, coping mechanisms of individual cells or even brain regions, drug treatments, and life history in general (Chan et al., 2011; Powchik et al., 1998). Furthermore, the length of antipsychotic medication treatment is strongly correlated with epigenetic alterations such as the methylation of MEK1 (Mill et al., 2008). Primary human non-neuronal tissue such as blood can only be used to extrapolate molecular function in disease related cell types. However, there are undeniable significant differences between the epigenome and transcriptome of different tissues or cell types.

Additionally, the iPSC-based approach offers the exciting potential for in depth family studies even from samples cryopreserved in biobanks for several years (Hoffman et al., 2017). Another advantage of the iPSCs model system over primary human material is, that

the material is not limited since iPSCs can be propagated and expanded. iPSCs can generate multiple disease relevant cell types which can be analyzed separately and in context with each other in 2D coculture systems (McPhie et al., 2018; Santos et al., 2017; TCW et al., 2017). In addition, iPSCs are able to form self-assembled 3D cerebral organoids (Lancaster et al., 2013) closely resembling the composition and 3D arrangement of fetal brain tissue (Luo et al., 2016). Thus, this is also one of the biggest drawbacks of the iPSC generated neurons, due to the fact that they resemble the maturation state of late fetal neurons (Nehme et al., 2018). Nevertheless, those neurons display synaptic contacts and are electrophysiological active: presence of action potentials, burst firing, network activity, and spontaneous miniature post synaptic potentials (Nehme et al., 2018; Y. Zhang et al., 2013). The immaturity of these cell types, however, also offers the great possibility to investigate the underlying mechanisms leading to the development of schizophrenia following the hypothesis of a neurodevelopmental component for disease development (E McKinney, 2014). The differentiation process of iPSCs to neurons *in vitro* mimics closely the developmental stages from early embryogenesis to late fetal neurons *in vivo* (Bennett et al., 2019). Therefore, critical developmental steps can be monitored throughout the differentiation process *in vitro*.

Similar to immortalized cell lines, iPSCs can be genetically engineered. The insertion of DNA fragments into the genome can be achieved with e.g., lentiviruses, zinc fingers, transcription activator-like effector nucleases (TALENs) or Cas9 nucleases. Zinc fingers, TALENs, and Cas9 nucleases produce double strand breaks in the gDNA which trigger host gDNA repair via non-homologous end joining (NHEJ), homology-directed repair (HDR), or a combination of these pathways (Adli, 2018; Xu & Qi, 2019). HDR can be used to specifically edit the genome according to a template with the desired exchange sequence allowing to insert genes or exchange SNPs. Targeted genetic engineering of iPSCs is a unique opportunity to compare isogenic cell lines from the same donor. The specificity of the Cas9 system (CRISPR; clustered regularly interspaced short palindromic repeats), is defined by the sequence of the guide RNA (gRNA). The Cas9 approach is also repurposed to alter the activity of GREs with CRISPR activation / interference (CRISPRa/i). For this approach, nuclease-dead Cas9 (dCas9) is fused to activating (VP64, p300) or inhibiting (KRAB, LSD1) effector proteins which open or close chromatin regions, respectively (Adli, 2018; L. S. Qi et al., 2013; Xu & Qi, 2019). The effector protein KRAB recruits methyl transferases

(DNMT3A, B) to gDNA. 5'-methylation of cytosines of the gDNA at CpG sites in promoter or distal regulatory regions represses transcriptional activation. Demethylation of those sites is performed by ten-eleven translocation (TET) proteins (TET1, TET2, and TET3) which can also be fused to dCas9. dCas9-effector can, furthermore, impact the state of the chromatin by posttranslational modification of histones. Fusion of LSD1 to dCas9 promotes histone methylation and, thus, deacetylation. On the other hand, de novo H3K27 acetylation by the dCas9-p300 fusion protein can induce transcriptional activation at GREs and promoters (Xu & Qi, 2019). This approach can be used to decrease gene expression when targeted to promoter regions to analyze the effect of decreased gene levels but also to experimentally link enhancer regions to gene expression.

In conclusion, iPSCs are a promising model system to explore molecular and cellular endophenotypes in relevant cell types of the complex psychiatric disease schizophrenia. This approach offers the unique possibility for the analysis of transcriptomics and epigenomics in case-control studies.

#### **2.2.4 Schizophrenia induced transcriptional and epigenetic alterations in iPSC derived neurons**

iPSC derived neurons are a valuable *in vitro* model system to investigate schizophrenia. A multitude of iPSC studies compare endophenotypes between healthy control and schizophrenia patient derived iPSC lines. This chapter of the introduction, due to the focus of this thesis, solely reports the results of previous case-control studies centered around transcriptomics and epigenetics.

In order to understand the impact of schizophrenia associated genetic aberrations on regulation of gene expression, ATAC-Seq can be used. Healthy control iPSC derived neurons differentiated *in vitro* for 6 weeks are enriched for open chromatin regions (OCRs) in proximity to genes regulating axon guidance, neuron migration, and synaptic transmission (Forrest et al., 2017). Those neuronal OCRs are specifically enriched for GWAS loci associated with schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2014). This enrichment supports mature neurons as a very important model system for schizophrenia (Forrest et al., 2017). Therefore, OCRs are compared to analyze differences in gene regulation between healthy controls and schizophrenia patient

derived neurons. The chromatin accessibility of 70 to 80 % of allele specific OCRs correlates with transcription factor binding strength in dependence of the allelic effect of the SNP. This correlation indicates that genetic variants in allele specific OCRs alter TF binding (S. Zhang et al., 2020). Isogenic proof of concept for impact of SNPs on accessibility of OCRs is published for the microRNA locus MIR137. At this location the presence of the alternative allele not only changes the accessibility of the chromatin but also affects MIR137 expression (Forrest et al., 2017). The effect size of several allele specific OCRs is proportional to the maturity of the generated neural cell types (NPCs vs 20 day of iNeuron differentiation) (S. Zhang et al., 2020).

An alternative approach to understand effects of schizophrenia is to investigate the transcriptome of disease relevant cell types. A neuronal differentiation study of four healthy control and four schizophrenia patient derived iPSCs reports 1,669 differentially expressed genes (DEGs) (Roussos et al., 2016). Likewise, in a set of 10 iPSC derived neurons (four childhood onset schizophrenia patients (COS), six healthy control), nearly 600 DEGs are detected with most of them displaying 1.3-fold change between the two cohorts. Pathway analysis implicates glutamate and Wnt signaling differences between case and control neurons. In general, the majority of DEGs in iPSC derived neuronal case-control studies are associated with synaptic transmission (Brennand et al., 2011, 2015). It is important, though, to take into consideration that the analyzed samples contain a varying degree of neuronal cell types: glutamatergic (~ 60 %), GABAergic (~ 30 %), and dopaminergic neurons (~ 10 % tyrosine hydroxylase (TH) positive). This cell heterogeneity of the differentiation protocols impedes the validity of the DEG set. Therefore, it is unclear, whether the reported associated pathways, are due to the difference in cell type composition or due to the difference between schizophrenia and healthy control neurons (Brennand et al., 2011). In a follow up study 36 and 40 independent cell lines from 10 unique COS and 9 unique healthy control individuals are differentiated to neurons via dual SMAD inhibition, respectively. After correction for the above mentioned cell type heterogeneity, only one gene remains significantly different with an FDR < 10 % (Hoffman et al., 2017). Likewise, a small study (n = 8) based on homogeneous cortical interneurons detected 5 DEGs (p.adj < 0.05) (Shao et al., 2019). This is in line with reports from postmortem dorsolateral prefrontal cortex (DLPFC) tissue from over 500 samples. For those postmortem brain samples, the cell types are evenly distributed in the bulk experiments

between healthy control and schizophrenia cases and over 600 DEGs are detected (Fromer et al., 2016).

The widely reported synaptic phenotype, and its amelioration by antipsychotics, indicates the need to investigate schizophrenia related transcriptional alterations in an activity dependent manner (Brennand et al., 2011, 2015; Grunwald et al., 2019; Kathuria et al., 2019). Furthermore, schizophrenia associated variants are especially enriched in enhancer, which decommission due to neuronal activity (Beagan et al., 2020). In active neuronal networks, calcium ions ( $\text{Ca}^{2+}$ ) enter neurons after action potential dependent induced depolarization of the post synapse through ion channels such as N-methyl-d-aspartate (NMDA) and  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors, and voltage gated calcium channels.  $\text{Ca}^{2+}$  is a second messenger that binds to  $\text{Ca}^{2+}$ /calmodulin-dependent protein kinase 2 (CAMK2). The activated CAMK2 activates CREB (cAMP response element-binding protein) signaling. CREB is a transcription factor that after being phosphorylated binds to CREs nearby transcription start sites (TSS) but also in exonic, intronic, and intergenic regions (West et al., 2001; Xinmin Zhang et al., 2005). Due to the stimulation, the abundance of H3K27ac sites increases primarily at locations previously annotated as H3K4me1 sites in primary neurons (Malik et al., 2014).

One approach to mimic such neuronal activity dependent depolarization in *in vitro* cell culture systems is to depolarize the neurons with high potassium chloride (hKCl) concentrations forcing positive charged calcium ions to enter the neurons through L-type voltage-sensitive calcium channels (Malik et al., 2014). Studies in primary murine neurons report time dependent alterations in transcription. Immediate early genes are induced after relatively short periods (10 s to 5 min of stimulation) with a peak expression shortly after 30 min. Immediate early genes such as *Fos* and *Egr1* as well as the above-mentioned *CREB* are well-known regulators of activity-dependent transcription (Malik et al., 2014; Sheng & Greenberg, 1990; Tyssowski et al., 2018). Delayed genes increase in expression or are induced hours after sustained depolarization (Tyssowski et al., 2018). Activity dependent gene regulation is conserved between species. In particular, the directionality of the change in gene expression is conserved between mouse and human neurons. However, the comparison of mouse primary, mouse ESCs, and human ESCs differentiated neurons denotes a human specific activity dependent gene regulation. The difference between mouse and human partly originates from divergent sequences from enhancer and

promoter regions for some of the genes (Qiu et al., 2016). This highly supports the need of cell models with a human genetic background to investigate human diseases. One study in iPSC derived neurons reports 1,199 DEGs after hKCl induced depolarization. Strikingly, only 59 genes are differentially expressed in the schizophrenia cohort ( $n = 4$ ) compared to 594 differentially expressed genes in the healthy control cohort ( $n = 4$ ) (Roussos et al., 2016).

In order to combine both types of data sets, chromatin conformation data, for instance HiC, can be used to link differential (allele specific) OCRs to neighboring genes. Thus, changes in OCRs can be used to predict changes in linked gene expression. Chromatin conformation sites do not only change upon activity but are altered throughout the differentiation from NPCs to neurons. Long range contacts between distal regions of the genome are more abundant in neurons than in NPCs. Those long-range contacts are especially enriched in genes important for chromatin remodeling and synaptic proteins. Interestingly, activity induced long range enhancer regions are enriched for schizophrenia associated SNPs (Beagan et al., 2020). Furthermore, chromatin interaction loci from neurons are enriched for schizophrenia loci compared to chromatin interaction loci from astrocytes (Rajarajan et al., 2018).

## 2.3 Aims of this thesis

The here presented work is motivated by one of the big unmet needs in biomedical science: to improve the functional annotation of the genetic component of complex diseases such as schizophrenia. Despite the vast research into the molecular mechanisms of schizophrenia little is known about this highly heritable and devastating disease. The overarching objective of this thesis was to annotate the directionality of the alteration in gene regulatory activity introduced by schizophrenia associated SNPs in dependence of cell types and states.

Due to the fact that gene regulatory activity is cell type dependent, these questions are addressed in a disease relevant human *in vitro* neural model. Therefore, one aim of this thesis was to reprogram several patient derived primary human cells into iPSCs. Successful reprogramming was monitored in order to verify the pluripotency state and the genetic integrity. Furthermore, the neurons derived from these iPSCs were inspected for their neuronal identity with immunocytochemistry, and bulk and single cell RNA sequencing. The

generation of a homogeneous neuronal culture by Ngn2 overexpression is particularly important for the here performed case-control studies. The establishment of the iNeuron differentiation protocol and the characterization of the iNeurons were part of this thesis.

In this dissertation, neurons were differentiated from the here reported (S1 – S5) and additional cell lines (C1-C6, S6-S8) to investigate the open chromatin landscape (ATAC-Seq) and the transcriptome (bulk RNA-Seq) of iPSC derived neurons (iNeurons). Additionally, this dissertation aimed to determine the impact of schizophrenia on high potassium induced depolarization of iNeurons. This depolarization of the iNeurons, and other stimuli on the primary mouse neurons solely used for the massively reporter assay (MPRA) experiments, is used to mimic activity induced depolarization of neurons.

These newly generated data sets of the iNeurons and other publicly available data sets were used to preselect enhancer regions tested by MPRA. MPRA identifies those investigated regions which are active and harbor an allelic effect on enhancer activity in a cell type and condition specific manner. I will present the incorporation of those allele specific enhancer regions with postmortem brain derived eQTLs and HiC data sets to link allele specific emVars to genes. Furthermore, I will apply the well-known CRISPRi technique to investigate gene regulation at the schizophrenia associated *TCF4* locus and verify the possibility of performing successfully a CRISPRi based screen in iNeurons. Taken together, this dissertation aims at a better understanding of schizophrenia genetics in order to enable the development of novel treatments in the future.

### 3 Materials and Methods

#### 3.1 Materials

##### 3.1.1 Cell Culture Plastics

Name	Catalog number	Manufacturer
10 cm dish	93100	TPP
15 cm dish	93150	TPP
15 cm dish	353025	BD Falcon
60MEA200/30iR-Ti-gr	60MEA200/30iR-Ti	Multichannel systems
Acrodisc Pes 0.45um 32mm steril	514-4133	VWR
Corning® 60 mm Ultra-Low Attachment Culture Dish	3261	Corning
Corning® 100 mm TC-treated Culture Dish	430167	Corning
Corning® Costar® TC-Treated Multiple Well Plates size 6 well	3516	Corning
Corning® Costar® TC-Treated Multiple Well Plates size 12well	3515	Corning
Corning® Costar® TC-Treated Multiple Well Plates size 24 well	3526	Corning
Corning® Costar® TC-Treated Multiple Well Plates size 48 well	3548	Corning
LUNA™ Cell Counting Slides, 1000 Slides	L12003	Logos Biosystems
Pre-Separation Filters (70µm)	130-095-823	Miltenyi Biotech
SepMate™-50 columns	RUO, 86450	StemCells Technologies
T-75 (Tissue culture flask)	90076	TPP

##### 3.1.2 Cell culture reagents

Name	Catalog number	Company / Vendor
(-)-Bicuculline methiodide	ab120108	Abcam
2-Mercaptoethanol (50 mM)	31350010	ThermoFisher Scientific
4-Aminopyridine (4-AP)	ab120122	Abcam
5x PEG-it	LV825A-1-SBI	SystemBio
Accutase	A6964-100ml	Merck
AlbuMAX™ I Lipid-Rich BSA	11020021	ThermoFisher Scientific
Attachment Factor Protein (1X)-100ml	S006100	ThermoFisher Scientific
B-27™ Supplement (50X)	17504044	ThermoFisher Scientific
B-27™ Supplement (50X), minus vitamin A	A12587010	ThermoFisher Scientific
BDNF (Brain derived neurotrophic factor)	130-096-286	Miltenyi Biotech
BrainPhys™ Neuronal Medium	5790	Stemcells Technologie
CNTF (Ciliary neurotrophic factor)	130-096-336	Miltenyi Biotech
Corning® Matrigel® hESC-Qualified Matrix, *LDEV-free, 5 mL	356278	Corning
Cytarabine	C3350000	Merck
D-AP5	0106	R&D Systems
DMEM (Dulbecco's Modified Eagle Medium, high glucose, pyruvate)	41966029	ThermoFisher Scientific
DMEM/F-12, HEPES	11330032	ThermoFisher Scientific
DAPT	130-110-489	Miltenyi Biotech
DMSO (Dimethyl sulfoxide)	A994.1	Carl Roth
DNase	LK003178	Worthington Biochem
DNase I, grade II	10104159001	Merck
Doxycycline hyclate	D9891-5G	Merck

DPBS (Dulbecco's phosphate-buffered saline, no calcium, no magnesium)	14190-144	ThermoFisher Scientific
EPZ004777 (EPZ 4477)	SEL-S7353	Biozol
Essential 8™ Medium	A1517001	ThermoFisher Scientific
FCS (fetal calf serum)	10270106	ThermoFisher Scientific
FCS (fetal calf serum) for primary neurons	10500064	ThermoFisher Scientific
Fibronectin bovine plasma	F1141-5MG	Merck
GDNF (Glial derived neurotrophic factor)	130-096-291	Miltenyi Biotech
Gentle Cell Dissociation Reagent	07174	Stem Cell Technologies
GlutaMAX™ Supplement 100x	35050038	ThermoFisher Scientific
Hexadimethrine bromide (polybrene)	H9268-5G	Merck
Horse serum	16050122	ThermoFisher Scientific
Hydrochloric acid	X896.1	Carl Roth
KnockOut™ DMEM	10829018	ThermoFisher Scientific
KnockOut™ Serum Replacement	10828028	ThermoFisher Scientific
L-Ascorbic acid	A4403-100MG	Merck
Laminin from Engelbreth-Holm-Swarm murine sarcoma basement membrane	L2020-1MG	Merck
LDN193189	130-103-925	Miltenyi Biotech
LipoD293	SL100668	SignaGen Laboratories
MEM-Non-Essential Amino Acids 100x	11140035	ThermoFisher Scientific
mTeSR™1	85850	Stem Cell Technologies
N-2 Supplement (100X)	17502048	ThermoFisher Scientific
NaB (sodium butyrate)	B-5887	Merck
Neurobasal™ medium	21103049	ThermoFisher Scientific
Papain	LK003172	Worthington Biochem
PBS, w/o Ca <sup>2+</sup> , Mg <sup>2+</sup>	L182-50	VWR
PD0325901	130-103-923	Miltenyi Biotech
PFTα	P4236-5MG	Merck
Poly-D-lysin	A-003-E	Merck
Poly-D-lysin (Alex)	P7886	Merck
Poly-L-ornithine	P3655-100MG	Merck
Puromycin dihydrochloride	P9620-10ml	Merck
RevitaCell 100x	A2644501	ThermoFisher Scientific
SB431542	130-105-336	Miltenyi Biotech
Sodium hydroxide solution, 2N	T135.1	Carl Roth
StemMACS™ iPS-Brew XF, human	130-104-368	Miltenyi Biotech
Strychnine hydrochloride	ab120416	Abcam
SU5402	3300/1	Bio-technne
TeSR™-E7™ Medium for Reprogramming (2-Component)	05910	Stem Cell Technologies
Tetrodotoxin citrate	ab120055	Abcam
Trypan blue solution 0.4 %	T8154-100ML	Merck
TrypLE	A1217701	ThermoFisher Scientific
Trypsin Solution 10X	59427C-100ML	Merck
Tetrodotoxin citrate (TTX)	1069	Bio-technne
UltraPure™ 0.5M EDTA, pH 8.0 (Ethylenediaminetetraacetic acid)	15575-020	ThermoFisher Scientific
Vitronectin (VTN-N) Recombinant Human Protein, Truncated	A14700	ThermoFisher Scientific
Water For Injection (WFI) for cell culture	A1287301	ThermoFisher Scientific
XAV939	3748/10	Bio-technne

### 3.1.3 Cell culture media compositions

Name	Abbre-viation	Media base	Ingredients	Media for
Astrocyte media	AM	DMEM	10 % horse serum	Primary astrocyte cultivation
BrainPhys media	BPM	BrainPhys	3 % FCS, 1x B-27, 1x GlutaMax, 1x MEM-NEAA, 10 ng/µl BDNF, 10 ng/µl CNTF, 10 ng/µl GDNF, 1 µg/ml Laminin	Neuronal differentiation
Fibroblast media	FBM	DMEM	10 % FCS, 1x GlutaMax	Fibroblast cultivation
Freezing media	FM	DMEM / F12	10 % DMSO, 50 % FCS	Freezing cells
HEK293T cells media	HM	DMEM	10 % FCS	HEK293T cells cultivation
hKCl-solution	hKCl	WFI	170 mM KCl, 150 mM HEPES, 1 mM MgCl <sub>2</sub> , 2 mM CaCl <sub>2</sub>	iNeuron stimulations
Infection media	IM	iPS Brew	1x RevitaCell, 6 µg/ml polybrene	Neuronal differentiation
Neuronal basal media	NBM	Neurobasal medium	2 % FCS, 1x B-27, 1x GlutaMax, 1x MEM-NEAA, 10 ng/µl BDNF, 10 ng/µl CNTF, 10 ng/µl GDNF	Neuronal differentiation
Neuronal progenitor media	NPM (N2)	DMEM / F12	1x N-2 supplement, 1x GlutaMax, 1x MEM-NEAA	Neuronal differentiation
Patterning media	PM (XLSB)	KnockOut DMEM	15 % KO serum, 1x GlutaMax, 1x MEM-NEAA, 50 µM 2-Mercaptoethanol, 2 µM XAV939, 10 µM SB431542, 100 nM LDN193189	Neuronal differentiation
Splitting media	SM	DMEM / F12	1 % BSA	iPSC - single cell splitting media

### 3.1.4 Chemicals

Name	Catalog number	Manufacturer
2-Propanol	6752.4	Carl Roth
Bovine Albumin Fraction V (7.5 % solution)	15260037	Thermo Fisher Scientific
CaCl <sub>2</sub> (calcium chloride)	T885.2	Carl Roth
Digitonin	G9441	Promega
Ethanol (≥99,8 %, p.a.)	9065.3	Carl Roth
Glutaraldehyde	G5882-50ml	Merck
HEPES (4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid, N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid))	HN78.1	Carl Roth
IGEPAL (Octylphenoxy poly(ethyleneoxy)ethanol, branched)	I8896-100ML	Merck
Isobutanol	1009841000	VWR
KCl (potassium chloride)	6781.3	Carl Roth
Low TE	12090015	Thermo Fisher Scientific
Magnesium acetate tetrahydrate	M5561-50G	Sigma Aldrich
MgCl <sub>2</sub> (magnesium chloride)	KK36.1	Carl Roth
NaCl (sodium chloride)	9265.1	Carl Roth
PFA (Paraformaldehyde)	7398.1	Carl Roth
Protease Inhibitor Cocktail for use	P8340-5ml	Sigma Aldrich
Saccharose	9097.1	Carl Roth

TBE (Tris-Borate-EDTA Buffer, 10X Powder Pack)	sc-296651	Santa Cruz Biotechnology
Tris	4855.2	Carl Roth
Triton™X-100	T8787-50ml	Sigma Aldrich
Tween20	655205-250ML	Merck

### 3.1.5 Enzymes, corresponding buffers, master mixes, premade solutions

Name	Catalog number	Manufacturer
5X Q5® Reaction buffer pack	B90275	NewEngland BioLabs
Agencourt Ampure XP magnetic beads	A63881	Beckman Coulter
BSA, acetylated (20 mg/mL)	AM2614	ThermoFisher Scientific
BsmBI	R0580L	NewEngland BioLabs
CutSmart® Buffer	B7204S	NewEngland BioLabs
dNTP mix (100 mM)	N0446S	NewEngland BioLabs
EcoRI	R0101S	NewEngland BioLabs
KpnI-HighFidelity enzyme	R3142S	NewEngland BioLabs
Luna Universal One-Step RT-qPCR Kit	E3005	NewEngland BioLabs
NEBuffer™ 3.1	B7203S	NewEngland BioLabs
NEBuilder® HiFi DNA Assembly Master Mix	E2621S	NewEngland BioLabs
NEBNEXT High-Fidelity 2X PCR Master Mix	M0541L	NewEngland BioLabs
Q5® High-Fidelity DNA Polymerase	M0493	NewEngland BioLabs
Q5 Hot Start High-Fidelity 2x Master Mix	M0494L	NewEngland BioLabs
RevertAid H Minus First Strand cDNA synthesis	K1632	ThermoFisher Scientific
RiboLock RNase Inhibitor, 400 units/ml	EO0382	ThermoFisher Scientific
rSAP (recombinant Shrimp Alkaline Phosphatase)	M0371S	NewEngland BioLabs
SfiI	R0123	NewEngland BioLabs
SsoAdvanced™ Universal SYBR® Green Supermix	1725272	BioRad
SuperScript IV reverse transcriptase	18090050	ThermoFisher Scientific
SuperScript™ IV VILO™ Master Mix	11756050	ThermoFisher Scientific
T4 Ligase	M0202S	NewEngland BioLabs
T4 Ligase Reaction Buffer	B0202S	NewEngland BioLabs
T4 PNK	M0201S	NewEngland BioLabs
TBE	Sc-296651	ChemCruz
TD buffer	15027866	Illumina
TDE1 tagment DNA enzyme	15027865	Illumina
TE buffer	12090015	ThermoFisher Scientific
TURBO DNA-free™ Kit	AM1907	ThermoFisher Scientific
UltraPure DNase/RNase-Free Distilled Water	10977035	ThermoFisher Scientific
XbaI	R0145S	NewEngland BioLabs

### 3.1.6 Kits

Name	Catalog number	Manufacturer
Bioanalyzer High Sensitivity DNA Analysis	5067-4626	Agilent
Cell Line Nucleofector™ Kit R	VCA-1001	Lonza
Cell Line Nucleofector™ Kit V	VCA-1003	Lonza
Chromium™ Single Cell 3' Library & Gel Bead Kit v2	PN-120237	10X Genomics
D1000 Reagents and ScreenTape	5067-5583 and 5067-5582	Agilent
DNA Clean & Concentrator-5	D4004	Zymo Research
Human CD34+ Cell Nucleofector™ Kit	VAPA-1003	Lonza
Infinium Global Screening Array-24 Kit	20030770	Illumina
Infinium OmniExpress-24 Kit	20024631	Illumina

KAPA Library Quantification Kit	KK4854 – 07960298001	Roche
Micellula DNA Emulsion & Purification Kit	E3600	EURx
MiSeq Reagent Kits v2 (300 cycles)	MS-102-2002	Illumina
MinElute Gel Extraction Kit	28604	QIAGEN
Poly(A)Purist MAG Kit-1	AM1922	ThermoFisher Scientific
QIAamp DNA Blood Mini Kit	51104	QIAGEN
QIAGEN Plasmid Plus Midi Kit	12943	QIAGEN
QIAquick Gel Extraction Kit	28704	QIAGEN
Quick-RNA Miniprep Kit	R1054	Zymo Research
QIAprep Spin Miniprep Kit	27106	QIAGEN
QuantSeq 3' mRNA-Seq Library Prep Kit FW	015.96	Lexogen
QIAshredder	79654	QIAGEN
Qubit dsDNA HS Assay Kit	Q33230	ThermoFisher Scientific
Quick-RNA Microprep Kit	R1050	Zymo Research
Quick-RNA Miniprep Kit	R1055	Zymo Research
Quick-RNA Viral	R1034	Zymo Research
SIRV-Set3 (Isomix E0 / ERCC)	51.01	Lexogen

### 3.1.7 Plasmid backbones

Plasmid name	Encoding	Catalog number	Manufacturer
Bidirect_dCas9-KRAB_M2rtTa	Tet-On system, dCas9-KRAB	NA	GenScript
FUW-M2rtTa	Reverse tetracycline-controlled trans activator	20342	Addgene (Hockemeyer et al., 2008)
GBX	Human eGFP and BCL-xL genes	64123	Addgene (Chou et al., 2015)
MIP 247 CoMIP 4in1 with shRNA p53	OCT4, KLF4, c-MYC, SOX2, shRNA against p53	63726	Addgene (Diecke et al., 2015)
MMK	Human MYK, human KLF4	64121	Addgene (Chou et al., 2015)
MOS	Human OCT4, human SOX2	64120	Addgene (Chou et al., 2015)
MPRA_nLuc	nanoLuc under minimal promoter, blasticidin resistance	NA	Davide Cacchiarelli
pAAVS1-NDi-RV-SpdCas9-KRAB	Tet-On system, dCas9-KRAB, targeting AAVS1 locus	NA	Rene Maehr
pHAGE-EF1a-AAVSZnFG-PGK_puro	Zinc finger, puromycin resistance	NA	Rene Maehr
pLVX-TETON-Cas9	Cas9	632633	TaKaRa
pMD2.G (VSVG)	VSV-G envelope	12259	Addgene
pMPRA1	Multiple cloning site	49349	Addgene (Melnikov et al., 2014)
pMPRALenti1	Multiple cloning site	61600	Addgene
pNL3.1 vector	Minimal promoter, nanoLuc	N1031	Promega
pNL3.1_minP	Minimal promoter, nanoLuc	NA	Davide Cacchiarelli
psPAX2	HIV-1 gag, HIV-pol (2 <sup>nd</sup> generation)	12260	Addgene
pTet-O-Ngn2-puro	Tet-ON promoter, Neurogenin-2, puromycin resistance	52047	Addgene (Y. Zhang et al., 2013)
sgOpti	gRNA scaffold and cloning site, puromycin resistance	85681	Addgene (Fulco et al., 2016)
sgOpti, modified	gRNA scaffold and cloning site, blasticidin resistance	NA	In-house

### 3.1.8 Bacteria work

Name	Catalog number	Manufacturer
Ampicillin sodium salt	K029.3	Carl ROTH
Endura electro competent cells	60242-2	Lucigen
LB Agar (Lennox)	X965.2	Carl ROTH
LB Broth (Lennox)	X964.2	Carl ROTH
MF-Millipore™ Membrane Filter, 0.05 µm pore size	VMWP01300	Merck
NEB® Stable Competent E. coli	C3040H	NewEngland BioLabs
One Shot® TOP10 Electrocomp™ E. coli bacteria	C404050	ThermoFisher Scientific
One Shot™ ccdB Survival™ 2 T1R Competent Cells	A10460	ThermoFisher Scientific
TubeSpin® Bioreactor 50	Z761028-180EA	TPP

### 3.1.9 Immunocytochemistry

Name	Species	Dilution	Catalog number	Manufacturer
Aqua-Poly/Mount	NA	NA	18606-20	Polysciences
β3-Tubulin (TUBB3)	Chicken	1:500	MA1-19187	ThermoFisher Scientific
Albumin Fraction V	NA	20 % in DMEM	8076.4	Carl Roth
Ctip2	Rabbit	1:300	ab28448	Abcam
Cux1 (Cutl1)	Mouse	1:300	ab54583	Abcam
DAPI (4',6-Diamidin-2-phenylindol)	NA	1:1000	6843.1	Carl Roth
Donkey IgG anti-Chicken IgY (H+L)-Alexa Fluor 488	Donkey	1:300	703-545-155	Dianova
Donkey IgG anti-Mouse IgG (H+L)-Alexa Fluor 594	Donkey	1:300	715-585-150	Dianova
Donkey IgG anti-Rabbit IgG (H+L)-Alexa Fluor 647	Donkey	1:300	711-605-152	Dianova
FoxG1	Rabbit	1:200	ab18259	Abcam
MAP2	Chicken	1:7000	ab5392	Abcam
Nanog	Mouse	1:250	MABD24	Merck
Anti-NeuN clone A60	Mouse	1:300	MAB377	Merck
Oct4 (POU5F1)	Rabbit	1:500	ABD116	Merck Millipore
Normal donkey serum	S30-100ML	Merck	MAB1598	Merck
Satb2	Mouse	1:200	ab51502	Abcam
Synapsin 1/2	Rabbit	1:300	106002	Synaptic Systems
Triton™ X-100	NA	1:1000	T8787-50ML	Merck

## 3.2 Cell culture

Cells were maintained in 5 % CO<sub>2</sub> and ambient oxygen levels at 37 °C. Prewarmed media was changed in a sterile hood according to the feeding schedule for the different cell lines. A summary of the components of the different media compositions is summarized in 0. The density of the cultures was monitored by inverted light microscopy. Cell culture was

performed in two different laboratories depending on the lentiviral load of the cultured cells with different equipment (Table 1).

	No infections	Cells with lentiviral infection
<b>Incubator</b>	MCO-20AIC-PE	MCO-19M-PE
<b>Centrifuge</b>	Allegra X-22R Refrigerated Benchtop centrifuge (Beckman Coulter)	Allegra X-22R Refrigerated Benchtop centrifuge (Beckman Coulter)
<b>Inverted microscope</b>	Leica DMi1 inverted microscope	Inverted Laboratory Microscope Leica DM IL LED
<b>Camera on microscope</b>	5 Megapixel HD-Microscope camera Leica MC170 HD	CCD Microscope Camera Leica DFC3000 G
<b>Sterile hood</b>	Maxisafe 2020	Maxisafe 2020

**Table 1 Description of cell culture equipment**

Depending on the lentiviral load of the HEK293T cells or iPSC derived cultures, different incubators, centrifuges, microscopes, and biosafety cabinets were used for the cell culture experiments.

For several experiments it was necessary to seed a specific number of cells per cm<sup>2</sup>. Therefore, cells were split single cell and a live-dead staining with Trypan blue solution was performed: 10 µl of the single cell solution were mixed with 10 µl trypan blue solution (0.4 %). HEK293T and primary astrocytes were counted manually in a Neubauer chamber. All other cell types were loaded into single-use counting chambers and counted by the automatized cell counter Luna II according to cell specific programs (Table 2).

Cell type	Dilution factor (1 – 100)	Noise Reduction (1 – 10)	Live Cell Sensitivity (1 – 9)	Roundness (0 – 100 %)	Min. cell size (3 – 59 µm)	Max. cell size (4 – 60 µm)	Declustering Level
Fibroblasts	2	7	1	20	8	60	Medium
PBMCs	2	5	7	60	3	30	Medium
iPSCs	2	6	3	0	3	16	High
NPCs day-4	2	1	9	39	3	60	High
Nuclei	5	1	1	10	3	60	Medium

**Table 2 Cell type specific settings of the automatized cell counter**

The settings of the automatized cell counter Luna II were altered cell type specific in order to account for differences in morphology, size, and the likelihood for the presence of cell multiplets.

### 3.2.1 Preparation of lentiviruses

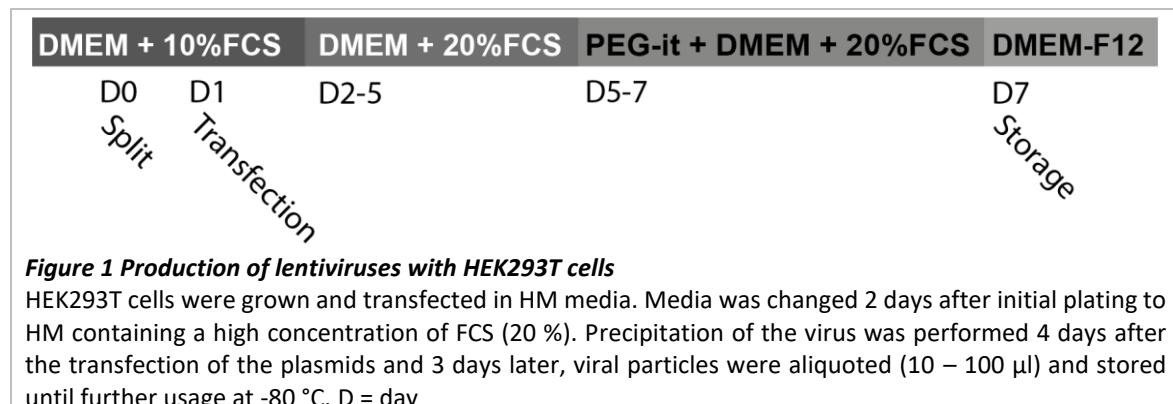
#### 3.2.1.1 Cultivation of the immortalized cell line HEK293T

HEK293T cells (ATCC® CRL3216™) were maintained in HEK media (HM: DMEM supplemented with 10 % FCS). Cells were splitted with 1x TrypLE when they reached 80 % confluency. To avoid undesired premature inhibition of the trypsin, cells were first washed with DPBS. When cells started to detach, trypsin was stopped by addition of fresh culture media (HM). Cells were separated by gentle pipetting and centrifuged at 300 rcf for 5 min

at RT. After the supernatant was discarded, the cells were resuspended in HM and replated for at a dilution of 1:10 to 1:25 unless they were seeded for lentiviral generation.

### 3.2.1.2 Production of infectious lentiviral particles

Lentiviral infectious particles were generated by HEK293T cells (Figure 1). HEK293T cells were counted one day before the transfection and seeded at a density of 7 mil cells / 10-cm dish. Cells were used for lentiviral production until passage 13 after initial storage of those cells. On the day of the transfection (day-1), the media was removed, and 8 ml HM were added to the 10-cm dish 30 min before the transfection. The DNA (15 µg of the plasmid of interest, 7 µg psPAX2 and 3.5 µg pMD2.G) was diluted in 500 µl DMEM. The transfection reagent (76.5 µl LipoD293) was diluted in 500 µl DMEM. The solutions were vortexed vigorously before and after addition of the LipoD293 containing solution to the DNA. The transfection solution was incubated for 10 min at RT. Afterwards, the transfection solution was added dropwise to the media taking care not to dislocate the cells.



16-18 h later (day-2), the medium was changed to 20 ml DMEM containing 20 % FCS to support the HEK293T cells during the production of the lentivirus. 2 days later (day-5), the supernatant was collected and sterile filtered (Acrodisc 0.45 µm filter). 5x PEG-it was added to precipitate and concentrate the virus for 3 days at 4 °C. After the precipitation, the lentiviral particles were centrifuged (1,500 × g for 30 minutes at 4 °C) and resuspend in 1/100 of the original volume in DMEM/F-12. The precipitated virus was aliquoted and stored at -80 °C. Thawed lentivirus aliquots were discarded after one week at 4 °C.

### 3.2.1.3 Quantification of the lentiviral particles

In order to transduce the cells with sufficient lentiviral particles, we quantified the particles by qPCR (protocol kindly provided by Dr. Davide Cacchiarelli, TIGEM, Italy). Absolute quantification of the number of RNA strands present was performed using the WPRE element present in the virus. As a standard for the absolute quantification, a lentiviral plasmid (pLVX-TETON-Cas9), coding for dCas9 under the doxycycline inducible promoter was used to obtain a standard curve. The plasmid was linearized by EcoRI digestion and purified from an agarose gel with columns according to manufacturer's protocols (QIAquick Gel Extraction Kit). The molecular weight of the linearized plasmid was calculated according to the size of the plasmid. Therefore, it was possible to calculate the number of WPRE fragments in one  $\mu$ l using the linearized plasmid concentration (Qubit).

In order to quantify the viral particles, the RNA of the lentivirus (10  $\mu$ l virus) was isolated with a column-based approach. In the beginning of the project, the RNA was isolated with the Quick-RNA Microprep and later on with the Quick-RNA Viral kit (both Zymo Research). For the gRNA viruses which were prepared at the same time, the measured RNA concentration (Nanodrop) was used to normalize the amount of virus per gRNA pool. Reverse transcription and qPCR were performed with Luna Universal One-Step RT-qPCR kit. Per reaction, 3.7  $\mu$ l extracted RNA or diluted WPRE plasmid DNA were mixed with 5  $\mu$ l Luna Universal One-Step Reaction Mix (2x), 0.5  $\mu$ l 20x Luna WarmStart RT Enzyme Mix, and 0.4  $\mu$ l (10  $\mu$ M) each reverse and forward primer. The WPRE element was specifically amplified with a forward primer (5'-ACACCTGTCAGCTCTTC-3') and reverse primer (5'-CAACACCACGGAATTGTCAG-3') pair. After an initial reverse transcription step for 10 min at 55 °C, and a denaturation step for 1 min at 95°C, the PCR cycle was repeated 45 times (95 °C for 10 sec, 60 °C for 30 sec). Standards and samples were pipetted in triplicates and at least two virus isolations for each virus were performed.

## 3.2.2 Primary cells derived from mouse brain tissue

### 3.2.2.1 Isolation of primary astrocytes from p1 mice

Primary mouse astrocytes were isolated from Crl:CD1(ICR) mice, following previously published protocols (Kaech & Bunker, 2006). In brief, mouse pups were sacrificed one day after birth. Next, the cortex was isolated taking care to remove the meninges. The tissue fragments of the cerebral cortex were washed with DPBS and digested with 2x trypsin and

DNase (grade II) to achieve a single cell solution. Cells were further separated by repeated pipetting and administration to a cell separation filter (70 µm). After a spin down, the single cells from 6-8 pups were plated in DMEM containing 10 % horse serum (astrocyte media, AM) on cell culture flasks treated with attachment factor for at least 1 h at 37 °C. On the next day, flasks were agitated vigorously to remove contaminating microglia and media was changed. Cells were grown until 95 % confluence and split twice with 2.5x trypsin before the coculture with neurons.

### 3.2.2.2 *Isolation of primary neurons from E15.5 mice*

Lysed primary cortical mouse neurons infected with the MPRA libraries were gratefully obtained from Dr. Alexander Herholt (laboratory of Prof. Moritz Rossner, LMU). In brief, murine embryos (E-15.5, C57Bl/6N) were collected after performing a Caesarean surgery and the brains were dissected to recover the cells located in the cerebral cortex. A single cell solution was prepared, and neurons were seeded at 8 mil cells per 15-cm dish (BD Falcon) previously coated with 0.1mg/ml Poly-D-lysine in Neurobasal™ supplemented with 2x B-27+A, 1x GlutaMAX and 5 % FCS. Primary mouse neurons (C57Bl/6N) were infected one day after plating (day in vitro 1, DIV1) during the necessary media change. The media was only changed once more to medium containing Neurobasal, 2x B-27, and 1x GlutaMAX on DIV6. On DIV12, the neuronal cultures were harvested with lysis buffer (3 – 4 ml / 15-cm dish). Each condition (see 3.2.7.1) was replicated 3 – 5 times and each replicate consisted of one 15-cm dish.

### 3.2.3 **Reprogramming induced pluripotent stem cells from primary human cells**

The different iPSC lines were generated from primary human cells with consent in written form. Reprogramming the fibroblasts from patients suffering from schizophrenia (S1 - S5) was part of this thesis and is, therefore, explained in more detail in 3.2.3.1. The control cell lines C1 and C6 (License C0135C) were generated by the laboratory of Dr. Micha Drukker (Helmholtz-Center Munich) (Cárdenas et al., 2018) and by the laboratory of PD Dr. Dietmar Spengler (MPI of Psychiatry, Munich), respectively. The controls C2 to C5 originated from the BECOME study performed at the MPI of Psychiatry. The CLM lines (S6 - S8) were derived from a patient cohort collected at the LMU Clinic for Psychiatry (Table 3).

Cell origin name	iPSCs cell line name	Donor name	Primary cell type	Gender	Origin of primary cells
GM01792	SCZ1	S1	Fibroblasts	Male	Coriell Institute
GM02497	SCZ2	S2	Fibroblasts	Male	Coriell Institute
GM01835	SCZ3	S3	Fibroblasts	Female	Coriell Institute
GM02038	SCZ4	S4	Fibroblasts	Male	Coriell Institute
GM02503	SCZ5	S5	Fibroblasts	Female	Coriell Institute
CLM0003	CLM3	S6	PBMCs	Male	MPI-P Biobank
CLM0007	CLM7	S7	PBMCs	Male	MPI-P Biobank
CLM0008	CLM8	S8	PBMCs	Female	MPI-P Biobank
BJ ATCC® CRL-2522	No1	C1	Fibroblasts	Male	ATCC
FOK0001BE00M32	FOK1	C2	PBMCs	Male	MPI-P Biobank
FOK0002BE00M32	FOK2	C3	PBMCs	Male	MPI-P Biobank
FOK0003BE00M32	FOK3	C4	PBMCs	Male	MPI-P Biobank
FOK0004BE00M32	FOK4	C5	PBMCs	Female	MPI-P Biobank
Human dermal fibroblasts, adult	HDF6	C6	Fibroblasts	Female	Thermo Fisher Scientific

**Table 3 Overview of the reprogrammed and utilized iPSC lines in this thesis**

7 human primary fibroblasts and 7 human primary peripheral blood mononuclear cells (PBMCs) were reprogrammed and used for experiments in this thesis.

### 3.2.3.1 Reprogramming from human skin fibroblasts

Primary human fibroblasts (Coriell) were shipped in fibroblast medium (FBM: DMEM, 10 % FCS, 1x MEM-NEAA) at RT. The cells were equilibrated to 37°C overnight. On the next day, the media was exchanged for the first time. Subsequently, FBM was completely changed every 3-4 days. Cells were split when the cultures where 100 % confluent: fibroblasts were washed with DPBS and incubated at RT with 0.5 mM UltraPure™ EDTA until (3 to 5 mins) cells started to change their morphology. 1x TrypLE fully replaced the EDTA and was incubated until the cells lifted off (3-10 mins). TrypLE was stopped with FBM, cells were centrifuged at 300 rcf for 5 mins and fibroblasts were seeded at a 1:4 dilution in FBM.

The reprogramming (Figure 2) of human primary skin fibroblasts was performed according to a previously published protocol (Diecke et al., 2014). On the day before the nucleofection, 2 mil cells were plated per 15-cm dish. On the next day (day-0), 2 mil cells were nucleofected with 20 µg CoMIP plasmid using the Amaxa Nucleofector II (program U23) and the cell line kit R. The CoMIP plasmid encodes *SOX2*, *POU5F1*, *MYC*, *KLF4*, and a short hairpin RNA against *TP53* (Diecke et al., 2015). Immediately after the nucleofection, cells were plated on 2 matrigel (1:100 in DMEM/F12) coated wells of a 6-well plate.

One and three days after the nucleofection, media was changed to FBM containing several reprogramming factors: 0.3 µM PFTα, 3 µM EPZ 4777, 0.2 mM NaB, 50 µg/ml ascorbic acid.

On day-4, the media was half-changed to a stem cell propagating media: supplemented TeSR™-E7™ Medium (E7 media). On day-5 to day-11, E7 media including inhibitors was replaced every day. On day 13, media was changed half to Essential 8™ Medium (E8 media) including inhibitors. Afterwards, media was changed every day with supplemented E8 media until the first iPSCs colonies appeared (around day-16 to 25). Only then, E8 media was added every day to the cells without inhibitors.

NaB		PFT $\alpha$ , EPZ 4777					
Fibroblast media		E7 media		E8 media		iPS Brew	
D-1	D0	D1	D2-4	D5-11	D12	D13	D19
Split	Nucleo- fection				Split	Split	D25...

**Figure 2 Timeline of the reprogramming process from primary fibroblasts to iPSCs**  
Human primary fibroblasts were reprogrammed into induced stem cells over a time of over 3 weeks. During this period, the media was changed repeatedly to support the different intermediate states of the cells. Several small molecules were added to support the generation of iPSCs.

Single colonies were picked and plated into one well of a 48-well plate coated with matrigel containing E8 supplemented with 1x RevitaCell. Colonies grew for 5 – 7 days until they could be propagated into 24-well, 12-well, and eventually 6-well plates. Colonies were split at this stage with 1x Gentle Cell Dissociation Reagent (more detailed splitting protocol in 3.2.4). After 6 to 8 passages cells were slowly habituated to the cultivation in iPS Brew.

In order to purify pluripotent stem cells in the iPSC culture, a Tra1-60 enrichment was performed (Dick et al., 2011; Grabundzija et al., 2013). First, cells were split single cell with 1x accutase, diluted with splitting media (SM, DMEM/F12 containing 1 % BSA), spun down (300 rcf for 5 min) and resuspended in 80  $\mu$ l E8 containing 1x RC and 20  $\mu$ l magnetic beads. After an incubation at 4 °C for 5 min, cells were diluted in 1 ml E8 containing 1x RC. The labelled cells were transferred to a magnetic separation column (MS Column). The column was washed three times with 0.5 ml E8 containing 1x RC. The labelled cells were immediately flushed out in 1 ml E8 containing 1x RC into a 1 well of a 6-well plate coated with matrigel. The unlabeled fraction was discarded.

### 3.2.3.2 Reprogramming human peripheral blood mononuclear cells

The biobank team of the MPI of Psychiatry had isolated the peripheral blood mononuclear cells (PBMCs) using SepMate™-50 columns from fresh blood sample. In brief, columns were

prefilled with 15 ml Lymphoprep solution and cells, diluted in 15 ml PBS supplemented in 2 % FCS, were added gently on top. The top layer was discarded after an initial centrifugation at 1200 rcf for 10 min at RT. PBMCs were washed twice with PBS / 2 % FCS and centrifuged at 300 rcf for 8 min. Cells were frozen down at 10 million cells per vial in freezing media (50 % FCS, 10 % DMSO, 40 % DMEM/F12).

The three patient derived donor lines (CLM lines) used in this thesis were reprogrammed by Anna Hausruckinger and Dr. Miriam Gagliardi with my help. The healthy control derived PBMCs (FOK1, FOK2, FOK3, FOK4) were reprogrammed in a collaborative effort between the laboratory of Prof. Elisabeth Binder (Vincenza Sportelli) and the laboratory of Prof. Michael Ziller (Anna Hausruckinger, Dr. Miriam Gagliardi, and I). In brief, cells were thawed at 37 °C in a water bath, resuspended dropwise in 10 ml DPBS supplemented with 2 % FCS. After pelleting (300 rcf for 8 min at RT), cells were resuspended in 1 ml Erythroid Expansion Medium and further diluted to 5 ml in a 25 cm<sup>2</sup> flask. On the next day, cells were transferred to a new flask. Afterwards, every other day, the media was changed completely until cells were nucleofected between day-12 to 14 (CD34+ Cell Nucleofector Kit, program T-016) with the reprogramming plasmids (4 µg MOS, 4 µg MMK, 2 µg GBX). The three plasmids, MOS, MMK, and GBX code for *SOX2* and *POU5F1*, human *MYC* and *KLF4*, and human *BCL2*, respectively. Directly after the nucleofection, cells were transferred in a 12-well plate and incubated for 2 days. Three days after the transfection, the media was changed to DMEM/10 % FCS and cells were spun down (30 min at 25 °C at 200 g) to attach on a 12-well plate coated with vitronectin (0.5 µg/cm<sup>2</sup>). Media was replaced on the next day with E8 media containing sodium butyrate (0.25 mM NaB). Cells were fed every other day with E8 containing NaB until iPSC colonies appear 10 to 14 days after the transfection. Then, cells were given E8 media every day until colonies were picked and expanded as described beforehand.

### 3.2.3.3 Generation of embryoid bodies with iPSCs

One 90 % confluent well of a 6-well plate was used for the embryoid body (EB) formation assay. This experiment was performed by Christina Fürle with the help of Susanne Maidl under my supervision. Each iPSC line was individually tested for its ability to form EBs. iPSCs were washed with DPBS. 1 ml Gentle Cell Dissociation Reagent was added to the culture for at least 5 min at RT until iPSC colonies started to lift from the matrigel. The well was

gently washed with 1 ml iPS Brew, and the colonies were spun down for 5 min at RT in a 15 ml falcon. The supernatant was discarded, and cells were resuspended (3 times pipetting up and down with a 5 ml pipette) in KO medium supplemented with 10 % KO serum, 1x MEM-NEAA, and 1 % L-glutamine. Small colonies were transferred to a 6 cm low-attachment dish. Every other day, cells were transferred to a 15 ml falcon to exchange the media. After 10-15 min (depending on the size of the EBs), EBs settled down in the bottom of the falcon and the supernatant was discarded and the fresh media was used to transfer the EBs back into the 6-cm dish. EBs were imaged with a bright field microscope on day-14 of the differentiation.

#### *3.2.3.4 Validation of the karyotype integrity of the generated induced pluripotent stem cells*

Genomic DNA (gDNA) was extracted before and after the successful reprogramming process in order to verify that no genomic abnormalities occurred during the reprogramming process. gDNA from one well of a 6-well plate (iPSCs) was extracted according to manufacturer's protocol (QIAamp DNA Blood Mini Kit). 200 ng were loaded on a microarray (Infinium Global Screening Array, v1 to v3) to analyze the abundance and allelic representation of SNPs distributed over the genome. In order to analyze the integrity of the karyotype of the samples post reprogramming, each sample was compared to the genotype of the original primary sample (Lucia Trastulla, AG Ziller) according to a previously published pipeline (Danecek et al., 2016).

### **3.2.4 Maintenance of induced pluripotent stem cells**

iPSCs were cultivated on matrigel (1:100 in DMEM/F12) in iPS Brew. Cells were split if they reached 70 % confluence with 1:4 accutase. For this purpose, cells were washed with DPBS and incubated with the diluted accutase until small colonies detached. Those small colonies were placed carefully in DMEM/F12 and spun down at 300 rcf for 5 min. Cells were plated on matrigel coated dishes in iPS Brew supplemented with 1x RevitaCell (RC). Cells were splitted depending on their growth rate between 1:5 to 1:10. Cells were only used for experiments below passage 50 to reduce appearance of CNVs in the genome (Garitaonandia et al., 2015). If cells needed to be counted, accutase was incubated undiluted, stopped with SM and cells were counted after resuspension in iPS Brew supplemented with RevitaCell (see 3.1.8).

In order to keep cell lines for repetitive usage, one well of a 6-well plate was frozen down in freezing media (50 % FCS, 10 % DMSO, 40 % DMEM/F12). One frozen vial was thawed at 37 °C and resuspended gently but fast in 6 ml of warm DMEM/F12. Cells were spun down (300 rcf, 5 min at RT) and plated onto one well of a 6-well plate in iPS Brew with 1x RC.

### 3.2.5 Generation of doxycycline inducible dCas9-KRAB cell lines

dCas9-KRAB lines were generated in-house by Dr. Miriam Gagliardi (AG Ziller). In brief, the dCas9-KRAB gene fragment was inserted by transposons (pAAVS1-TetOn-dCas9-KRAB, pHAGE-EF1a-AAVSZnFG-PGK\_puro). Transient puromycin resistance was used to select single positive colonies which were verified for insertion on the genomic, mRNA, and protein level by PCR, qPCR, and western blot, respectively (data not shown).

### 3.2.6 Neural differentiations of the patient and healthy control derived iPSCs

#### 3.2.6.1 *Differentiation into cortical neural progenitor cells*

Ingredients	Day-0&-1	Day-2&-3	Day-4&-5	Day-6&-7&-8
KnockOut™ Serum Replacement	7.50 %	7.50 %	5 %	2.50 %
GlutaMAX™ Supplement	1x	1x	1x	1x
MEM NEAA	1x	1x	1x	1x
2-Mercaptoethanol (50 mM)	100 µM	100 µM	100 µM	100 µM
B27:N2 mix (2:1)	-	-	1 %	2 %
LDN193189	250 nM	250 nM	250 nM	250 nM
SB431542	10 µM	10 µM	10 µM	10 µM
XAV939	5 µM	5 µM	5 µM	5 µM
PD0325901	-	1 µM	8 µM	8 µM
SU5402	-	10 µM	10 µM	10 µM
DAPT	-	10 µM	10 µM	10 µM

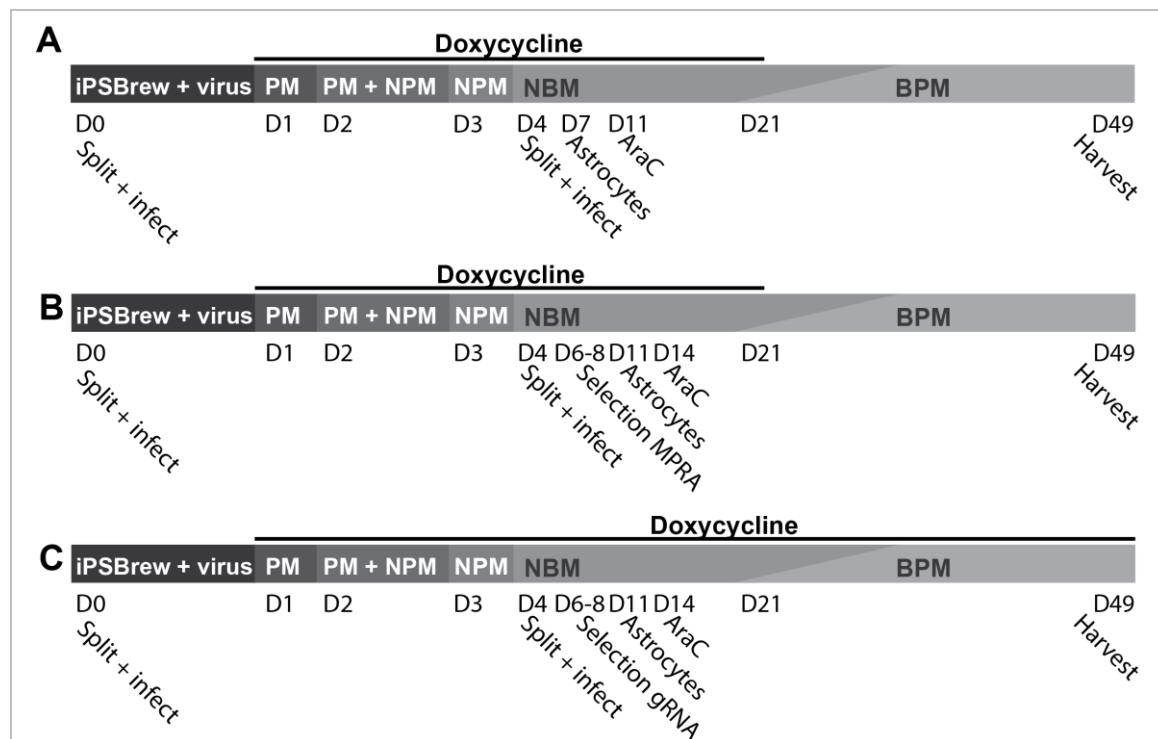
**Table 4** Media composition for the differentiation of iPSCs into neural progenitor cells

Media composition for the neural progenitor differentiation was altered every other day until cells were harvested on day-9. The media contained patterning factors to induce a forebrain excitatory neuronal fate. Neuronal progenitor cells were differentiated by Dr. Ruhel Ahmad (AG Ziller) following a previously published and adapted protocol (Y. Qi et al., 2017). In brief, 2 mil cells were plated on matrigel (1:30 in DMEM/F12) on one well of a 6-well plate. In order to differentiate the stem cells into forebrain cortical neuronal progenitor cells, patterning factors were added to the differentiation media (media composition in Table 4). Every day, for 8 days, media was changed until cells were harvested on day-9. Human iPSC derived NPCs were splitted after 5 days of differentiation with 1x accutase. Cells were resuspended

in day-6 media with 8  $\mu$ l MPRA virus and cells were replated in two thirds of the original area. Each sample was prepared from two 10 cm dishes (6 – 8 mil cells).

### 3.2.6.2 Differentiation of human iPSCs into induced neurons

Neuronal cultures were generated by overexpression of the doxycycline inducible mouse Neurogenin-2 (*Ngn2*) transcription factor (Y. Zhang et al., 2013) and inhibition of the SMAD- and WNT pathways during the first days of differentiation (Figure 3).



**Figure 3 Timeline of the differentiation process from iPSCs to iNeurons**

**A-C**) iNeurons were differentiated for 49 days after an initial infection with the tetracycline responsive system and the mouse derived neuronal transcription factor *Ngn2*. The media was changed initially every day to different patterning media (PM, NPM). 4 days later, cells were split, infected with virus to investigate gene regulation, and seeded single cell. **A)** Astrocytes were added 3 days later and AraC treatment (selection against proliferative cells) was performed on day 11. Cells were fed every 3-4 days afterwards, and *Ngn2* induction was stopped after day 24. Starting from day 21 on until day 49, cells were fed with BPM media. **B,C)** Astrocyte addition and AraC treatment took place after initial selection with blasticidin for the presence of the MPRA fragment or the gRNA, respectively. **C)** In order to maintain the expression of dCas9, doxycycline was supplemented in the media throughout the differentiation.

Single cell solutions of iPSCs were infected in infection media (IM) containing iPS Brew supplemented with 1x RevitaCell and 6  $\mu$ g/ml polybrene. Cells were infected at a concentration of 500,000 cells / ml with 3 - 15  $\mu$ l virus at RT for 10 min and plated at a density of 1.2 - 2.5 mil (depending on the growth rate) on a 10 cm dish coated with matrigel. iPSCs were infected simultaneously with two viruses (FUW-M2rtTA and pTet-O-*Ngn2*-puro) consisting of the TET-ON system (reverse tetracycline-controlled trans activator and TET

responsive promoter), mouse *Ngn2*, and a puromycin resistance gene. Starting from day-1 until day-21, *Ngn2* expression was induced by addition of doxycycline (2 µg/ml) to the differentiation media. This is in contrast to previously published protocols where *Ngn2* expression was maintained until the day of harvesting. Changes in length of the doxycycline induced *Ngn2* expression, without patterning factors, does not lead to the generation of different cell types but affects the proportion of generated cell types (Lin et al., 2020).

In order to enrich for forebrain excitatory neurons, a patterning step (dual SMAD, Wnt inhibition) was introduced. Patterning occurred on day-1 with full concentration of patterning chemicals and on day-2 with half of the small molecule concentration (see 0). In comparison with published protocols, which were not present at the time of the final decision of the differentiation procedure described in this thesis, this duration of patterning was substantial lower (2 days instead of 3 or even 6 days) (M. Chen et al., 2020; Nehme et al., 2018). The patterning media (PM, see 0) contained three inhibitors (2 µM XAV939, 10 µM SB431542, 100 nM LDN193189) to generate forebrain like neurons. The elongation of dual SMAD inhibition alone (LDN193189 and SB431542) is sufficient to decrease PNS (PRPH) and RG (PTN) markers and increase glutamatergic markers such as ELAVL4 and FOXG1 (M. Chen et al., 2020). The small molecule XAV939 present for the first two days of the differentiation inhibited the Wnt-pathway necessary to generate forebrain GPM6A and POU4F1 positive cells. In order to deplete the cell culture from *Ngn2* negative cells, puromycin selection was performed. However, since a splitting step was also introduced, cells were selected for 2 days with 10 µg/ml and not for 5 days with 3 µg/ml puromycin as described elsewhere (M. Chen et al., 2020). Therefore, puromycin (10 µg/ml) was added to the media on day-2 (1:1 of PM and neuronal progenitor media, NPM, see 0), and on day-3 (NPM). The NPC state was supported on day-3 by NPM containing puromycin and doxycycline.

One necessary alteration of the original protocol was to introduce a splitting step in order to account for the different functional analyses (MEA wells, coverslips for immunocytochemistry, normal plates for RNA- and ATAC-Seq, second infection for MPRA and CRISPRi experiments) and the different growth rates of the cell lines (M. Chen et al., 2020). Therefore, on day-4, NPC-like cells were incubated with 1x accutase until cells lifted off. Cells were dissociated to single cells with a P1000 pipette and centrifuged (5 min, 300 rcf, RT) in splitting media. Cells were resuspended in Neurobasal medium (NBM) containing

no FCS but doxycycline (2 µg/ml) and plated on previously coated neuronal differentiation plates (80,000 cells / 48-well). For the neuronal differentiation plates, 12.5 µg/ml Poly-L-ornithine in DPBS was incubated overnight at 37 °C. Afterwards, the plates were washed 3 times with DPBS and incubated 6 h to over night with laminin (1 µg/ml) and fibronectin (2 µg/ml) diluted in DPBS. The field recording chambers, and the coverslips were coated with double of the coating reagent concentration. Additionally, the number of plated cells was increased to 180,000 cells/ recording chamber or 200,000 cells / cover slip considering the lower attachment rate of the cells on glass. For the MPRA experiments as well as the CRISPRi experiments, DNA fragments, and gRNAs had to be expressed in the iPSC derived neurons. Therefore, the intermediate neuronal cell population was infected while replating on day-4. The cell suspension was diluted to 1 mil cells / ml and 1.4 - 8 µl virus were added depending on the concentration of the virus. An equivalent virus amount (75 ng viral RNA) was applied to independent differentiation wells of a 6-well plate during the day-4 split of the iNeurons for the CRISPRi experiments. The viruses were incubated for 30 min at RT before the cells were plated.

To further mature the differentiated neurons, astrocytes were added at day-7. 5,000 primary mouse astrocytes (3.2.2.1) per 0.95 cm<sup>2</sup> were plated onto the neuronal culture while performing a full media change with NBM containing 2 % FCS. The astrocyte plating time point is 3 days later than in Nehme et al., because astrocytes were added after the antibiotic selection and after the splitting step. On the other hand, astrocytes were added to the culture 4 days earlier than in Zhang et al. and, therefore, addition of FCS to the differentiation media was started earlier (day-7 instead of day-11). Even though FCS is known to inhibit spontaneous AP firing in neuronal network, FCS was added to the media composition up to 2 or 3 % of the total media volume due to the fact that the astrocytes need to support and mature the generated neuronal network for a total of 7 instead of 3 weeks (Y. Zhang et al., 2013). 4 µM cytarabine in NBM was added on day-11 of the differentiation to enrich for postmitotic neurons and prevent unlimited expansion of primary astrocytes. Afterwards, 2/3 of neuronal media were changed every 3-4 days until cells were harvested after 7 weeks of differentiation. NBM media was used for the first 3 weeks and after day-24 the base of the neuronal media was changed to BrainPhys (media composition see BPM media). BrainPhys media increases spontaneous excitatory and inhibitory firing rate, and neurite branching points in 2D cultures (Bardy et al., 2015; Satir

et al., 2020). BPM media was supplemented with Laminin (1 µg/ml) to facilitate adherence and maturation of the neurons (Figure 3 A).

For the single cell RNA-Seq experiment, cells underwent the protocol described above, however, BPM was depleted of FCS in order to facilitate electrophysiological maturation (Bardy et al., 2015). However, as the MEA and the patching experiments indicated, this depletion of FCS was not necessary in line with results from previous Ngn2 based differentiation protocols (Nehme et al., 2018; Y. Zhang et al., 2013). Additionally, depletion of FCS could affect the constitution of the cocultured astrocytes since then, the media would be depleted for important growth factors and nutrients. An insufficient physical interaction of astrocytes with neurons and influence on the media composition affects the maturity of the neurons (Burke et al., 2020; Meijer et al., 2019; Muratore et al., 2014; Nehme et al., 2018; Tang et al., 2013).

In the case of the dCas9-KRAB experiments to link and validate functional enhancers, doxycycline was added to the media every other day throughout the differentiation process at a higher concentration (4 µg/ml) to ensure a stable expression of dCas9-KRAB. Due to the fact that the patterning and the overexpression of Ngn2 via doxycycline was performed the same way for the alternative differentiation protocols (Figure 3 B, C), lineage specificity of the individual neuron should have not been affected. Lineage specification is reported to be determined after the first days while the first 3 weeks of the differentiation protocol were the same to the normal protocol (Lin et al., 2020). The use of blasticidin S in the culture media for a total of 3 days was sufficient to select for those cells infected with at least one of the gRNAs from the gRNA pool (Figure 3 B). For the knockdown experiments, one sample equals one well of a 6-well plate. One sample for the MPRA experiments consisted of 4-6 dishes (10 cm diameter) or 6-well plates.

### 3.2.7 Stimulation of neuronal networks

#### 3.2.7.1 Stimulation of primary mouse neurons

Primary mouse neuronal cultures were silenced (“silent” condition) the day before the harvest (DIV11) for 24 h, artificially excited for 4 h (“excited” condition) on DIV12, or not treated (“baseline” condition) before harvesting (Table 5).

Condition	Drugs	Effect on
Baseline	-	-
Silent	1 $\mu$ M TTX	Block of sodium gated potassium channels
	100 $\mu$ M D-APV	No NMDA receptor transmission
Excited	100 $\mu$ M 4-AP	Block of voltage gated potassium channels
	100 $\mu$ M glycine	Co-agonist of NMDA receptor
	1 $\mu$ M strychnine	Inhibition of glycine receptor alpha subunit
	50 $\mu$ M bicuculline	Block of GABA <sub>A</sub> receptor

**Table 5 Overview of the applied stimuli on mouse primary neurons**

Three different stimuli were applied to the primary mouse neurons before the cells were harvested on DIV12.

### 3.2.7.2 Stimulation of human iNeurons

iNeurons were treated on day-49 with a high-potassium solution (hKCl) to depolarize the cells to investigate the impact of depolarization on gene expression. Cells were stimulated (“hKCl” condition) in BPM supplemented with KCl-solution (170 mM KCl, 150 mM HEPES, 1 mM MgCl<sub>2</sub>, 2 mM CaCl<sub>2</sub>) for 5 h. Cells were harvested directly after the stimulation. As a control, cells were not stimulated and kept in BPM (“baseline” condition).

For the MEA experiments, cells were either hKCl stimulated with different concentrations or the synaptic transmission in the culture was altered. In order to block excitatory glutamatergic synaptic transmission while allowing the generation of action potentials (APs), APV (50  $\mu$ M) and NBQX (10  $\mu$ M) were applied in BPM. AP generation was afterwards blocked with tetrodotoxin (TTX, 1  $\mu$ M).

## 3.3 Electrophysiological characterization of the generated iNeurons

Electrophysiological experiments were performed at the Max Planck Institute of Psychiatry in close collaboration with Dr. Mathias Eder. Patch experiments were performed by Barbra Hauger. Raw data extraction and mEPSC measurements were performed by Dr. Mathias Eder.

### 3.3.1 Multi electrode array field recordings of mature iNeurons

Electrical field activity was recorded after 7 weeks of neuronal differentiation. The recording chamber was prewarmed to 37 °C and the CO<sub>2</sub> levels were equilibrated to 5 % prior to recording. The chamber containing the neuronal culture was placed into the recorder and equilibrated for 10 min. 59 recording electrodes (30  $\mu$ m electrode diameter,

200  $\mu$ m electrode spacing) were recording neuronal activity in parallel. Baseline network activity was measured for 3 min. The different hKCl stimulation concentrations were applied and the effect on the neuronal cultures were measured once for 3 min. APV, NBQX, TTX effects were recorded for 1 – 3 minutes after applying the reagents.

Recordings were low-pass filtered and digitized with MCS-IFB 3.0 multiboot (MultiChannel Systems). Recordings for APV, NBQX and TTX induced changes were performed with MC\_Rack (MultiChannel Systems). Recordings (50 kHz sampling rate) for the hKCl induced changes were performed with Multi Channel Suite (MultiChannel Systems). Raster plots and close-ups of single channels were generated offline with the Plexon Offline Sorter software. For action potential detection, a threshold value of 5.8x RMS noise was applied to the data.

### 3.3.2 Single cell recording of mature iNeurons

iNeurons grown on coverslips were placed in a submerged recording chamber and equilibrated for some minutes. Recordings were performed at RT. Coverslips were continuously superfused with carbonated (95 % O<sub>2</sub> / 5 % CO<sub>2</sub>) ACSF at a flow rate of 2 - 3 ml / min. ACSF contained the following ingredients: 121 mM NaCl, 4.2 mM KCl, 1.1 mM CaCl<sub>2</sub>, 1 mM MgSO<sub>4</sub>, 29 mM NaHCO<sub>3</sub>, 0.45 mM NaH<sub>2</sub>PO<sub>4</sub>-H<sub>2</sub>O, 0.5 mM Na<sub>2</sub>HPO<sub>4</sub>, and 20 mM D-glucose (Bardy et al., 2015). The patch pipettes had an open tip resistance of 3-5 M $\Omega$  and a seal resistance greater than 1 G $\Omega$ . Recordings were low-pass filtered (3 kHz) and digitized at 15 kHz using an EPC9 amplifier and stored with the acquisition program Pulse, version 8.5 (Heka Elektronik, Lambrecht, Germany). Recordings were only started 10 min after break-in to the cell. 8 mV liquid junction potential correction was considered for the analysis.

Whole-cell current clamp (<20 M $\Omega$  series resistance) was performed in order to analyze electrophysiological properties of the neurons. In order to inhibit inhibitory and excitatory input onto the neurons, ACSF contained 100  $\mu$ M picrotoxin and 5  $\mu$ M NBQX. The intra cellular solution for the patch pipette was 135 mM KMeSO<sub>4</sub>, 8 mM NaCl, 10 mM HEPES, 2 mM Mg<sub>2</sub>ATP, 0.3 mM Na<sub>3</sub>GTP, 0.3 mM EGTA at a pH of 7.3. Whole cell current clamp tracks were visualized in Pulse and directly exported. Square-pulse current injections hyper- or depolarized the neurons.

Whole-cell voltage clamp (-70 mV) experiments were analyzed in order to investigate the occurrence and properties of mEPSCs. The experiments were performed in the above mentioned ACSF with an increase of the calcium concentration from 1.1 mM to 2 mM and the presence of 1 µM TTX. The intracellular solution (pH 7.2) for the patch pipette consisted of the following: 125 mM Cs methansulfonate, 10 mM HEPES, 0.5 mM EGTA, 8 mM NaCl, 20 mM Na<sub>2</sub>-phosphocreatine, 4 mM Mg<sub>2</sub>ATP, and 0.3 mM Na<sub>3</sub>GTP. Properties of mEPSCs were analyzed offline with IgorPro v.6.1 software (WaveMetrics, Inc., Portland, Oregon). Data was plotted in R with ggplot2 package (Wickham, 2016).

## 3.4 Detection of proteins in iPSCs and iPSC derived neurons

### 3.4.1 Immunocytochemistry

iPSCs and iNeurons were plated on coverslips in 24-well plates. To fix the cells, the coverslips were washed with DPBS. Afterwards, 4 % PFA in DPBS was applied for 20 min at RT. Cells were washed three times with DPBS. To block unspecific binding of proteins, cells were incubated with blocking buffer (1 % donkey serum, 0.1 % BSA, 0.1 % Triton™ X-100 in DPBS) for 1 h at RT. Primary antibodies were diluted in blocking buffer and incubated overnight at 4 °C (dilutions see 3.1.4). Cells were washed three times with 0.1 % Triton™ X-100 diluted in DPBS for 5 min at RT. Then, the secondary antibodies (1:300) and DAPI (4',6-Diamidin-2-phenylindol, 1 µg/ml) were incubated for 3 h at RT. Secondary antibodies were raised in donkey and recognized the epitope of mouse, rabbit, and chicken. Afterwards, the cells were washed again three times with the above-described washing buffer. Coverslips were mounted on glass slides with Aqua-Poly/Mount until imaged.

Pluripotency was verified with antibodies against NANOG and OCT4 (*POU5F1*). Neuronal identity was verified with antibodies recognizing epitopes from the following proteins: SYNAPSIN1/2, MAP2, β3-TUBULIN (*TUJI*, *TUBB3B*), NEUN (*RBFOX3*), CUTL1 (*CUX1*), CTIP2 (*BCL11B*), SATB2, and FOXG1.

### 3.4.2 Confocal imaging

The exposure time was fixed for each staining condition between different cell lines and different experiments. The exposure was set, that cells only stained with the secondary antibody did not show any specific signal. Coverslips were imaged on a confocal microscope

(Leica TCS SP8 confocal laser scanning microscope) equipped with a 10x (dry, 0.3 numerical aperture, NA, HC PL FLUOTAR), 20x (0.75 NA; immersion, HC PL APO CS2) and a 40x (water immersion, 1.10 NA, HC PL APO CS2) objective. Several images were taken at multiple z-planes and merged into maximal intensity projections. The pinhole was set to 1.0 airy units. Per picture 1024 x 1024 or 2048 x 2048 pixels were imaged, and the picture was averaged two times. Images were taken at RT. Lasers were separately adjusted to improve signal to noise ratio. The different fluorophores were excited consecutively to minimize false positive signal. The LAS X software (version 2.0.1.14392, Leica) controlled intensity, offset, filters, and laser intensities.

### 3.4.3 Analysis of intensity of expressed markers

Confocal and epifluorescent pictures were processed (i.e. tonal corrections, addition of scale, and merging) in FIJI (Schindelin et al., 2012) and assembled in Adobe Illustrator (Adobe Systems, San Jose). All pictures of one experiment were treated the same.

For quantification of marker proteins of the iNeurons, 9 – 10 images were counted per cell line per marker. Confocal images were taken with a 10x objective, and the z-plane was selected for a bright staining of the nucleus (DAPI channel, 408 nm). Image analysis was performed using the below annotated semi-automated batch-macro in ImageJ (Wayne Rasband, NIH, open source). In order to run the below code, the images had to be sorted in a separate folder for each channel. The DAPI channel was named ch0 for this code. chX is a placeholder for any other channel to quantify the intensity. The macro “Background intensity” was used to quantify the intensity of a channel over the whole image.

```
macro "Background intensity"{
//input folder DAPI
file1 = getDirectory("ch0_Select folder containing files to process 0");
list1 = getFileList(file1);
n1 = lengthOf(list1);
//Input folder staining to analyse
file3 = getDirectory("chX_Select folder containing files to process X");
list3 = getFileList(file3);
n3=lengthOf(list3);
//the output folder
file4 = getDirectory("output_Select an output folder merge");
list4 = getFileList(file4);
n4 = lengthOf(list4);
// Insert a ROI at 0/0 position with 1024/1024 pixel size
small = n1;
open(file1 + list1[1]);
run("ROI Manager... ");
makeRectangle(0, 0, 1024, 1024);
roiManager("Add");
```

```
//quantify intensity of ROI
for(i = n4; i < small; i++) {
  open(file3 + list3[i]);
  run("Invert LUT");
  roiManager("Show All without labels");
  roiManager("Measure");
}
}
```

The macro “ROI intensity” measured the intensity in the regions of interest (ROI) automatically defined by the DAPI channel on the other channels of the same image. It also saves automatically the quantifications performed by IMAGEJ or FIJI as a .csv file in the preselected folder.

```
macro "ROI_intensity"{
  setBatchMode(true);
//Input folder DAPI
file1 = getDirectory("ch0_Select folder containing files to process 0");
list1 = getFileList(file1);
n1 = lengthOf(list1);

//Input folder staining to analyze
file3 = getDirectory("ch2_Select folder containing files to process 2");
list3 = getFileList(file3);
n3 = lengthOf(list3);

//the output folder
file4 = getDirectory("output_Select an output folder merge");
list4 = getFileList(file4);
n4 = lengthOf(list4);

//Looping over adjustments to make the DAPI nuclei detectable automatically
small = n3;
for(i = n4; i < small; i++) {
  open(file1 + list1[i]);
  run("Invert LUT");
  run("Morphological Filters", "operation=Opening element=Octagon radius=2.5");
  run("Extended Min & Max", "operation=[Extended Minima] dynamic=15 connectivity=4");
  run("Invert LUT");
  run("Analyze Particles...", "size=45-115 pixel show=Outlines display exclude clear summarize add");
  selectWindow("Results");
  close("Results");

//transfer the ROIs from DAPI to the channel of interest
  open(file3 + list3[i]);
  run("Invert LUT");
  roiManager("Show All without labels");
  roiManager("Measure");
  name = list1[i]+list3[i];
  selectWindow("Results");
  saveAs("Results", file4 + name + ".csv");
  close("*");
}
  showMessage(" -- finished --");
}
```

Aggregates of neurons were excluded from the analysis because a maximum nuclei size was set in order to exclude astrocytic nuclei. These settings also automatically excluded overlapping nuclei from analysis which would have been counted otherwise as one cell.

This also harbors the advantage that a positive and an overlapping negative nucleus were not labeled as an intermediate expressing cell. Generated measurements of the intensity of the markers were processed in Excel, R (version 4.0.3 (2020-10-10), R Foundation for Statistical Computing, Vienna, Austria), and GraphPad Prism (version 9.0.0 for Windows, GraphPad Software, San Diego, California, USA).

### 3.4.4 Western Blot with WesSimple

TCF4 protein expression in iNeurons harvested on day-49 was analyzed by Western Blot. Neuronal networks from day-49 iNeurons were flash frozen in 1.5 ml tubes. The whole protein was extracted with DPBS supplemented with 1x Protease inhibitor and 0.1 % IGEPAL. Protein content was measured with Bradford assay on the MPM6 machine (BIO-RAD). In brief, 10 µl of each sample of the whole protein was diluted in a 1:5 mixture of Bradford solution and water. Each sample was analyzed in duplicates and the average value was calculated. A standard curve with BSA was used to correlate color intensities and protein concentration. The intensity of the lysis solution was subtracted from each of the measured color intensities. 1 µg protein of each sample was analyzed for the protein abundance of TCF4 and MAP2. Preparation of the samples was performed according to the protocol published for the EZ Standard Pack1 with a measurable protein size between 12 - 230 kDa. The Western blot, the image export, and intensity quantification were performed with Compass for Simple Western (ProteinSimple, San Jose, California, USA). Intensities were plotted with R (ggplot2).

## 3.5 RNA levels in iNeurons

### 3.5.1 Bulk RNA isolation

For total RNA isolation, frozen neuronal networks, frozen cell pellets, or living cultures were incubated with lysis buffer. Total RNA was isolated following the instruction of the Quick-RNA Microprep Kit. At least two wells of a 12-well plate or one well of a 6-well plate were used for RNA isolation on one column following manufacturer's protocol. The optional DNA digestion step on the column was performed as indicated. Quality and concentration of the isolated RNA was checked on a NanoDrop or the TapeStation to verify correct optical density (OD) or RIN values, respectively.

### 3.5.2 Quantitative real time PCR to analyze gene expression

400, or 1000 ng total RNA were used to prepare cDNA for the schizophrenia vs control, or the knockdown with dCas9 experiments, respectively. In order to further reduce DNA contamination, a DNase step (TurboDnase kit) was introduced into the protocol. 1.8  $\mu$ l DNase buffer and 1  $\mu$ l DNase were added to 16  $\mu$ l total RNA and incubated for 30 min at 37 °C. The DNase was stopped and separated from the RNA by a bead purification step. For this purpose, 2  $\mu$ l beads were added to the digestion and incubated for 2 min at RT. Beads were centrifuged and the supernatant was collected.

cDNA was reverse transcribed according to manufacturer's recommendation (RevertAid H Minus First Strand cDNA synthesis kit). For this purpose, RNA diluted in water (final volume 12.5  $\mu$ l) was incubated with 1  $\mu$ l oligo dT-primer for 5 min at 65 °C. Before the addition of 4  $\mu$ l Reaction buffer, 1  $\mu$ l RNase Inhibitor, 2  $\mu$ l 10 mM dNTPs and 1  $\mu$ l reverse transcriptase, the reaction was cooled down on ice. Afterwards, the reverse transcription started, with a 5 min incubation step at 25 °C, followed by an incubation at 45 °C for 60 min. Afterwards, the enzyme was denatured at 70 °C for 5 min.

To verify differential gene expression, quantitative real time PCRs (qPCR) were performed. Gene expression was normalized to *MAP2* abundance to compare gene of interest levels between different samples. Each sample was analyzed in triplicates or quadruplicates and the mean was calculated before calculating the fold change. qPCR was performed on the LightCycler® 480 using SYBRGreen as a fluorescent dye to quantify the amount of dsDNA (SsoAdvanced™ Supermix). cDNA was amplified following this protocol: 95° for 2 min; [95° for 15 s, 60° for 10 s, 72° for 10 s]x40; 60° for 1 s; 45° for 1 s; melting curve. 2  $\mu$ l of a 1:6 to 1:10 diluted sample were used in a 10  $\mu$ l reaction including 5  $\mu$ l SsoAdvanced master mix, 0.5  $\mu$ l FW- and 0.5  $\mu$ l RV-primer (10  $\mu$ M initial concentration), and 2  $\mu$ l water.

### 3.5.3 Bulk 3'RNA sequencing

For bulk RNA sequencing, 500 ng total RNA were used to generated 3' mRNA sequencing libraries following manufacturer's protocol (QuantSeq 3' mRNA-Seq). In order to normalize the libraries, spike-ins were added during library preparation (SIRV-Set3). Libraries were amplified for sequencing with 14 PCR cycles. Each library was examined for correct size distribution and concentration with the Bioanalyzer or the tape station. Library

preparations were performed by Susanne Maidl (AG Ziller). Up to 50 libraries were pooled equimolar and sequenced on a NovaSeq6000 sequencer (50 bp paired-end sequencing).

Bioinformatic analysis of the bulk RNA sequencing data was performed by Prof. Dr. Michael Ziller. Paired-end raw reads were aligned to GRCh 37 or GRCh38 human genome assembly (GENCODE version 27 transcriptome definition) utilizing the STAR aligner (version 2.7.3a):

```
--outFilterType BySJout --outFilterMultimapNmax 20 --outFilterMismatchNmax 999 --outFilterMismatchNoverReadLmax 0.04 --alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --alignSjoverhangMin 8 --alignSjDboverhangMin 1 --sjdbScore 1 --readFilesCommand zcat --outSAMtype BAM Unsorted SortedByCoordinate --quantMode GeneCounts --outSAMstrandField intronMotif. Duplicates were discarded and library quality metrics were calculated with picard tools (http://broadinstitute.github.io/picard/). Read counts on the gene level were generated with the Rsubread package collapsing the exon level reads onto gene level features: countMultiMappingReads=FALSE, ignoreDup=FALSE, sGTFAnnotationFile=TRUE, isPairedEnd=T. Genes were excluded if their total read counts, considering all samples, were below the 20th or above the 99th percentile. Further exclusion criteria were genes from mitochondria and a read count less than 2 or more than 52,220 summed over all samples. Differential expression analysis was performed with DESeq2 (version 1.22.2) correcting for experimental batch for diagnosis or hKCl stimulation effect with the design formula ~batch+condition+diagnosis or ~batch+donor+condition, respectively. Differential genes showed an adjusted p-value  $\leq 0.05$  and a minimal absolute log2FC  $\geq 0.5$ . For plotting, 3' mRNA counts were vst normalized.
```

I plotted the metadata from the samples using ggplot2. PCA analysis was performed with PCAtool (Blighe K, Lun A (2021). PCAtools: PCAtools: Everything Principal Components Analysis. R package version 2.6.0, <https://github.com/kevinblighe/PCAtools>) generating the biplot of PC1 and PC2, the eigencor and the component loading plot. Significant difference between conditions was tested using the non-parametric Mann Whitney test in GraphPad prism (version 9.1.0). Normality was tested with Shapiro-Wilk test with GraphPad. I plotted the marker quantification and differential expressed genes after hKCl stimulation with the open source pheatmap package (Version 1.0.12, Raivo Kolde, 2019) and hierarchical clustered according to RNA expression.

### 3.5.4 Single cell RNA sequencing

Neuronal networks were digested with a papain (10 units/ml) – DNase (100 units/ml) mixture diluted in DMEM/F12 to facilitate generation of a single cell suspension. After 10 min at 37 °C, the network was mechanically disrupted with a P1000 pipette tip. DMEM/F12 supplemented with 20 % FCS was added for a centrifugation step at 400 rcf for 5 min (RT). Cells were resuspended in DMEM/F12 supplemented with 1 % BSA and 1x RC and spun down (400 rcf for 5 min at RT). Final resuspension was in sterile filtered 1 % BSA in DPBS supplemented with RNase Inhibitor, 1x RC, 3 mM Mg(Ac)<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. After filtration to remove network debris (70 µm cell strainer), cells were counted, and 12,000 cells (3,000 from each cell line) were loaded on the 10X Chromium machine (10X Genomics). Libraries from single cells were prepared following manufacturer's instructions for the Chromium™ Single Cell 3' Library & Gel Bead Kit v2.

Sequencing output files (FASTQ format) were loaded and modified with Cell Ranger version 3.1.0 (10X Genomics) by Vanessa Murek. Data was aligned against the human genome GRCh38. Clustering of the scRNA sequencing data was performed by Prof. Dr. Michael Ziller in Seurat v3.2.2 (Stuart et al., 2019). Clustering of cells was performed using the following parameters: reduction.type = "pca", dims.use = 2:20, resolution = 0.3, nn.eps = 0.5). Genes were excluded from the analysis if they were expressed in less than one cell. Cells were analyzed if their feature count was between 1000 and 10,000, and they contained less than 15 % mitochondrial reads. Subsequently, 20 PCAs were selected for dimensionality reduction based on inspection of 100 PCs in an elbow plot analysis (data not shown). Data was normalized by sctransformation. Seurat was used to obtain a UMAP embedding using standard parameters and a min\_dist parameter of 0.2, followed by NearestNeighbor identification.

Vanessa Murek deconvoluted the genotypes of the 4 cell lines present in the single cell experiment presented here with demuxlet (H. M. Kang et al., 2018). For this purpose, the fastq output files of the single cell experiment were mapped with CellRanger count to hg19. Laura Jiménez Barron generated a master vcf file containing all karyotyped and imputed SNPs from all cell lines present in the laboratory. This master vcf file was filtered for the 4 used cell lines. This filtered vcf was further down-sampled for the relevant SNPs located in the 3'UTR exons. Annotations for the SNPs in the 3'UTR exons were taken from <http://genome.ucsc.edu/cgi-bin/hgTables> (downloaded 09.02.2021). The final vcf

encompassed 97,809 sites to be used for deconvolution. Deconvolution was run with demuxlet with the following parameters: demuxlet --alpha 0.0 --alpha 0.0 --vcf vcf\_1ps\_christine\_hg19.filtered.recode.sorted.vcf --field GT --sam mpg\_L20019-5\_Christine-iN-5-scRNA-Seq\_hg19/outs/possorted\_genome\_bam.bam --out demuxlet\_mpg\_L20019-5\_Christine-iN-5-scRNA-Seq\_hg19\_GT --group-list mpg\_L20019-5\_Christine-iN-5-scRNA-Seq\_hg19/outs/filtered\_feature\_bc\_matrix/barcodes.tsv.gz.

Results were mapped to the previously generated Seurat object (hg38 processed) and all barcodes were preserved. Vanessa Murek performed automatic classification of the cell types present in the differentiated neurons utilizing a single cell mouse data set comprising of neurons from the central and peripheral nervous system (Zeisel et al., 2018).

Manual marker gene selection for the individual clusters was performed according to the 30 most enriched genes for each cluster. Additionally, neuronal subtype specific genes were considered. The selected markers were analyzed for their specificity for the mouse nervous system on the online platform [www.mousebrain.org/genesearch.html](http://www.mousebrain.org/genesearch.html) accessed between the 22.12.2020 – 01.05.2022. The heatmap depicting scRNA-Seq expression of the cluster - and brain specific genes was plotted with pheatmap package. Cells belonging to individual clusters were plotted as neighbors without any further hierarchical clustering. Genes were clustered hierarchically with the pheatmap algorithm. The color scheme originates from the ColorBrewer package in R.

### 3.6 Bulk open chromatin

ATAC-seq (assay for transposase-accessible chromatin using sequencing) was performed in bulk cell pellets according to previously published protocols (Corces et al., 2017). The neuronal network of one 12-well was frozen down in freezing buffer at -80 °C. The frozen approach required an initial thaw and resuspension in 1 ml PBS. The network was washed 3 times with 1 ml PBS and pelleted (400 rcf, 5 min, 4 °C) before the normal protocol was followed to lyse the nuclei. For the freshly isolated cells, 100,000 cells were pelleted before the first lysis step. Detergents for the lysis buffer were added freshly to the basic RSB buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>): RSBI (RSB supplemented with 100 µg/ml digitonin, 1:100 Tween<sup>TM</sup>20, 1:100 IGEPAL) and RSBII (1:100 Tween<sup>TM</sup>20). Cell pellets were pipetted 4 times up and down in 50 µl RSBI with the P200 pipette and lysed on ice for 3 min. To stop the lysis, 1 ml of RSBII was added and the tubes were gently mixed 6 times

with a P1000 pipette. The nuclei were pelleted at 500 rcf for 10 min at 4 °C. In order to count the nuclei from the frozen tissue, lysed nuclei were resuspended in 100 µl nuclei buffer (DPBS supplemented with 3 mM magnesium acetate, 5 mM CaCl<sub>2</sub>, 1 % BSA), and automatically counted in trypan blue. 70,000 nuclei were pelleted (550 rcf at 4 °C for 10 min) and used for the transposition reaction. The supernatant was carefully removed and the transposition mix (25 µl TD buffer, 2.5 µl TDE1 tagment DNA enzyme, 16.5 µl PBS, 0.5 µl Tween™20, 0.5 µl digitonin [1 % stock], 5 µl water) was mixed ten times with the nuclei. The transposition reaction was incubated at 37 °C for 30 min whilst mixing (1000 rpm) on a thermomixer. The transposition was stopped with binding buffer (DNA Clean & Concentrator kit). The fragmented DNA was column-purified according to manufacturer's instructions and eluted in 22 µl elution buffer provided by the kit. For the amplification of the library, 20 µl of the purified gDNA were mixed with 25 µl NEBNext High-Fidelity master mix, 2.5 µl FW primer (25 µM, 5'-AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG-3') and 2.5 µl of the RV primer with the library barcode (barcodes denoted in Table 6, 25 µM, 5'-CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTCTCGTGGGCTCGGAGATGT-3') (Buenrostro et al., 2015). Amplification was performed for 5 min at 72°C, 98°C for 30 sec, and 8 cycles of 98 °C for 10 sec, 63 °C for 30 sec, 72 °C for 1 min, and a final elongation step of 5 min at 72 °C.

ATAC		Barcodes			
TAAGGCGA	GTAGAGGA	ACCACTGT	CCGTTTGT	GAGGGGTT	AGGTTGGG
CGAGGCTG	GTCGTGAT	TGGATCTG	TGCTGGGT	TGGGTTTC	TGGTCACA
AAGAGGCA	CCACTCCT	AGGCAGAA	TCCTGAGC	GGACTCCT	TAGGCATG
GTGTGGTG	TTGACCCCT	CTCTCTAC	CAGAGAGG	GCTACGCT	

**Table 6 Summary table of the barcodes used for the open chromatin libraries**

23 barcodes were used to amplify and single index the open chromatin libraries.

After the amplification, a bead cleanup with 0.55x of the volume and 0.45x of the volume was performed. Before the pooling, libraries were quantified and verified for a successful transposition on a Bioanalyzer with a high sensitivity DNA chip. Up to 23 libraries could therefore be equimolar pooled for sequencing on a NovaSeq (50 bp paired-end sequencing), respectively. The final pool was purified with a double-sided cleanup of 0.5x and 0.45x bead volume of the library volume and eluted in 20 µl low-TE buffer if the final concentration was below 2 nM.

Computational analysis of the open chromatin data set was performed by Prof. Dr. Michael Ziller. In brief, paired-end raw ATAC-Seq reads were aligned against the GRCh37 (for MPRA

analyses) or GRCh38 (bulk ATAC) genome with bowtie2 version 2.3.5 (Langmead & Salzberg, 2012). Duplicates were removed using picard tools. Peaks were called with macs2 with standard human genome size (hs parameter). Afterwards, all peak files were merged with diffbind into one union peak set excluding those peaks from further analysis which occurred in less than 6 samples. These fragments which overlapped with peaks by at least one base were counted with the bedtools multicov function and were summarized in a matrix peak X fragment count matrix. Sequencing depth normalization and size factor estimation was performed using the effective library size for which only those reads were considered which overlapped with at least one peak of the union peak set. The dispersion was estimated using DESeq2's function: `estimateDispersions(obj, fitType='local')`. Differential ATAC peak enrichment was performed with a paired design using DESeq2: design formula `~donor+Factor` and `nbinomWaldTest` function. Peaks which were assigned a Benjamini-Hochberg corrected p-value  $\leq 0.05$  and a minimal absolute  $\log_{2}FC \geq 1$  between hKCl and baseline or in the case control comparison were defined as differential and used in the subsequent analysis.

Partitioned heritability scores were calculated by Laura Jiménez Barrón on ATAC peaks of day-49 iNeurons as published previously (Finucane et al., 2015). Only those ATAC peaks were considered which showed a stimulation dependent alteration in activity. Intersected ATAC regions were subgrouped for those regions which increased or decreased in accessibility comparing baseline with hKCl stimulated iNeuron data sets. hKCl dependent ATAC peak activity was compared pairwise independent of disease status. Genomic locations were intersected with LDSC annotation files using bedtools version 2.27.1-1-gb87c465 and LD scores were calculated using LDSC regression version 1.0.0. Subsequently, a cell-type specific analysis estimated the disease association for the following traits: major depressive disorder (MDD), antidepressant treatment resistance, any psychotic experience, schizophrenia, ever smoked, autism spectrum disorder (ASD), bipolar disorder (BD), coronary artery disease (CAD), height, type 1 diabetes (T1D), type 2 diabetes (T2D), Neuroticism, low density lipoprotein (LDL), celiac, fasting glucose, and years of education. All GWAS summary statistics were obtained from publicly available data sets. For schizophrenia, the GWAS summary statistics were obtained from Ripke et al. (2014). Significant enrichment ( $FDR < 0.05$ ) in a modified baseline model (version 1.1) containing

52 categories was corrected for multiple testing. Z-scores were calculated from the coefficient p-value and plotted in a heatmap (Finucane et al., 2018).

### 3.7 Targeted repression of enhancer activity using dCas9-KRAB

Region name	Location of locus within TCF4 gene	Complementary gDNA sequence (5' to 3')
SNP1a	rs61023811	GTCAAGAAAACCGCTGAGC
SNP1a	rs61023811	TATGTTGCCCTGGACCACGAA
SNP1b	rs61023811	GCTGTGTGCGGTGAACGATT
SNP1b	rs61023811	AGCCCAGAATTCTATAATGGG
SNP1b	rs61023811	TAATACAAGGTATCGATCA
Prom_short	Promoter for short transcript	CAATGTGTCTGCCTTATTAA
Prom_short	Promoter for short transcript	TACAATCTCCACAACATGTC
Prom_short	Promoter for short transcript	AATCAAGATCGCTCTGTCAC
Prom_short	Promoter for short transcript	TACTTCTTATAGCCTTAATA
SNP2	rs75741749	GGAAAGTCTAGGTAGAAGGA
SNP2	rs75741749	TCAACAAGTATTAAGACTGT
SNP2	rs75741749	CAAGTATTAAGACTGTTGGC
SNP3	rs78322266	CTTTAATTGTAAGAGGAG
SNP3	rs78322266	AATTGACATGCTAATT
SNP3	rs78322266	CACCGTCAGGCCAGATCCCC
SNP4	rs72930740	AGTAGCTGAACAATTATCG
SNP4	rs72930740	TAGTCTGCTCGTGTAGGT
SNP4	rs72930740	GTATTCCACTAGTTATCAGA
Prom_long_c	Promoter for long transcript	CTCTTCTCGACGTATCTAG
Prom_long_c	Promoter for long transcript	GCCCCACTTGGAAAGGCAGTT
Prom_long_c	Promoter for long transcript	CGGTAGGTACCGCGCCTGC
Prom_long_c	Promoter for long transcript	CTAAGGCAGCCATTGCTGT
Prom_long_c	Promoter for long transcript	CTGGTCGACCACGCCCTCCTC
Prom_long_d	Promoter for long transcript	TTGCCAGGAAACGTAGCCCT
Prom_long_d	Promoter for long transcript	TTCCAAGAGAGACGATCAAAC
Prom_long_d	Promoter for long transcript	TTGCCCGTTGCATCCCTCGG
Prom_long_d	Promoter for long transcript	TTATCAATGTGACTCCACGG
Prom_long_a	Promoter for long transcript	TTATCAATGTGACTCCACGG
Prom_long_a	Promoter for long transcript	TACCGCCCGCGCGGAGAAG
Prom_long_b	Promoter for long transcript	TCGGGTAGGCGTCGCGCGTG
Prom_long_b	Promoter for long transcript	CCGCAGTCCCGGATGTGAA
Prom_long_b	Promoter for long transcript	TGGAGTGAAAACACGTCAA

**Table 7** CRISPRi targeted regions and their gRNA targeted gDNA sequence

Annotation of the short region name, the target of the gRNA and the corresponding sequence of the binding sequence of the gRNA on the genomic DNA (gDNA).

In order to functionally link enhancers which showed differential gene regulatory activity, several enhancers and promoters putatively regulating gene expression in the TCF4 gene locus were chosen. The CRISPRi system was utilized to repress promoter / enhancer activity

by targeting KRAB via gRNA guided dCas9 to the specific sites on the genome. gRNAs were selected based on the publicly available online platform (<http://www.crispor.tefor.net>; version 4.92). gRNAs were scored within the program according to their sensitivity and specificity. The scorings as well as the vicinity to the SNPs and enhancer regions were considered to select gRNAs. For each gene regulatory element or SNP at least 3 different gRNAs were selected. Rosalba Proce, a master student intern, performed this step under my supervision for 2 of 10 regions of interest. gRNAs were ordered as single complementary oligos (metabion GmbH). gRNA sequences were edited like the following: gRNA oligos containing the sequence with the protospacer adjacent motif NGG were preceded by CACC. In case the first nt of the gRNA was not a G, gRNA was preceded by CACCG. gRNAs containing the complementary sequence were added a G at the 3' end of gDNA sequence and were preceded by AAAC or AAA in case the 5' end of the complementary sequence started with a cytosine (Table 7).

A lentiviral system was used in order to express the gRNAs in the iNeurons. Due to the fact that the iNeurons were positive for the puromycin resistance, the resistance of the expression vector (sgOpti) was changed from puromycin to blasticidin. I defined the cloning strategy and Susanne Maidl cloned the backbone in the laboratory (data not shown). The generated vector is referred to as “modified sgOpti”. Apart from the blasticidin S resistance it also encodes for the U6 promoter which drives gRNA expression.

gRNA annealing and subsequent cloning steps as well as sequencing steps were performed by Christina Fürle under my close supervision. The modified sgOpti vector was digested with BsmBI in NEBuffer™ 3.1 for 5 h at 55 °C. After an additional dephosphorylation with rSAP for 30 min at 37 °C, the 7,110 bp long vector backbone was separated according to its size from its 1,885 bp fragment via gel electrophoresis. The digested backbone was purified with the QIAquick Gel Extraction Kit according to manufacturer's recommendation. 1 µl FW- and RV-oligo (each 100 µM) were annealed and phosphorylated in 6.5 µl water, 1x T4 ligation buffer, and 0.5 µl T4 PNK (5 units). The oligos were first heated to a temperature of 37 °C for 30 min and to 95 °C for 5 min. Subsequently, the temperature was decreased by 5 degrees every minute until the temperature reached 25 °C. The annealed oligos and the backbone were ligated for 5 h at RT (ligation reaction described in Table 8). Afterwards, the ligation enzyme was inactivated by an incubation step at 65 °C for 15 min (Ran et al., 2013).

Ingredients	Volume in $\mu$ l for negative control	Volume in $\mu$ l for ligation
T4 DNA Ligase Reaction Buffer	0.25	0.5
T4 Ligase	0.60	1.2
Backbone vector	50 ng	100 ng
Annealed oligos (1:100)	-	2 $\mu$ l
Water	Add to 6.0 $\mu$ l	Add to 12 $\mu$ l

**Table 8 Description of the cloning reaction for the gRNA plasmids**

Negative control included the backbone vector but not the insert to verify the specific integration of the gRNAs into the cloning vector.

2  $\mu$ l of the ligation were transformed into 25  $\mu$ l chemically competent bacteria (NEB® Stable Competent E. coli). For this purpose, bacteria were thawed on ice for 10 min, and the DNA was mixed gently with the bacteria. Bacteria were incubated with the DNA for 30 min on ice. Afterwards, bacteria were heat shocked at 40 °C for 30 s and immediately placed on ice for 5 min. Bacteria grew in 500  $\mu$ l of outgrowth media for 1 h at 37°C whilst shaking at 300 rpm on a thermomixer. 200  $\mu$ l of the transformant was plated on an agarose plate containing ampicillin. On the next day, colonies were picked and grew for 14 - 18 h at 37 °C with shaking in ampicillin containing LB medium. The day after, plasmid DNA was isolated using the QIAprep Spin Miniprep Kit following manufacturer's instructions.

Region name	Chr. 18 Start	Chr. 18 End	Targeted element	Number of gRNAs	$\mu$ l virus used for 765,000 cells
<b>SNP1a</b>	52967979	52969975	rs61023611	2	1.61
<b>SNP1b</b>	52967979	52969975	rs61023611	3	1.76
<b>Prom_short</b>	52985201	52987383	Prom_short	4	1.01
<b>SNP2</b>	53060208	53060642	rs75741749	3	1.35
<b>SNP3</b>	53062740	53064215	rs78322266	3	1.72
<b>SNP4</b>	53144124	53146026	rs72930740	3	2.17
<b>Prom_long_a</b>	53254291	53258077	Prom_long	2	1.18
<b>Prom_long_b</b>	53254291	53258077	Prom_long	3	1.52
<b>Prom_long_c</b>	53253089	53257620	Prom_long	5	0.96
<b>Prom_long_d</b>	53253089	53257620	Prom_long	4	3.55

**Table 9 Lentiviral particles for CRISPRi mediated inactivation of gene regulation**

Description of the start and end site of the region targeted by 10 gRNA pools. Each pool contains 2-5 gRNAs. According to the concentration of the viral RNA a specific volume was used for each virus for iNeuron (day-4) transduction.

The isolated DNA was sequenced (Sanger sequencing) at Eurofins genomics with a primer that binds to the U6 promoter (5' ACTATCATATGCTTACCGTAAC 3'). Sanger sequencing results were verified in NCBI nucleotide blast to ensure correct insertion of the desired gRNA sequence (Figure 39 C, D). In case the NCBI blastn algorithm detected mismatches, insertions, or deletions, fluorescent intensity raw data was consulted to verify correct

annotations of nt (Boratyn et al., 2013). Sequence verified bacterial colonies were inoculated to Bioreactor tubes containing 30 ml LB media supplemented with ampicillin. On the next day, plasmid DNA was isolated with a column-based method (QIAGEN Plasmid Plus Midi Kit). Plasmid DNA coding for different gRNAs at the same locus were pooled equimolar (Table 9) before the generation of lentiviral particles (3.2.1) and transduction of day-4 iNeurons(3.2.6.2).

RNA from infected day-49 iNeurons was isolated (3.5.1) and reverse transcribed (3.5.2) by Christina Fürle under my supervision. Analysis of the quantitative expression was performed on 1 – 3 replicates from 2 independent viral transductions and differentiations to validate enhancers responsible for the regulation of *TCF4* expression on mRNA levels. The *TCF4* primer pair binds to the 3'UTR of *TCF4* and amplifies all possible human splicing variants (Sepp et al., 2011). It was generated and tested for human specificity by Rosalba Proce (data not shown). *TCF4* expression was normalized to *MAP2* expression. For the samples of batch16, the expression of *dCas9* and the gRNA scaffold (data not shown) were additionally verified by quantitative PCR readout (Table 10).

Gene	FW primer (5' to 3')	RV primer (5' to 3')
TCF4	GGAAAGTGGACATCGGAGGA	CTTCCTCCAAACCAGCAACC
MAP2	CGAAGCGCCAATGGATTCC	TGAACATACCTTGCAGACACCT
gRNA scaffold	TGCTGGAAACAGCATAGCAAG	GACTCGGTGCCACTTTTCA
<i>dCas9</i>	ATCCCCAGCAGCTTTCAC	AAAGACCGAGGTGCAGACAG

**Table 10** Primer sequence to quantify gene expression after gene knockdown

Primer sequences (FW = forward, RV = reverse) used for quantitative real time PCR.

I performed the qPCR and tested the significance of each gRNA pool with a linear model (lme4 package in R) using batch correction according to the following formula: fold change expression value ~ gRNA pools + differentiation batch (Bates et al., 2015). Normality was tested using GraphPad Prism (Shapiro-Wilk test).

## 3.8 Genome browser representations with IGV

IGV was used to visualize genomic and epigenomic features (Thorvaldsdóttir et al., 2013). ATAC-Seq data was converted by Prof. Dr. Michael Ziller with makeTDF from IGVTools to the tdf format. IGV version 2.4.16 was used to generate graphs in Figure 21, Figure 33, Figure 36, and Figure 40.

## 3.9 Massively parallel reporter assay

### 3.9.1 Selection of the enhancer fragments *in silico*

The fragments to be analyzed with the MPRA assay were selected by Prof. Dr. Michael Ziller according to the following criteria: containing a schizophrenia associated SNP, SNP location in at least one annotated open chromatin region or eQTL and nearby a gene expressed in the (developmental) brain or in iPSC derived neurons.

The cutoff for the p-value from the summary statistics of the genome wide association of each SNP with schizophrenia was set at  $10^{-5}$  (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2014). Candidate SNPs were cross-referenced for their location in open chromatin regions. Therefore, in house generated OCR maps from scATAC-Seq (Dr. Miriam Gagliardi, Vanessa Murek) of human postmortem brain tissue (Brodmann area 10/11) were intersected with candidate SNPs in several annotated cell populations: excitatory cortical layer neurons (layer II-III, layer IV-V, layer VI), inhibitory (parvalbumin (PARV), somatostatin (SST), or vasoactive intestinal peptide (VIP) expressing) neurons, oligodendrocytes and oligodendrocyte precursor cells, microglia, and astrocytes. In addition, candidate SNPs were taken into further consideration when they overlapped with OCRs from iPSC derived forebrain neurons (dNeurons, Dr. Ruhel Ahmad) at day-63 (Yuan et al., 2015) or iNeurons at day-49 or with a H3K27ac chromatin immunoprecipitation sequencing (ChIPSeq) signal in at least one of the below publicly available human (neural) tissue derived data sets (Table 11).

Tissue type	Sample Name	GEO accession number
Brain tissue	Brain, anterior caudate	GSM772832
Brain tissue	Brain, inferior temporal lobe	GSM772995
Brain tissue	Brain, cingulate gyrus	GSM773011
Brain tissue	Brain, mid frontal, Brodmann area 9/46, dorsolateral prefrontal cortex	GSM773015
Brain tissue	Brain, angular gyrus	GSM773016
Brain tissue	Brain, hippocampus middle	GSM773020
Brain tissue	Brain, hippocampus middle	GSM916035
Brain tissue	Brain, substantia nigra	GSM997258
Brain tissue	Brain, substantia nigra	GSM1112778
Brain tissue	Brain, hippocampus middle	GSM1112791
Brain tissue	Brain, angular gyrus	GSM1112807
Brain tissue	Brain, mid frontal, Brodmann area 9/46, dorsolateral prefrontal cortex	GSM1112810
Brain tissue	Brain, anterior caudate	GSM1112811

Brain tissue	Brain, inferior temporal lobe	GSM1112812
Brain tissue	Brain, cingulate gyrus	GSM1112813
Control tissue	Duodenum smooth muscle	GSM772908
Control tissue	Adrenal gland	GSM896163
Control tissue	Heart, aorta	GSM906392
Control tissue	Esophagus	GSM906393
Control tissue	Adipose	GSM906394
Control tissue	Lung	GSM906395
Control tissue	Heart, left ventricle	GSM906396
Control tissue	Pancreas	GSM906397
Control tissue	Spleen	GSM906398
Control tissue	Heart, left ventricle	GSM908951
Control tissue	Gastric	GSM910555
Control tissue	Psoas muscle	GSM910556
Control tissue	Heart, right atrium	GSM910557
Control tissue	Sigmoid colon	GSM910559
Control tissue	Small intestine	GSM915330
Control tissue	Sigmoid colon	GSM915331
Control tissue	Skeletal muscle	GSM916064
Control tissue	Stomach smooth muscle	GSM916065
Control tissue	Adipose nuclei	GSM916066
Control tissue	Ovary	GSM956009
Control tissue	Gastric	GSM1013122
Control tissue	Lung	GSM1013123
Control tissue	Heart, right ventricle	GSM1013124
Control tissue	Thymus	GSM1013125
Control tissue	Adrenal gland	GSM1013126
Control tissue	Esophagus	GSM1013127
Control tissue	Gastric	GSM1013128
Control tissue	Pancreas	GSM1013129
Control tissue	Psoas muscle	GSM1013130
Control tissue	Small intestine	GSM1013131
Control tissue	Spleen	GSM1013132
Control tissue	Bladder	GSM1013133
Control tissue	Colonic mucosa	GSM1112779
Control tissue	Colon smooth muscle	GSM1112780
Control tissue	Duodenum mucosa	GSM1112790
Control tissue	Rectal mucosa	GSM1112795
Control tissue	Rectal smooth muscle	GSM1112796
Control tissue	Kidney	GSM1112799
Control tissue	Rectal mucosa	GSM1112801
Control tissue	Colonic mucosa	GSM1112802
Control tissue	Kidney	GSM1112806
Control tissue	Liver	GSM1112808

Control tissue	Liver	GSM1112809
Control tissue	Spleen	GSM1120338
Control tissue	Adrenal gland	GSM1120339
Control tissue	Psoas muscle	GSM1127171
Control tissue	Small intestine	GSM1127172
Control tissue	Heart, left ventricle	GSM1127173
Control tissue	Muscle, trunk, fetal day115 f	GSM1160189
Control tissue	Adrenal gland, fetal day97 m	GSM1160190
Control tissue	Heart, right ventricle	GSM1220280
Control tissue	Gastric	GSM1227053
Control tissue	Psoas muscle	GSM1227054
Control tissue	Heart, aorta	GSM1227055
Fetal tissue	Thymus, fetal day110 f	GSM1027289
Fetal tissue	Large intestine, fetal day108 m	GSM1058765
Fetal tissue	Small intestine, fetal day108 m	GSM1058766
Fetal tissue	Muscle, leg, fetal day110 f	GSM1058767
Fetal tissue	Stomach, fetal day96 f	GSM1102783
Fetal tissue	Placenta, day 113	GSM1102784
Fetal tissue	Spinal cord, fetal day108 f	GSM1220561
Immortalized cell line	IMR90 cell line	GSM469966
Immortalized cell line	IMR90 cell line	GSM469967
Immune cells	CD4 naive primary cells	GSM772835
Immune cells	CD8 memory primary cells	GSM772880
Immune cells	CD34 mobilized primary cells	GSM772885
Immune cells	CD34 mobilized primary cells	GSM772894
Immune cells	CD4 naive primary cells	GSM772934
Immune cells	CD8 naive primary cells	GSM772949
Immune cells	CD4 memory primary cells	GSM772963
Immune cells	CD8 naive primary cells	GSM772976
Immune cells	CD4 memory primary cells	GSM772997
Immune cells	CD4+ CD25- CD45RA+ naive primary cells	GSM773004
Immune cells	CD4+ CD25int CD127+ Tmem primary cells	GSM916026
Immune cells	CD4+ CD25+ CD127- Treg primary cells	GSM997233
Immune cells	CD4+ CD25- Th primary cells	GSM997239
Immune cells	CD4+ CD25int CD127+ Tmem primary cells	GSM997260
Immune cells	CD19 primary cells	GSM1027287
Immune cells	CD56 primary cells	GSM1027288
Immune cells	CD3 primary cells	GSM1058764
Immune cells	CD8 primary cells	GSM1102781
Immune cells	CD14 primary cells	GSM1102782
Immune cells	CD4 primary cells	GSM1220560
iPSC derived cells	H1 BMP4 derived mesendoderm cultured cells	GSM753425
iPSC derived cells	H1 BMP4 derived mesendoderm cultured cells	GSM753426
iPSC derived cells	H1 derived neuronal progenitor cultured cells	GSM753429

iPSC derived cells	H1 derived mesenchymal stem cells	GSM767341
iPSC derived cells	H1 derived mesenchymal stem cells	GSM767342
iPSC derived cells	H1 derived neuronal progenitor cultured cells	GSM767343
iPSC derived cells	H1 derived neuronal progenitor cultured cells	GSM818031
iPSC derived cells	H1 BMP4 derived mesendoderm cultured cells	GSM864035
iPSC derived cells	H1 BMP4 derived mesendoderm cultured cells	GSM864799
iPSC derived cells	H1 derived neuronal progenitor cultured cells	GSM896162
iPSC derived cells	H1 derived neuronal progenitor cultured cells	GSM956008
iPSC derived cells	hESC-derived CD56+ ectoderm cultured cells	GSM1112824
iPSC derived cells	hESC-derived CD56+ mesoderm cultured cells	GSM1112825
iPSC derived cells	hESC-derived CD56+ ectoderm cultured cells	GSM1112829
iPSC derived cells	hESC-derived CD184+ endoderm cultured cells	GSM1112830
iPSC derived cells	hESC-derived CD184+ endoderm cultured cells	GSM1112831
iPSC derived cells	hESC-derived CD56+ mesoderm cultured cells	GSM1112832
iPSC derived cells	ES_D0	GSM1316300
iPSC derived cells	ES_D0	GSM1316301
iPSC derived cells	ES_D2	GSM1316307
iPSC derived cells	ES_D2	GSM1316308
iPSC derived cells	ES_D5	GSM1316314
iPSC derived cells	ES_D5	GSM1316315
iPSC derived cells	ES_D7	GSM1316321
iPSC derived cells	ES_D7	GSM1316322
iPSC derived cells	ES_D10	GSM1316328
iPSC derived cells	ES_D10	GSM1316329
iPSC derived cells	H7_ESC	GSM1273645
iPSC derived cells	Anterior primitive streak cells	GSM1273650
iPSC derived cells	Definitive endoderm cells	GSM1273655
iPSC derived cells	Anterior foregut cells	GSM1273660
iPSC derived cells	Posterior foregut cells	GSM1273665
iPSC derived cells	Midgut/hindgut cells	GSM1273670
iPSC derived cells	Embryonic stem cells (ESCs)	GSM1521721
iPSC derived cells	Embryonic stem cells	GSM1521726
iPSC derived cells	Neuroectoderm cells	GSM1521730
iPSC derived cells	Neuroectoderm cells	GSM1521735
iPSC derived cells	Early radial glia	GSM1521740
iPSC derived cells	Early radial glia	GSM1521745
iPSC derived cells	Midneurogenic radial glia	GSM1521750
iPSC derived cells	Midneurogenic radial glia	GSM1521755

**Table 11 Overview of tracks to verify location of SNPs in open chromatin**

Table annotates gene expression omnibus (GEO) numbers of open chromatin tracks, the cell or tissue type used and the corresponding human tissue type.

The DevLib contained 2,559 putative enhancer regions overlapping with at least one open chromatin region from H3K27ac ChIPSeq of human postmortem brain, H1 derived neurons,

and H9 derived neural precursors. Included in this open chromatin data set were the following human postmortem brain regions: substantia nigra, anterior caudate, mid frontal lobe, angular gyrus, cingulate gyrus, hippocampus middle, inferior temporal lobe. iPSC derived neural precursor stages included neuroectoderm (NE), early (ERG), and midneurogenic radial glia (MRG). Additionally, the top 900 schizophrenia associated SNPs were included if overlapping with at least one H3K27ac ChIPSeq derived open chromatin region from various tissues. This SNP inclusion criteria resulted in many SNPs which were in open chromatin regions from immune cells. 85 positive control fragments of the mouse *reverse orientation splice acceptor locus (ROSA26)*, the cytomegalovirus (CMV) promoter, the *human eukaryotic translation initiation factor 4A1 (hEIF4A1)*, and active promoters from the following genes were included in the eQTLLib: *human ubiquitin C (UBC)*, *glyceraldehyde 3-phosphate dehydrogenase (GAPDH)*, minimal *EF1alpha*, *phosphoglycerate kinase 1 (PGK)*, *beta kinesin*, and *beta actin*. 52 negative control fragments originated from 6 regions with no epigenomic annotation in any of the considered data sets and a repeat content below 30 %. Those regions were tiled into EFs with a length of 160 bp. 100 negative control fragments were randomly generated with a first order Markov model based on the union of all H3K27ac peak sequences from the considered data sets.

The eQTLLib contained 4,325 SCZ associated SNPs with two alleles each. SNPs were situated in or designated as an eQTL with an FDR  $\leq 0.05$  derived from the GTEX data set specific for the dorsolateral prefrontal cortex (DLPFC) (Fromer et al., 2016). 59 negative and 78 control fragments were selected based on closed chromatin regions and no eQTL effect. The ValLib contained a subset of pEFs from the DevLib and 100 randomly selected SNPs associated with schizophrenia with a p-value greater than 0.5 as negative controls for SNP effect. A subset of positive and negative controls from the DevLib and eQTLLib were also included based on their activity pattern.

### 3.9.2 Low bias amplification and purification of the enhancer fragments

All selected putative enhancer fragments (pERs) were screened for the presence of XbaI, KpnI, and SfiI restriction sites and were replaced with altered nucleotide sequences to prevent restriction enzyme digestion during the cloning steps. Internal cutting sites for XbaI or KpnI were replaced from GGTACC or TCTAGA to CGTACG or AGTAGT, respectively. Each

putative enhancer fragment was included in the original and reverse complement annotation with and without the schizophrenia annotated SNP. The SNP of interest was located in the middle of the sequence fragment.

The lyophilized oligo pools were dissolved in low TE (100 nM). To prepare the enhancer fragments for subsequent cloning steps, the oligo pool was amplified with primers containing Sfil restriction sites and a unique nucleotide sequence of 16 bp length: 5'-GCCAGAACATTCTCTGGCCTAACTGGCCGCTTGACG-3' (MPRA\_v4\_amp\_FW) and 5'-CCGACTAGCTTGGCCGCCGAGGCCGACGCTTCCGATCTNNNNNNNNNNNNNTAGAGGTACC GCAGGAGCCGCAGTG-3' (MPRA\_v4\_amp\_RV). In order to avoid PCR bias and to maintain high sample diversity, an emulsion PCR was used to amplify pEFs in separate reaction compartments. For the emulsion PCR and the amplicon purification, the Micellula DNA Emulsion & Purification Kit was used with a modified protocol. Four emulsion PCR reactions were prepared and aliquoted on ice per library. The oil phase contains three different ingredients: 1100 µl Component 1, 100 µl Component 2, 300 µl Component 3. The water phase was prepared separately and contained: 65 µl 5X Q5® Reaction buffer, 3 µl 100mM dNTP mix, 6.5 µl 100 µM MPRA\_v4\_amp\_FW, 6.5 µl 100 µM MPRA\_v4\_amp\_RV, 3.5 µl Q5® High-Fidelity DNA Polymerase, and 221 µl DNasefree water. Of the water phase mix, 47 µl were transferred into a well containing either 3 µl water, or 2.56 µl water and 0.44 µl oligo pool as a negative or positive control, respectively. To the remaining water phase mix, 2.0 µl oligo pool, 9.5 µl water and 2.25 µl acetylated BSA (20 mg/ml) were added and mixed by pipetting. Afterwards 100 µl water phase were mixed with 600 µl oil phase at 4 °C for 5 min on a vortex at maximum speed. For the thermocycler reaction 50 µl of the emulsion PCR

reaction were aliquoted into a 96-well plate. Amplification was performed at 95 °C for 30 sec, and 25 cycles of 95 °C for 10 sec, 58 °C for 30 sec, 72 °C for 1 min, and a final elongation step of 5 min at 72 °C.

175 µl of the emPCR product were mixed with 1 ml isobutanol (2-methylpropan-1-ol) to release the PCR products. After an additional vortex step at RT, 400 µl of Orange-DX buffer were added, inverted for 3 min at RT and the colored water phase was spun down (2 min, 16,000 rcf, RT). The colored water phase and the milky interphase were applied onto previously activated columns (incubation of 40 µl Activation Buffer DX for 5 mins). The column was washed consecutively with 500 µl of Wash-DX1 buffer and 650 µl of Wash-DX2 buffer. The flow-through was discarded each time. Each column was eluted with 25 µl preheated (80 °C) Elution-DX buffer.

### 3.9.3 Cloning of the barcoded enhancer fragments into a cloning vector

To ligate the amplicon into a cloning vector (pMPRA1), the insert and the cloning vector had to be digested by the Sfil restriction enzyme. The cloning vector was amplified in ccdB survival bacteria. This vector allows a selection for backbones with the insert in non-ccdB survival bacteria. Restriction digests were carried out for 5 h at 50 °C (Table 12). To avoid religation, the cloning vector was subjected to a dephosphorylation step at 37 °C for 30 min (5 µl rSAP per digest). All enzymes were heat inactivated at 80 °C for 10 min.

Ingredients	Volumes for PCR amplicon digest in µl	Volumes for pMPRA1 digest in µl
DNA	160	20
CutSmart buffer	20	20
Sfil enzyme	12	5
Water	8	155

**Table 12** *Restriction digest and dephosphorylation of the pMPRA1 plasmid*

pMPRA1 and the emPCR product were digested with the Sfil restriction enzyme in 1x CutSmart buffer.

The digested amplicon or backbone were run on a 2 % or 0.8 % gel (E-Gel EX) to visualize the size of the fragments and cut out the correct fragments. The fragments were isolated from the gel with a column-based approach (MinElute Gel Extraction Kit, or QIAquick Gel Extraction Kit) following manufacturer's instructions. Only the digested emPCR of the ValLib was purified with a double-sided bead cleanup (0.45x volume, 0.35x volume). In brief, 0.45x the volume of the pooled PCR product of magnetic beads was bound to the PCR product for at least 10 min at RT. After magnetic separation, the supernatant was incubated

for at least 10 min at RT with 0.35x the volume of the original pooled PCR product. The magnetic beads were washed twice with 200 µl 80 % ethanol. To elute the dsDNA of the correct size, the beads were dried, and dsDNA was eluted with 20 µl low TE buffer. The concentration of the fragments (digested amplicon and backbone) was measured by Qubit™ dsDNA HS Assay Kit according to the manufacturer's protocol. Ligation was performed overnight at 4 °C or for 5 h at RT (Table 13).

Ingredients	For negative control	For ligation
T4 DNA Ligase Reaction Buffer	1.25 µl	2.50 µl
T4 Ligase	0.50 µl	1.00 µl
Backbone vector	50 ng	100 ng
Enhancer-barcode	0 ng	200 ng
Water	Add to 12.5 µl	Add to 25 µl

**Table 13 Reaction to ligate the barcoded emPCR product into pMPRA1**

T4 Ligase was used to ligate the Sfil digested emPCR product with pMPRA1 in 1x T4 ligation buffer.

Before the transformation of the ligated product into One Shot™ TOP10 Electrocomp™ E. Coli bacteria the ligation enzyme was inactivated at 60 °C for 15 min. To increase the transformation efficiency and reduce the number of transformations, the ligation for the ValLib was desalting via osmosis (MF-Millipore™ Membrane Filter) for 30 min at RT before the electroporation.

### 3.9.4 Transformation of the library and transformation efficiency

1.3 – 3 µl of the ligated vector were transformed into 50 µl One Shot® TOP10 Electrocomp™ E. coli bacteria with an electroporator using a voltage shock of 1.8 kV (Eppendorf Eporator®). Immediately afterwards, bacteria were gently flushed out of the cuvette with 1.5 ml prewarmed SOC media. The freshly transformed bacteria were shaken for 1 h at 37 °C on a thermomixer at 300 rpm.

Dilution	µl LB media	µl bacteria	Origin of bacteria solution
1:100	990	10	1.5ml stock of transformation
1:1000	270	30	From 1:100
1:5000	240	60	From 1:1000
1:10000	270	30	From 1:1000
1:20000	285	15	From 1:1000

**Table 14 Serial dilutions to estimate the number of transformed bacteria colonies**

Dilutions of the transformed bacteria in SOC media were performed to count the number of colony forming units on Agar plates to estimate the number of necessary transformations to achieve the required complexity of the plasmid libraries.

LB media and agar plates, both containing ampicillin (100 µg/ml) as a selection reagent, were prewarmed to 37 °C to facilitate bacteria survival. In order to estimate the number of successful transformations, serial dilutions of the SOC medium containing the transformed

bacteria were plated onto the agar plates (Table 14). The remaining solution was split evenly into two 50 ml bioreactor tubes containing 25 ml LB media with ampicillin and were shaken at 37 °C (190 rpm) for 16 h. The bacteria were pelleted at 6,000 rcf at RT in a centrifuge and stored at -20 °C for further analysis. Colony forming units (CFU) were calculated according to the following formula:

$$CFU = \text{No. of colonies} * \text{dilution factor} * \frac{1000 \mu l}{\text{plated volume in } \mu l} * \text{total volume in ml}$$

The isolation of the plasmid DNA from the bacteria was performed for each transformation separately with a column-based approach (QIAGEN Plasmid Plus Midi Kit) following the high-yield protocol. Each transformation was eluted in 200 µl EB provided with the kit. The concentration was measured with the NanoDrop™ 2000 to pool the midi preparations according to their CFU:

$$\mu l \text{ DNA} = \frac{CFU * 100 \mu g}{1000000 * \text{concentration DNA in } \mu g/\mu l}$$

### 3.9.5 Subcloning the reporter gene into the initial library for the intermediate library

After verification of the initial library complexity, a minimal promoter and a gene were cloned between the enhancer and the barcode. For this purpose, an expression vector for nLuc under the regulation of a minimal promoter (minP) kindly provided by Dr. Davide Cacchiarelli (TIGEM, Italy) was used. The vector is a modified version of pNL3.1 with a deletion of the Sfil site in the nano luciferase gene (nLuc).

Ingredients	MinP-nanoLuc	Cloned vector	Cloned vector - XbaI test	Cloned vector - KpnI-HF test
DNA	10 µg	20 µg	0.5 µg	0.5 µg
CutSmart buffer	10 µl	20 µl	1 µl	1 µl
XbaI enzyme	5 µl	10 µl	0.5 µl	-
KpnI-HF enzyme	3 µl	6 µl		0.3 µl
Water	Add to 100 µl	Add to 200 µl	Add to 10 µl	Add to 10 µl

**Table 15 Digest of the nanoLuciferase fragment and pMPRA1-enhancer-barcode**

pMPRA1-enhancer-barcode and the nanoLuciferase containing pNL3.1 plasmid were digested with the KpnI high fidelity and XbaI restriction enzyme in 1x CutSmart buffer. Test restriction digests of pMPRA1-enhancer-barcode were performed with only one enzyme.

Restriction digests of the modified pNL3.1 vector and the vector generated previously with the barcoded pEFs were carried out for 5 h at 37 °C. To verify that both enzymes cut in the previously cloned vector, a KpnI and XbaI control digest were performed in parallel (Table 15). Dephosphorylation of the previously cloned vector was performed at 37 °C with 5 µl

rSAP for 30 min. All enzymes were inactivated at 60 °C for 15 min. Fragments were size selected and purified from an agarose gel as described beforehand. After measuring the concentration by Qubit, the fragments were ligated overnight at 4 °C (Table 16) and heat-inactivated at 60 °C for 15 min before the transformation. 1.8 – 3 µl were transformed into electro competent bacteria as described beforehand. Several transformations had to be performed in order to achieve a CFU count higher than 10,000 times the number of elements in the ordered pool.

Ingredients	Negative control	Ligation
T4 DNA Ligase Reaction Buffer	0.75 µl	1.50 µl
T4 Ligase	0.50 µl	1.00 µl
Backbone vector	50 ng	100 ng
MinP-nano Luciferase	-	200 ng
Water	Add to 7.5 µl	Add to 15 µl

**Table 16 Ligation of pMPRA1-enhancer-barcode and minimalPromoter-nanoLuc**

T4 Ligase was used to ligate the KpnI / XbaI digested pMPRA1-enhancer-barcode backbone with the insert coding for a minimal promoter and the reporter gene nanoLuciferase in 1x T4 ligation buffer.

### 3.9.6 Subcloning enhancer-minP-nLuc-barcode fragments into lentiviral vector

The plasmid DNA from each transformation was isolated separately and pooled according to CFU count as described in 3.9.4. Afterwards, the enhancer-minP-nLuc-barcode fragments were cloned into a lentiviral expression vector (pMPRAgenti1) in exchange of the ccdB selection site. Sfil restriction digests were incubated for 5 h at 50 °C (Table 17), followed by a rSAP incubation (5 µl only for the pMPRAgenti1) for 30 min at 37 °C. All enzymes were inactivated at 85 °C for 15 min.

Ingredients	Enhancer-minP-nLuc-barcode fragment	pMPRAgenti1
DNA	20 µg	25 µg
CutSmart buffer	15 µl	10 µl
Sfil enzyme	10 µl	5 µl
Water	Add to 150 µl	Add to 100 µl

**Table 17 Restriction digest of pMPRAgenti and the MPRA coding insert**

pMPRAgenti and the MPRA coding insert (enhancer-minP-nLuc-barcode) were digested with the restriction enzyme Sfil in 1x CutSmart buffer.

DNA fragments were size selected over an agarose gel (0.8 %) and purified and analyzed using the methods mentioned previously.

Ingredients	Negative control	Ligation
T4 DNA Ligase Reaction Buffer	0.75 µl	1.50 µl
T4 Ligase	0.60 µl	1.20 µl
Backbone vector	100 ng	200 ng
Enhancer-barcode	-	100 ng
Water	Add to 7.5 µl	Add to 15 µl

**Table 18 Ligation reaction of digested MPRA coding insert and pMPRAgenti**

T4 ligase was used to ligate the Sfil digested enhancer-minP-nLuc-barcode into the pMPRAgenti plasmid.

The ligation was carried out for at least 5 h at RT and heat inactivated at 65 °C for 10 min (Table 18). To increase the yield and decrease the recombination activity between the repetitive elements, special library preparation bacteria were used for the lentiviral transformation (Endura electro competent bacteria). The procedure of the transformation is the same as above mentioned with the exception that only 25 µl bacteria were transformed with 1.5 µl - 2.8 µl ligation. The number of transformations necessary was depending on the CFU from each transformation. The CFU count of all transformations had to be above 25,000 times the number of elements in the original oligo pool. The pooled DNA (according to CFU as described beforehand) from the midipreps from the lentiviral vector were used to prepare lentiviral particles (see 3.2.1.2).

### 3.9.7 Association by sequencing of the barcodes with their respective enhancer

As a next step, we investigated the association between the enhancers and the barcodes present in the pool. Therefore, the inserted fragments were amplified to sequence them on a MiSeq (Illumina) sequencing machine. The same forward primer (M30\_FW 5' AATGATACGGCGACCACCGAGATCTACACTTTCCCTACACGACGCTCTCCGATCT 3') was used for all barcode amplifications. The reverse primer (5' CAAGCAGAAGACGGCATAC GAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTTCCGATCTGGTGCCAGAACATTCT CTG 3') was barcoded so that several libraries could be sequenced at the same time. For each sample 112.5 µl Q5 Hot Start High-Fidelity 2x Master Mix were mixed with 11.25 µl 10 µM M30\_FW, 11.25 µl 10 µM barcoded RV primer, 800 pg plasmid. Water was added to a reaction volume of 225 µl. The amplification was performed in 25 µl PCR reactions. Each pool was amplified in eight separate PCR reactions. Amplification was performed at 98 °C for 30 sec, and 16 cycles of 98 °C for 10 sec, 60 °C for 30 sec, 72 °C for 20 sec, and a final elongation step of 2 min at 72 °C.

The PCR product was 367 bp in length and purified with Agencourt Ampure XP magnetic beads with a double-sided selection (0.55x, and 0.35x the original volume of the DNA). The correct size and the concentration of the library were analyzed with the Bioanalyzer (High Sensitivity DNA Analysis). Library quantification was performed by qPCR on the Illumina sequencing adapters according to manufacturer's protocol (KAPA Library Quantification) on a Roche LightCycler® 480 System. In brief, the purified PCR product was diluted 10,000x, 100,000x and 1,000,000x in water. 6 µl of the Master mix with 2 µl of each standard, each

dilution, and negative controls (water) were analyzed in triplicates. Amplification was performed at 95 °C for 5 min, and 40 cycles of 95 °C for 30 sec, 60 °C for 45 sec. A melting curve analysis was performed from 60°C to 95 °C with a gradual temperature increase of 1 °C per sec. 6-10 % PhiX were spiked into the final library for sequencing. Paired-end (150 bp each side) sequencing was performed with the MiSeq Reagent Kit (v2, 300-cycles).

Data analyses of the MPRA barcode associations were performed by Prof. Dr. Michael Ziller. Reads were aligned to a custom build reference “MPRA genome” containing the synthesized pEF sequences using bwa version 0.7.17 in paired end mode (H. Li & Durbin, 2009). Subsequently, the resulting .bam files were converted to .txt files retaining high quality reads. Reads were only considered if they aligned to the custom MPRA genome, contained the investigated SNP in the center of the pEF and the sequence contained less than 5 clipped or mismatched bases. Retained reads were collapsed onto individual barcodes. Barcodes were only included for the final analysis if at least two reads contained the same barcode, and each barcode was associated with one unique pEF.

### 3.9.8 Generation of sequencing libraries from different cell types

The activity of the enhancers selected for the MPRAs was analyzed in different cell types (Table 19). NPCs were infected at day 5 during their differentiation and kept 3 more days in culture. iNeurons were kept in culture for 45 days after they were infected with the MPRA library on day-4 during the differentiation. Primary mouse neurons were infected while already plated and kept in culture for 10 additional days.

Cell type	Plated cells	Species	Conditions
NPCs	8 mil	Human	Baseline
iNeurons	15- 25 mil	Human	Baseline, hKCl
Primary mouse neurons	8 mil	Mouse	Baseline, silent, excited

**Table 19 Overview of the cell types and stimulation conditions used for the MPRA**

3 different neural cell types (2 human and 1 mouse derived) were analyzed in the MPRA after different stimulation conditions.

All cell types were directly lysed during the harvest. The supernatant of the cells was aspirated, and the lysis buffer was directly added to the dishes. The lysis buffer of different dishes was pooled to samples and the lysate was homogenized by a centrifugation step (1 min, RT, 6000 rcf) through QIAshredder columns. Total RNA from the different cells was isolated following the instruction of the Quick-RNA Miniprep Kit. For each sample one column was used. Only for the virus (20 µl each sample), one column of the Quick-RNA

Microprep kit (DevLib) or Quick RNA Viral kit (ValLib and eQTLlib) was used. DNA contamination was reduced by performing a column-based DNase digestion. All columns were eluted in 100 µl per sample but the virus which was eluted in 32 µl.

Before the mRNA isolation (Poly(A)Purist MAG kit) from the total RNA samples was started, the concentration and purity of the samples was measured with the NanoDrop. According to this concentration the volume of the purification beads was calculated individually for each sample. Before the beads were incubated with the total RNA, the beads were washed twice with WashSolution1 (0.5x volume of the original bead volume) and were resuspended in 2xBindingBuffer (1x volume of the original bead volume). After the addition of 2xBindingBuffer to the samples, the mix was incubated at 70 °C for 5 min and at RT for 1h on a rotator (SB3, Stuart). Each washing step of the magnetic beads consisted of the application of washing solutions (2x 100 µl WashSolution1, 2x 100 µl WashSolution2) as well as a gentle mix of the beads with the respective solutions. After the last wash, beads were dried and resuspended in 35 µl preheated RNA Storage Solution (80 °C). mRNA purity and concentration were measured by NanoDrop. The generated mRNA or the RNA from the viruses were reverse transcribed into cDNA. 30 µl mRNA were mixed with 2 µl SuperScript IV reverse transcriptase and 8 µl SuperScript™ IV VILO™ Master Mix. Reverse transcription was performed at 25 °C for 10 min, 50 °C for 3 min, and 85 °C for 5 min.

MPRA barcodes				
CTCTACTT	GATCGTGT	TCGGAACA	CGATCATG	TGGTAACG
CCTTACCT	CCATTGTT	GATACAGT	TAGATCCT	TCGCCAGA
ATACCTGT	CTAACTGG	CAGTTGGT	TCTGGACC	TTGAGCCT
CCTGGTAG	TAAGCATG	AGATGTGC	GTCGAGCA	GAATTGCT
TCCTTGGT	AATGCGTT	TGCGACCT	AGTTGCTT	AGGCGAAG

**Table 20** Summary table of the barcodes used for MPRA library indexing

25 barcodes with a length of 8 bp were used to amplify and single index the MPRA barcode libraries.

In order to analyze the activity of each enhancer in the pool separately, the barcodes were amplified in a quantitative manner from the cDNA. The PCR protocol described in 3.9.7 was adopted but the cycle numbers was increased to 22. The cDNA samples generated from the viruses were diluted (1:100) before the PCR reaction and the number of PCR reactions was decreased to 12. For all the other samples, 0.5 µl cDNA were used per reaction (8 reactions of 25 µl per sample). M30\_FW and MPRA\_RV (5'-CAAGCAGAAGACGGCATA CGAGATNNNNNNNTGACTGGAGTTCAGACGTGTGCTTCCGATCTTATCGACGAGCGCCT G, Table 20) were used to amplify the barcodes in 1x Q5 master mix (3.9.7). 25 different

reverse primers (MPRA\_RV) were used barcoding the individual samples (Table 20). The PCR product was 259 bp in length and purified with Agencourt Ampure XP magnetic beads with double sided size selection (0.5x and consecutively 0.35x bead volume of the PCR volume). Libraries were pooled equimolar according to their molarity measured by the Bioanalyzer or the Tapestation. In case of primer dimers or remaining cDNA contamination, the pooled library was repurified according to the above stated protocol and the concentration was reassessed on the Bioanalyzer high sensitivity chip. The sequencing was performed on a NovaSeq6000 at the NGS core facility at the Max-Planck Institute for Molecular Genetics (Berlin).

Data analysis of the MPRA barcode recovery in the different analyzed cell types was performed by Prof. Dr. Michael Ziller. In brief, 3' mRNA sequencing data were processed to only retain those reads with less than 5 nt below a phred score of 20 and plasmid backbone sequence elements such as restriction sites at the expected positions of the read. Additionally, only those barcodes (16 bp length) were considered which were part of the final recovered enhancer barcode association table from 3.9.7. I plotted the barcode abundance plots in R (version 4.0.3) with ggplot2, and ggVennDiagram and exported them as pdf files with ggpubr.

Depending on the subsequent analysis, all retained barcodes were collapsed on the (allele specific) pEF level for further analysis. For quality control, each replicate of the MPRA library in the same cell type was analyzed separately and normalized to the library's individual read depth. The plasmid association library was normalized in an analogous way. This plasmid library was used to normalize the pEF abundance in the different analyzed conditions: log2 ratio of mRNA cpm over plasmid based cpm (+1). pEFs were filtered according to a minimal pEF abundance across conditions of pEFs  $\log_2 (\text{cpm} + 1) \geq 4$  (5 for eQTLLib). This excluded low abundant pEFs with a low reproducibility between libraries and conditions.

### 3.9.9 Computational analysis of differentially active elements in the MPRA

Data analysis of the MPRA differential activity was performed by Prof. Dr. Michael Ziller. MPRA analysis to investigate allele and conditions specific activity of pEFs was performed

with MPRAalyze R package (Ashuach et al., 2019). Raw count data per pEF were used as input for each MPRA library separately to the generalized linear model (GLM).

A condition specific normalization factor for the activity of pEFs (independent of the allele status) was estimated by the size factors for each replicate of RNA and DNA counts separately (estimateDepthFactor function). Negative control sequences were included. Quantification of pEF activity was performed with GLMs with the analyzeQuantification function (condition as factor for RNA design, no factor for DNA design). Activity was described by the resulting alpha values. Condition specific p-values of pEF activity are pairwise calculated between single pEFs and the 147 negative controls (testEmpirical function). Differential activity testing between conditions for each pEF was performed with pairwise comparisons (analyzeComparative). Condition was the only factor in the full RNA design and one factor in the reduced GLM design for each condition pair. Those pEFs with an FDR less than 0.01 were defined as differentially active between conditions.

Allele specific analysis utilized fitted GLMs and considered each variant pair of one pEF. Differential allele analysis was based on the WaldTest separately for each condition (analyzeComparative function). Replicate and status or only replicate were included in the full or reduced model, respectively. Reproducibility, sensitivity, and specificity of the MPRA libraries was empirically evaluated utilizing the high complexity ValLib (> 400 barcodes per pEF). The ValLib contained 215 pEFs from the DevLib and 100 randomly selected SNPs not associated with schizophrenia to empirically calibrate FDR and p-value thresholds for differential allele activity. Therefore, the fraction of randomly selected SNPs of the ValLib which were assigned differential allele activity were correlated with the FDR cutoff and the minimum pEF activity p-value threshold. For this comparison, the mouse baseline condition of the ValLib was used to calculate the MAD based p-value threshold over the negative control element activity distribution. According to this analysis, the FDR cutoff was set to 0.01 for the differential allele test and the MAD p-value cutoff for minimal activity was set to 0.005 for the ValLib. Those thresholds were used to annotate the 215 pEFs of the ValLib to generate a ground truth emVar set. This ground truth emVar set was utilized to evaluate true and false positive rate (TPR and FPR). A pEF was defined as active if at least one of the alleles had a MAD p-value below the minimal threshold of 0.005. An emVar was annotated if the differential allelic activity was below a p-value of 0.005. Condition specific emVars were only detected in one specific condition (cell types). Stimulus dependent emVars were

investigated only in neuronal activity active culture systems: iNeurons hKCl, iNeurons baseline, mouse primary neurons baseline, and mouse primary neurons excited.

### **3.9.10 Intersection of differential active MPRA regions with genomic data sets**

For further enrichment analysis and intersections with other data sets, SNP activity results (emVar status) from the DevLib and eQTLLib were merged (containing the DevLib results in case the SNP was tested in both libraries). Analyses were performed on hg19 by Prof. Dr. Michael Ziller.

Overlap of emVars with open chromatin region (ATAC, H3K27ac) were considered as overlap if the emVar was located within 200 bp to the open chromatin region. Enrichment of specific emVars was tested against all other non-emVars using Fisher's exact test. Overlap with an eQTL was considered if emVars directly overlapped with the eQTL region. Results are reported as odds ratios, p-value, or multiple testing corrected p-values. Additionally, concordance of effect direction was tested between eQTL measurements of the DLPFC (Fromer et al., 2016) and MPRA assay. The sign of the effect of the alleles on all associated genes in the eQTL data set (FDR < 0.05) were compared for each MPRA condition separately. Concordant directionality was defined that at least one eQTL gene was affected in the same direction as the MPRA results indicated.

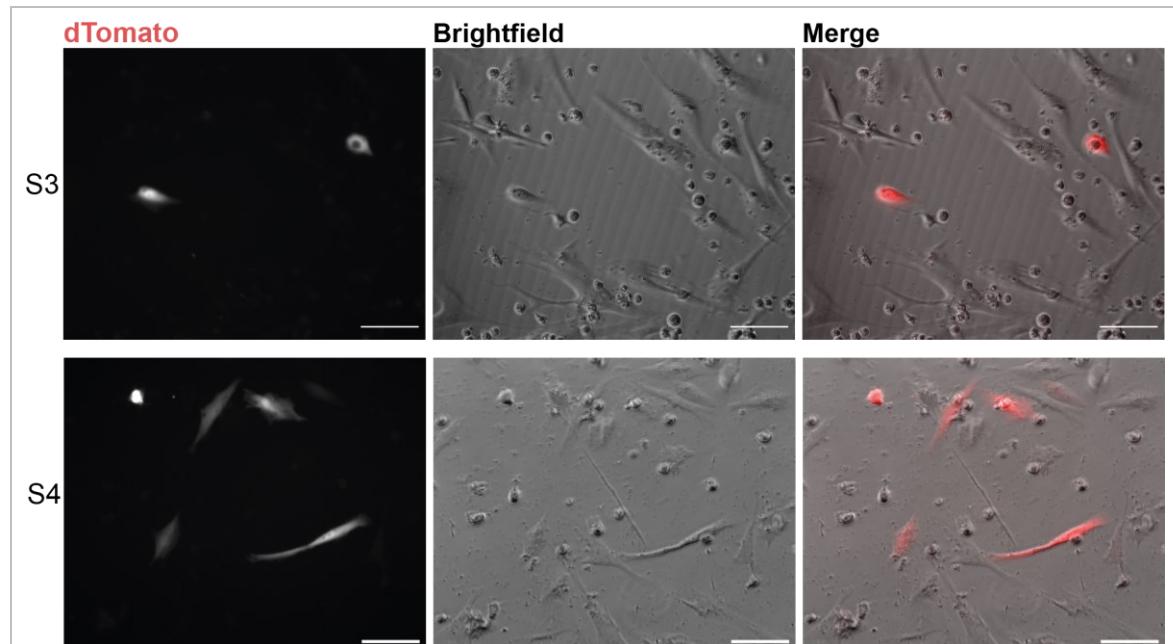
emVars were associated with putative target genes over several sources: 1) Intersection of end-coordinate of loops provided by the promoter centric HiC obtained from PsychENCODE (<http://resource.psychencode.org/>) (D. Wang et al., 2018). 2) Intersection of emVars with eQTL from the DLPFC and annotating the target gene to the emVar as the gene regulated by the eQTL. 3) Intersection of emVar loci with open chromatin marks (H3K27ac ChIP) that are linked to promoter of genes based on HiC data using the abc gene model (Nasser et al., 2021).

## 4 Results

### 4.1 Generation of induced pluripotent stem cells from fibroblasts

One aim of this thesis was to generate human iPSCs from patients and healthy controls to directly analyze the genetic impact of schizophrenia on disease development in a case - control study in disease relevant cell types. This part of the thesis will focus on the work of reprogramming primary human fibroblasts and quality control of iPSCs. The reprogramming of human peripheral blood mononuclear cells (PBMCs) was a team effort of the wet lab in the Ziller group. To ensure a successful reprogramming of the fibroblasts, iPSCs were investigated for their potential to form embryoid bodies and their expression of the pluripotency markers OCT4 and NANOG. Furthermore, the analysis of the integrity of the genome of the reprogrammed iPSC lines guaranteed that downstream assays of those lines were a result of the schizophrenia related genome and not reprogramming induced genomic aberrations such as CNVs.

#### 4.1.1 Generation of pluripotent stem cells from human skin fibroblasts

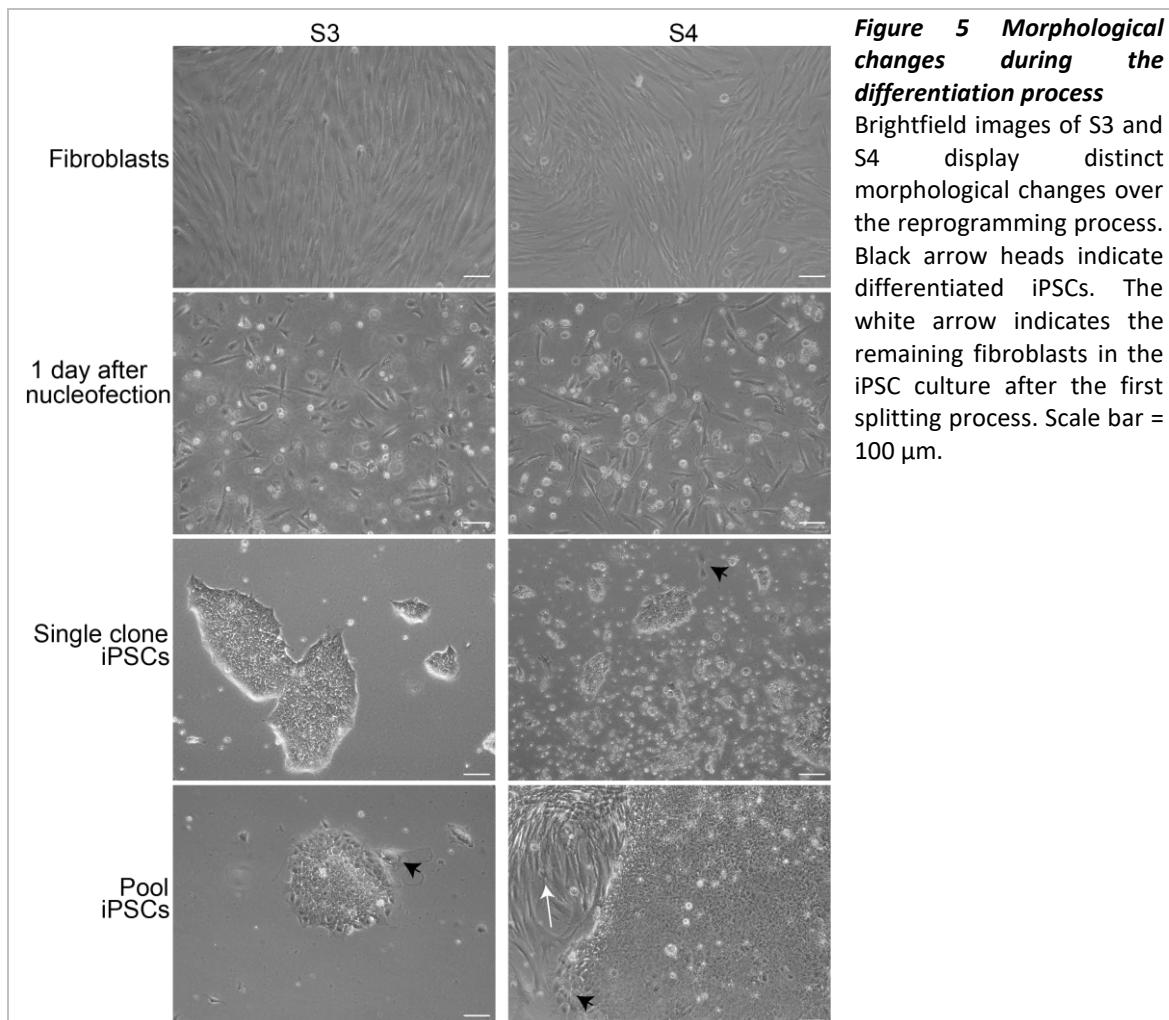


**Figure 4** Visualization of successfully nucleofected primary human fibroblasts

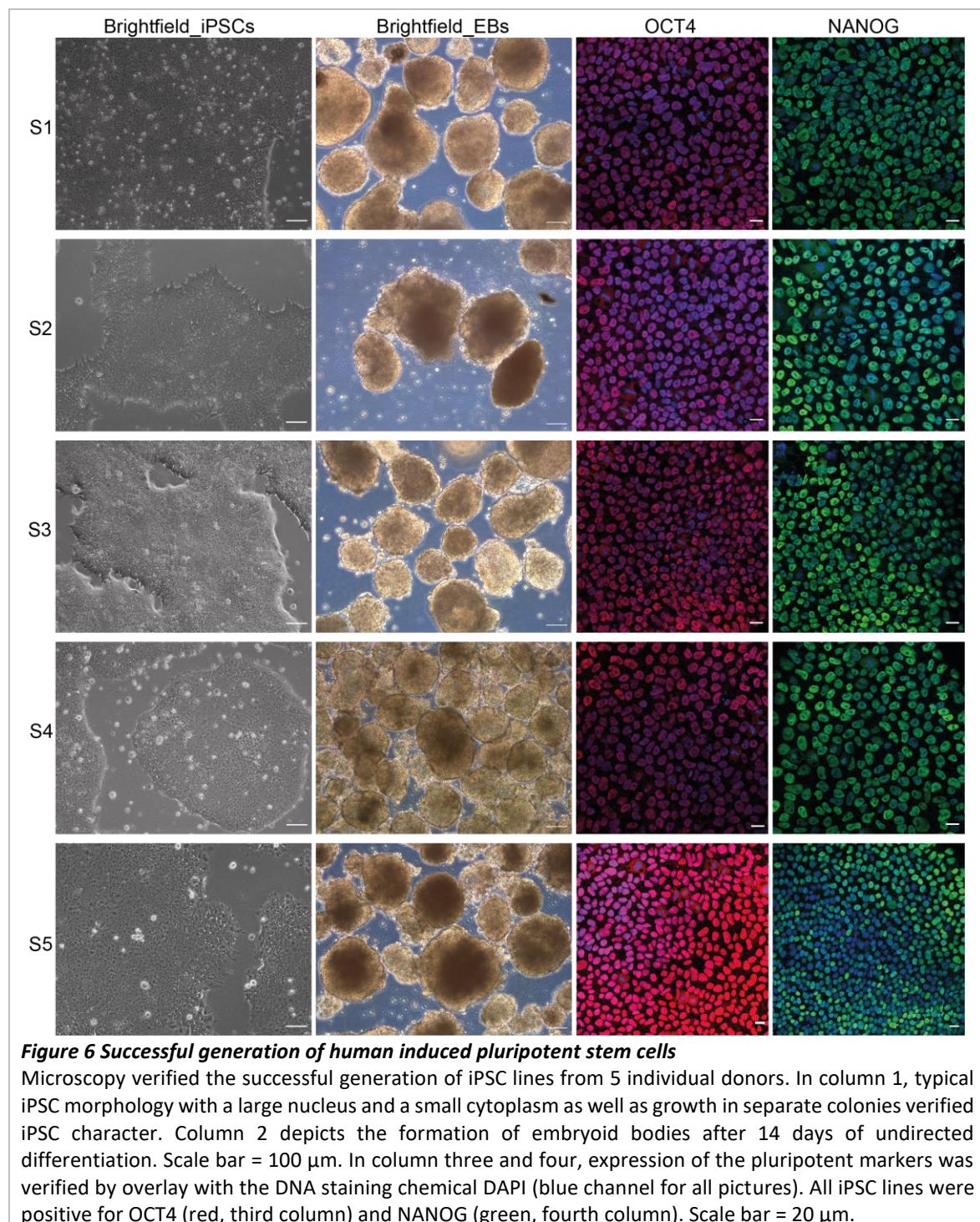
2 or 8 days after nucleofection of the CoMIP plasmid, expression of red fluorescence protein (dTOMATO, white arrow heads) was visualized by epifluorescent microscopy in the here depicted cell lines S3 or S4, respectively. The fluorescent protein was encoded by the reprogramming plasmid also coding for *SOX2*, *POU5F1*, *MYC*, *KLF4*, and a short hairpin RNA (shRNA) against *TP53*. Merged brightfield and fluorescent image of the same location shows a small number of successfully transfected cells. Scale bar = 100 µm.

Primary fibroblasts from 5 donors diagnosed with schizophrenia were expanded and nucleofected with a plasmid coding for the reprogramming factors: *SOX2*, *POU5F1*, *MYC*,

*KLF4*, and a short hairpin RNA (shRNA) against *TP53*. Fibroblast media was gradually exchanged for stem cell media while monitoring the progressive morphological changes throughout the reprogramming process by brightfield microscopy. Due to the fact that the reprogramming plasmid codes for a red fluorescent protein (dTomato), successful nucleofection was verified by epifluorescent imaging (Figure 4) after nucleofection.



Three days after nucleofection the primary human fibroblast started to alter their morphology from elliptical to rounded. The two characteristic nucleoli of the iPSCs emerged 6 days after nucleofection and colony formation was observed around day-9 (Figure 5). Newly generated iPSCs expanded until day-13 to 16, when single clones were picked, and the remaining colonies were split for the pooled cell line. Each nucleofection generated between 5 – 20 individual colonies. Remaining fibroblasts and partly differentiated or reprogrammed cells were selected against by several rounds of splitting as well as by selection for the stem cell surface marker *Tra1-60* (glycoprotein on the podocalyxin protein) with magnetic beads (Bharathan et al., 2017).



After several passages and cryopreservation, the individual iPSC lines were subjected to various quality control checks to verify pluripotency. Brightfield images displayed growth of individual colonies as well as iPSC morphology (round cytoplasm, large nucleus, and distinct nucleoli) for all five reprogrammed cell lines. The five reprogrammed cell lines were capable to form embryoid bodies (Figure 6). Further validations of the generated iPSC lines were performed with immunocytochemistry for the stem cell markers OCT4 and NANOG. All cell lines expressed those transcription factors in the nucleus. The nucleus was visualized

by 4',6-diamidino-2-phenylindole (DAPI), which binds specifically to the minor groove of double-stranded DNA (dsDNA).

#### 4.1.2 Karyotypic analysis reveals several genomic aberrations caused by the reprogramming process

In order to identify iPSC lines with genomic aberrations caused by the reprogramming process, we analyzed the SNP abundance throughout the genome with a SNP microarray. To account for the genomic aberrations already present in the patient derived fibroblasts, the reprogrammed cell lines were compared to the genome of the original fibroblasts. Changes in copy numbers were called if more than 10 SNPs in a window larger than 200 kb in one genomic locus showed an increase (duplication) or decrease (deletion) in fluorescent signal on the array (Danecek et al., 2016).

Cell line	Pool/clone	Passage at gDNA isolation	CNV between 0.2 and 1 Mbp	CNV > 1 Mbp	CNV on chromosome number	Usage
S1	Single clone	7	1	-	3	Kept, not used
S1	Single clone	7	-	-	-	Used
S1	Pool	7	-	-	-	Kept
S2	Single clone	7	6	-	2, 2, 2, 3, 3, 14	Kept, not used
S2	Pool	6	1	-	2	Used
S3	Single clone	7	-	-	-	Used
S3	Single clone	7	-	-	-	Used
S3	Pool	6	4	-	X, X, X, X	Kept, not used
S4	Pool	17	7	-	1, 3, 3, 4, 4, 6, 15	Not used
S4	Single clone	19	1	-	7	Not used
S4	Pool	7	2	1	8, 8, 8	Not used
S4	Pool	10	3	-	1, 11, 11	Not used
S5	Single clone	7	1	-	-	Kept
S5	Pool	7	1	-	16	Used

**Table 21 Annotation of copy number variations in reprogrammed iPSC samples**

The genome of reprogrammed single clones and pools were investigated with microarrays for the appearance of copy number variations (CNVs) in comparison to the fibroblast genome. CNVs were categorized according to their base pair length into two groups below and above 1 Mbp. Only those cell lines without a CNV were used for experiments.

For the reprogramming of the five patient derived fibroblasts reported in detail in this thesis, every donor presented in at least one of the corresponding iPSC lines an additional CNV in the karyotype. CNVs longer than 0.2 Mbp and shorter than 1 Mbp occurred with a higher frequency than CNVs spanning more than 1 Mbp. For those cell lines reported here, only one cell line (S4) showed a deletion larger than 1 Mbp (Table 21). Only iPSC lines without CNVs were used for further analysis if possible.

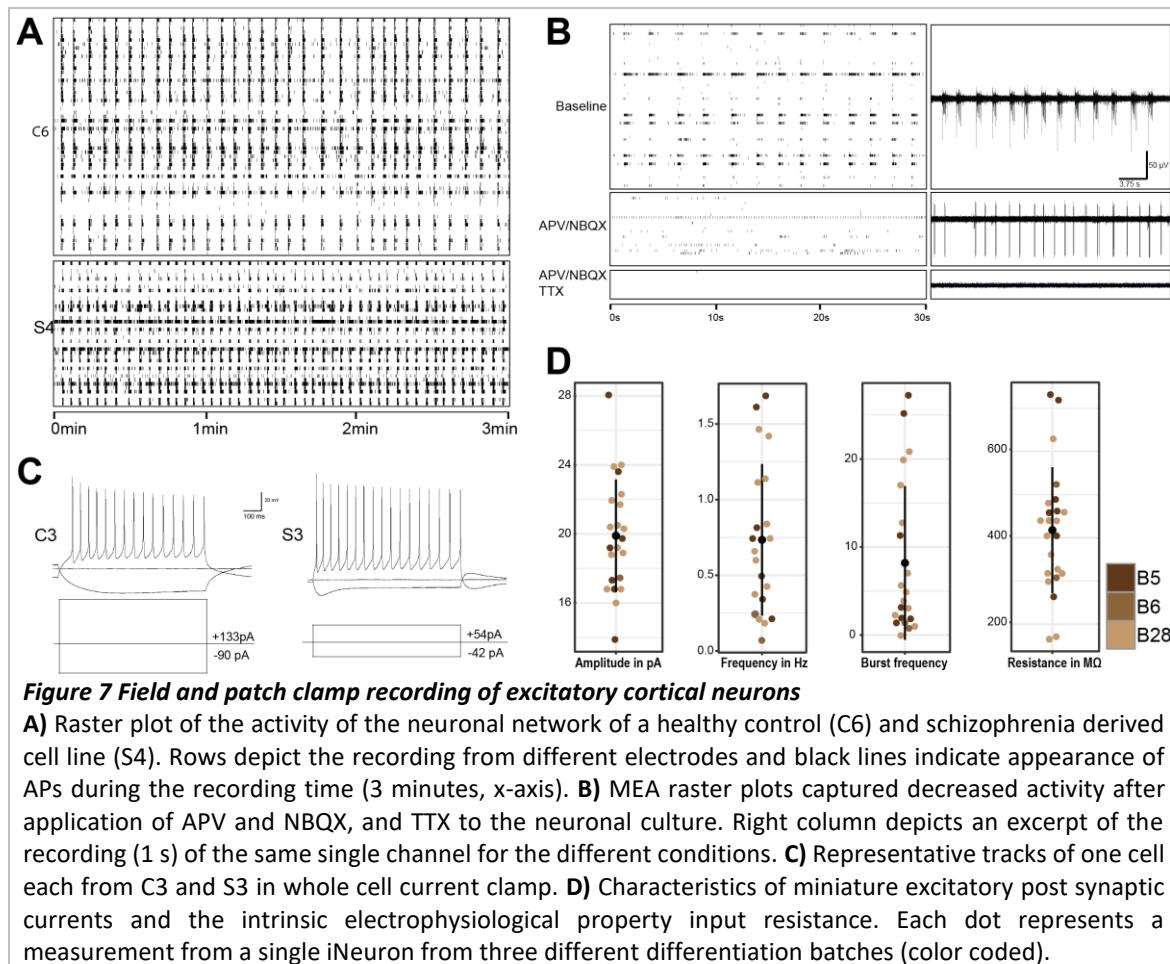
## 4.2 Characterization of the neuronal identity of iNeurons

After successfully reprogramming schizophrenia and healthy control derived primary human cells into iPSCs, the next objective of this thesis was to generate excitatory forebrain neurons in order to investigate schizophrenia genetics in a disease relevant cell type. The here reported neuronal differentiation protocol utilizes lentiviral overexpression of mouse Ngn2 and patterning factors. Differences between published protocols and the protocol applied in this study occurred because neuronal differentiation of iPSCs for this study started before the respective alterations to the original protocol (Y. Zhang et al., 2013) were published (M. Chen et al., 2020; Lin et al., 2020; Nehme et al., 2018). In consideration of the fact that the iNeurons described here are differentiated with an altered iNeuron protocol, the homogeneity of marker presence in the generated neuronal cultures were analyzed with immunocytochemistry, bulk and single cell RNA-Seq. Quality control of the generated iNeurons is crucial, because case control comparisons in iPSC studies rely heavily on the repeatable generation of comparable neuronal cultures to assess small changes in open chromatin accessibility and gene expression.

### 4.2.1 Electrophysiological active burst firing excitatory neurons

One property of neurons is synaptic transmission between cells. In order to verify the differentiated iNeurons were capable of firing action potentials and receive glutamatergic input on excitatory synapses, electrophysiological recordings were performed. The properties of the signaling changes alongside the maturation of neurons. Electrical maturity of the generated cells was analyzed by field recording, whole cell current clamp, and the analysis of spontaneous evoked miniature excitatory postsynaptic currents (mEPSCs) (Figure 7). Schizophrenia patient and healthy control derived neurons were differentiated on multi electrode arrays (MEA) with 60 field electrodes (59 recording and one internal reference electrode). All electrodes were recorded simultaneously. iNeurons from healthy control and schizophrenia patient derived iPSCs displayed network activity and recurrent bursting behavior over a recording time of 3 min (Figure 7 A). Application of inhibitors such as NBQX and APV disturbed coordinated network burst activity by selectively inhibiting AMPA receptors and NMDA receptors, respectively. NBQX and APV disturbed glutamatergic synaptic transmission and the network desynchronized. Only a fraction of cells showed electric activity after NBQX and APV application. Application of tetrodotoxin

(TTX) which blocks voltage gated sodium channels necessary for action potential generation resulted in no measurable electrical changes at the field electrodes (Figure 7 B).



In order to confirm electrical activity of single cells and to assess the functionality and maturity of the generated neurons, whole cell current clamp was performed. The two cells from 2 donors, displayed a resting membrane potential lower than -70 mV. Positive depolarization of individual iNeurons from C3 or S3 by positive current injection resulted in a train of 17 or 15 action potentials, respectively. Negative current injection hyperpolarized the neurons. Membrane potential before the stimulation protocols slightly differed from the potentials after the stimulation (Figure 7 C).

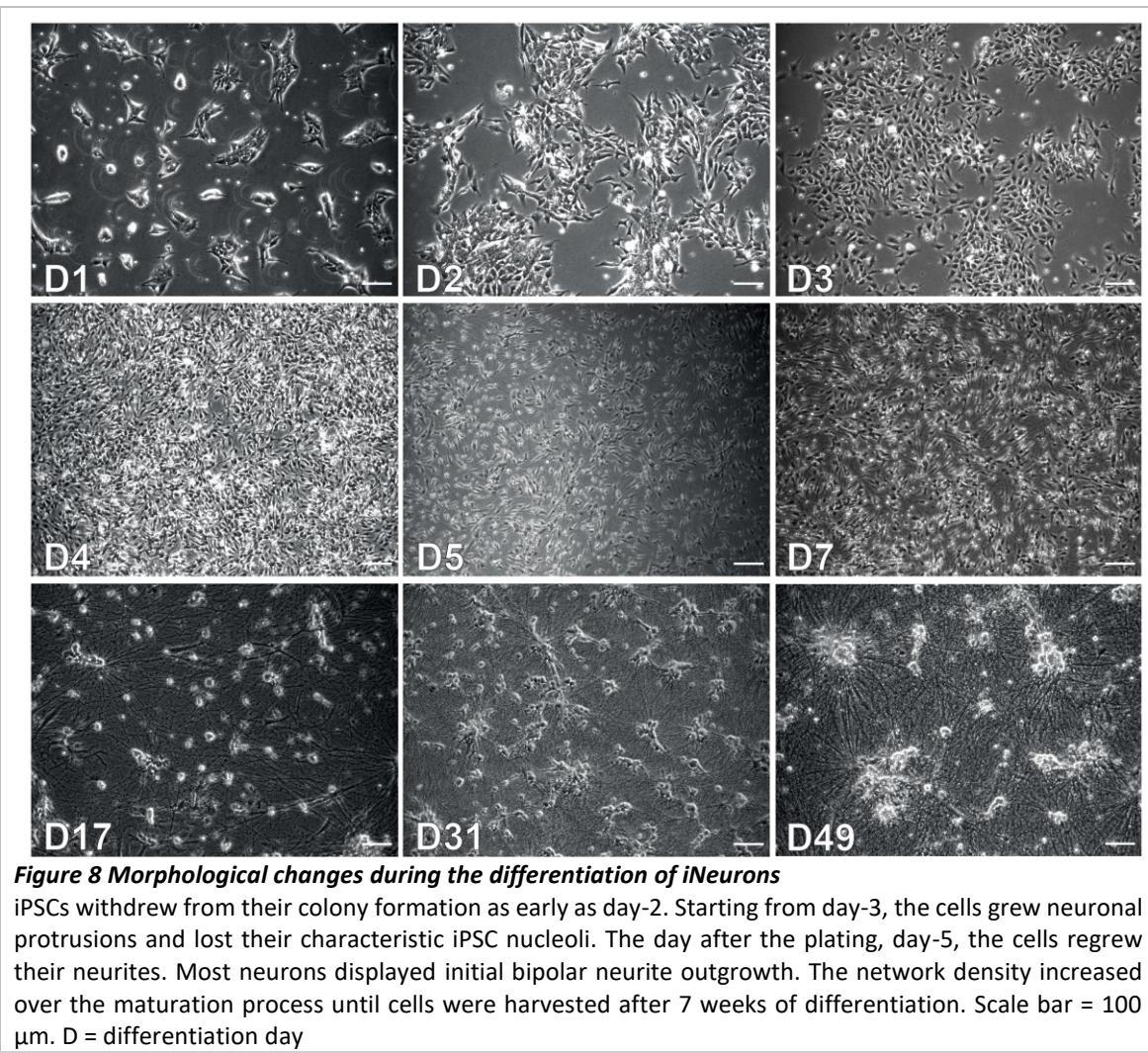
Intrinsic electrical properties of the post synapse were measured in whole cell voltage clamp mode. The described method records the effect on the post synapse from spontaneous vesicle release at the presynapse. According to definition, each measured event, named a miniature postsynaptic current (mEPSC), is due to the fusion from one unit of glutamatergic vesicle with the presynaptic membrane. Upon vesicle fusion, glutamate is released into the synaptic cleft. With the application of TTX to the external patching

solution, action potential generation was prevented to visualize small potential changes in the neuronal membrane. These small changes in membrane potential are triggered by glutamate that binds to ionotropic glutamate receptors e.g., kainate, AMPA, and NMDA receptors located at post synaptic densities. Binding of glutamate to its receptors allows positively charged ions to enter into the cytoplasm of the neuron. This influx of positive charges alters the ion concentration inside and outside of the cell which depolarizes the neuron. 4 healthy control iPSC derived iNeurons between day-42 and day-54 of differentiation from 3 batches were analyzed for their electrophysiological properties. The mean frequency of a spontaneous miniEPSC was 0.7343 Hz and ranged from 0.07 to 1.69 Hz. For several neurons the occurrence of a train of at least 3 miniEPSCs with an inter-mEPSC interval of less than 200 ms was recorded. This special firing pattern was defined as a burst. This burst phenomenon occurred on average 8.19 times per second, with 8 or 7 of 21 cells bursting  $\leq$  2 or more than 11 times per second, respectively. The average generated change in injected current in order to keep the clamped membrane potential of -80 mV was  $19.89 \pm 3.261$  pA (mean  $\pm$  SD). In whole cell voltage clamp it is possible to measure the input resistance ( $R_{in}$ ) of individual neurons.  $R_{in}$  of the iNeurons was on average at  $413.3 \pm 145.7$  M $\Omega$  (mean  $\pm$  SD). (Figure 7 D).

In summary, the changes in AP frequency upon addition of inhibitors for synaptic transmission verified neuronal glutamatergic activity. Additionally, synaptic glutamatergic vesicle induced postsynaptic depolarization was detected.

#### **4.2.2 iNeurons positive for neurite and cortical layer markers**

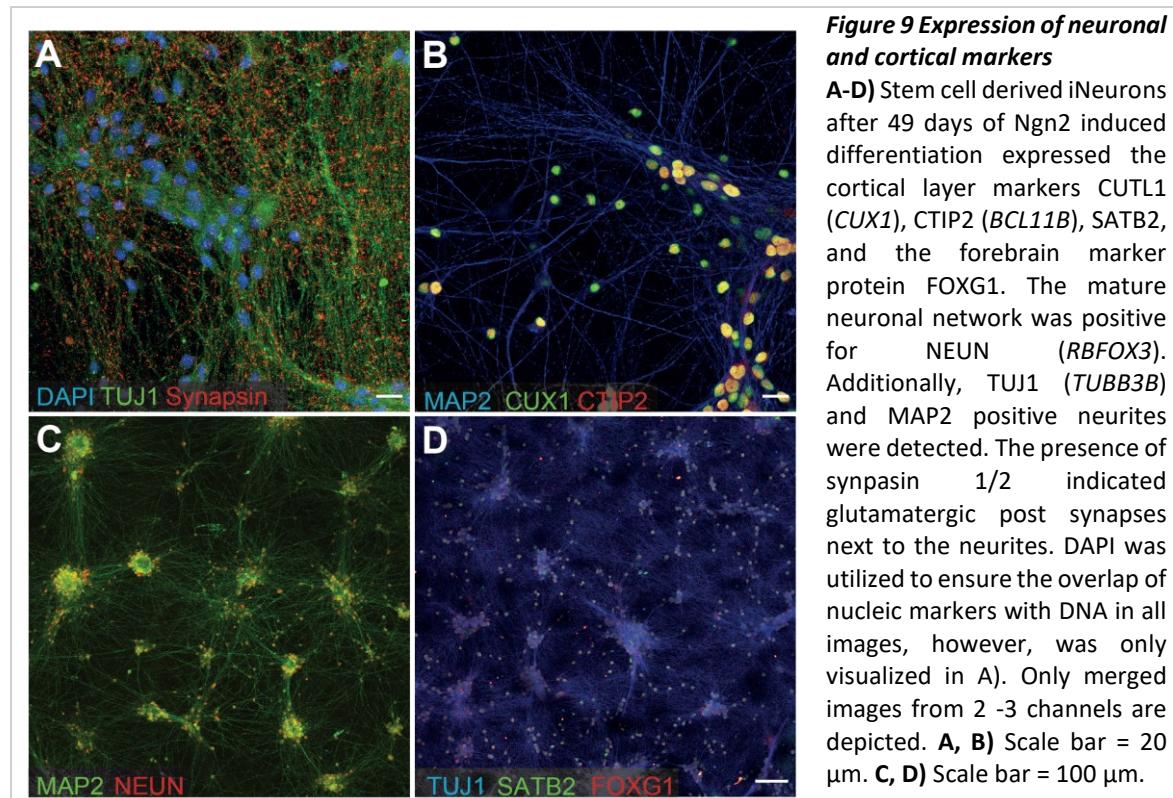
As indicated beforehand, a successful neuronal differentiation does not only depend on whether the neurons generate active neuronal networks but also whether the desired cell type is generated. In order to observe the differentiation process, neurons were closely monitored for changes in their morphology. After the harvest at day-49 of the differentiation, I checked for the expression of neuronal and region-specific cortical markers. As a next step, I quantified the expression of proteins used for cortical layer identification. This cortical layer identity of individual neurons for several iPSC lines and batches allows to quantify the homogeneity between individual differentiations. Only a homogeneous neuronal differentiation system enables the detection of subtle changes in gene expression or chromatin accessibility expected in schizophrenia.



During the first four days after the initial infection of the cells with the lentivirus coding for mouse Neurogenin-2 (Ngn2), the cells lose their close contacts characteristic for their stem cell state (Figure 8). At the day of the splitting procedure, day-4, cells have clearly elongated, and initial neurites are formed. This intermediate state results in a bipolar outgrowth of neurites and subsequently, neurite branching and network formation. The density of the network is heavily supported by the layer of mouse astrocytes, and it increased until cells were harvested or analyzed for their electrophysiological properties.

The dense neuronal network can be visualized by the general neurite marker TUJ1 ( $\beta$ 3-Tubulin, TUBB3) and the dendrite specific marker microtubule-associated protein 2 (MAP2). The successful generation of mature neurons was further verified by the presence of the pan-neuronal marker *RBFOX3* (NEUN) (Sarnat, 2013). Excitatory glutamatergic input of the generated neurons was confirmed with immunocytochemistry for the excitatory synaptic markers synapsin1/2. The generated neurons showed expression of special AT-rich sequence-binding protein 2 (SATB2), B-cell lymphoma/leukemia 11B (CTIP2, *BCL11B*),

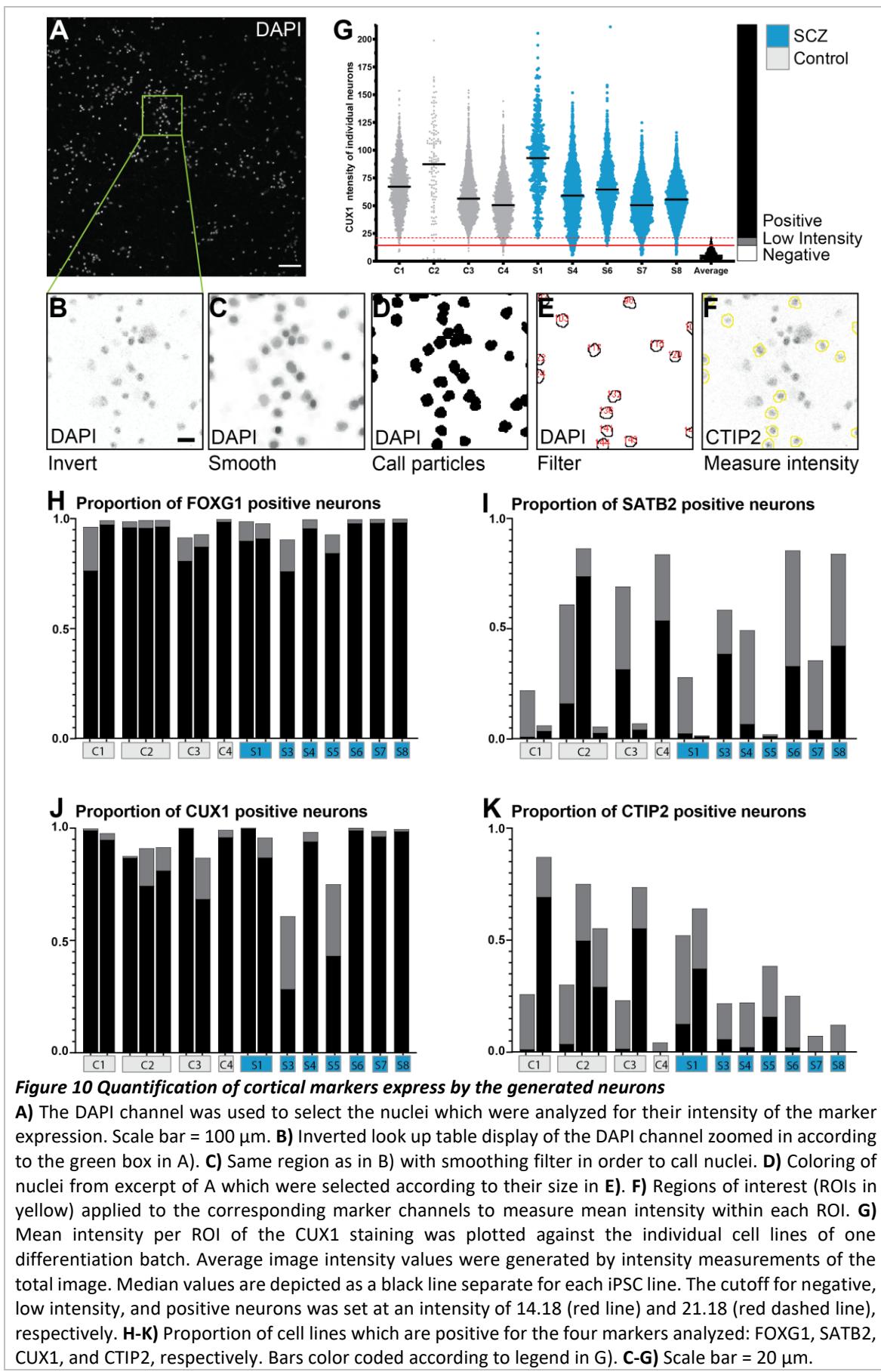
Cut Like Homeobox 1 (CUTL1, *CUX1*) and forkhead box G1 (FOXG1). These markers primarily are known for their specificity for cortical layer subtypes of excitatory neurons (Figure 9 A-D).



**Figure 9 Expression of neuronal and cortical markers**

**A-D)** Stem cell derived iNeurons after 49 days of Ngn2 induced differentiation expressed the cortical layer markers CUTL1 (*CUX1*), CTIP2 (*BCL11B*), SATB2, and the forebrain marker protein FOXG1. The mature neuronal network was positive for NEUN (*RBFOX3*). Additionally, TUJ1 (*TUBB3B*) and MAP2 positive neurites were detected. The presence of synapsin 1/2 indicated glutamatergic post synapses next to the neurites. DAPI was utilized to ensure the overlap of nucleic markers with DNA in all images, however, was only visualized in A). Only merged images from 2 -3 channels are depicted. **A, B)** Scale bar = 20  $\mu$ m. **C, D)** Scale bar = 100  $\mu$ m.

DAPI was used as a marker of the nucleus of the cells present in the field of view, including the mouse astrocytes (Figure 10 A). One z-layer focus per field of view was taken and analyzed in ImageJ. Images were automatically inverted, smoothed and the contrast was increased with the plugin MorphoLibJ (Figure 10 B-D). Particles above a pixel size of 45 and below 115 were selected with the region of interest (ROI) tool. The pixel range was set in order to avoid measuring partial cells or cell debris (below 45), astrocytes, and overlapping nuclei of two or more cells (above 115). ROIs were applied as a mask onto the images taken with the forebrain markers and the average intensity of the area of each ROI was measured separately (Figure 10 E, F). The intensity of all pictures ranged between a minimum of 0 and a maximum of 255 units corresponding to the resolution of an 8-bit image. The intensity of each ROI directly translated to the average marker intensity of one neuron. The intensity cutoff for a neuron to be considered negative for the markers with a high specificity was 14.18. This threshold equals the value at the 95 % percentile of the intensity of the whole image of all the markers used as a background intensity measure. The threshold for a positive image was set to 21.18 which corresponds to the maximum



intensity value measured at the average intensity values for all the images (Figure 10 G).

The range between positive and negative staining was also implemented to account for

cells which cannot be specifically annotated as positive or negative for the investigated marker. The selected cutoffs were supported by manual subjective counting. The annotation of a neuron positive for a marker was symbolized by a two-color scheme with black representing the cutoff as the maximum value of the averaged frames and grey indicating intensity values between the 95 % quantile of the average frame intensity and the maximum value.

In total, 36,707 or 39,422 cells were analyzed for their expression of CTIP2 and CUX1, or SATB2 and FOXG1, respectively (Figure 10 H-K). According to these cutoffs, the marker for forebrain cortical neurons, FoxG1 was evenly expressed in all differentiated iPSC lines. An average of 91.19 % of all analyzed neurons was positive for FOXG1. The lower and upper confidence limit of the median percentage of 95.69 % ranged between 84.29 % and 97.80 %. No expression of the forebrain marker was found in 0.11 to 9.57 % (minimum and maximum) of the neurons. On average  $84.08 \pm 21.29$  % (mean  $\pm$  standard deviation) of the neurons were positive for CUX1. CTIP2 and SATB2 levels varied highly between samples but did not associate with case-control status of the iPSC line. Furthermore, the cortical layer identity differed in proportion between the same cell line of replicate differentiations. 6 from 16 samples expressed mostly only one marker, while 8 or 6 samples from a total of 16 samples were mainly positive ( $> 60$  %) for SATB2 or CTIP2, respectively.

#### **4.2.3 RNA expression verified protein profiles and presence of peripheral neurons**

To further verify in depth the identity of the generated neurons, RNA-Seq was applied to two differentiation batches. In total, 6 healthy control derived and 5 schizophrenia patient derived iPSC lines were independently differentiated two times. iNeurons were treated on day-49 with hKCl stimulation or harvested with baseline neuronal activity (see also Figure 19 A). Bulk 3'RNA-Seq libraries were prepared after total RNA isolation. B12 and B13 were treated independently from differentiation, RNA isolation, NGS libraries preparation, to sequencing.

On average, 10,692,785 high quality reads were aligned to each library (Table 22). Number of reads did not significantly differ (Mann Whitney test,  $p = 0.2411$ ) between B12 (mean = 12,499,101.5) and B13 (mean = 9,316,543.2). Duplication rate differed (Mann Whitney test,

Donor	Stimulation	Batch	Diagnosis	Aligned reads
C1	Unstimulated	B12	Healthy control	8,072,286
C1	Unstimulated	B13	Healthy control	7,232,965
C1	hKCl	B12	Healthy control	11,720,174
C1	hKCl	B13	Healthy control	12,424,398
C2	Unstimulated	B12	Healthy control	9,230,946
C2	Unstimulated	B13	Healthy control	13,094,787
C2	hKCl	B12	Healthy control	9,502,376
C2	hKCl	B13	Healthy control	3,750,452
C3	Unstimulated	B12	Healthy control	11,577,012
C3	Unstimulated	B13	Healthy control	10,486,742
C3	hKCl	B12	Healthy control	13,063,156
C3	hKCl	B13	Healthy control	8,815,293
C4	Unstimulated	B12	Healthy control	10,813,220
C4	Unstimulated	B13	Healthy control	10,153,228
C4	hKCl	B13	Healthy control	7,410,359
C5	Unstimulated	B13	Healthy control	11,587,855
C5	hKCl	B13	Healthy control	6,452,523
C6	Unstimulated	B12	Healthy control	11,253,106
C6	Unstimulated	B13	Healthy control	13,817,655
C6	hKCl	B12	Healthy control	13,059,872
C6	hKCl	B13	Healthy control	7,739,987
S1	Unstimulated	B12	Schizophrenia	23,040,306
S1	Unstimulated	B13	Schizophrenia	8,291,096
S1	hKCl	B13	Schizophrenia	6,341,459
S4	Unstimulated	B13	Schizophrenia	11,103,734
S4	hKCl	B12	Schizophrenia	14,822,114
S6	Unstimulated	B12	Schizophrenia	21,604,376
S6	Unstimulated	B13	Schizophrenia	10,738,887
S6	hKCl	B12	Schizophrenia	11,519,156
S6	hKCl	B13	Schizophrenia	10,312,263
S7	Unstimulated	B12	Schizophrenia	11,442,658
S7	Unstimulated	B13	Schizophrenia	7,761,148
S7	hKCl	B12	Schizophrenia	10,624,660
S7	hKCl	B13	Schizophrenia	7,802,089
S8	Unstimulated	B12	Schizophrenia	8,640,206
S8	Unstimulated	B13	Schizophrenia	12,124,050
S8	hKCl	B13	Schizophrenia	8,206,437

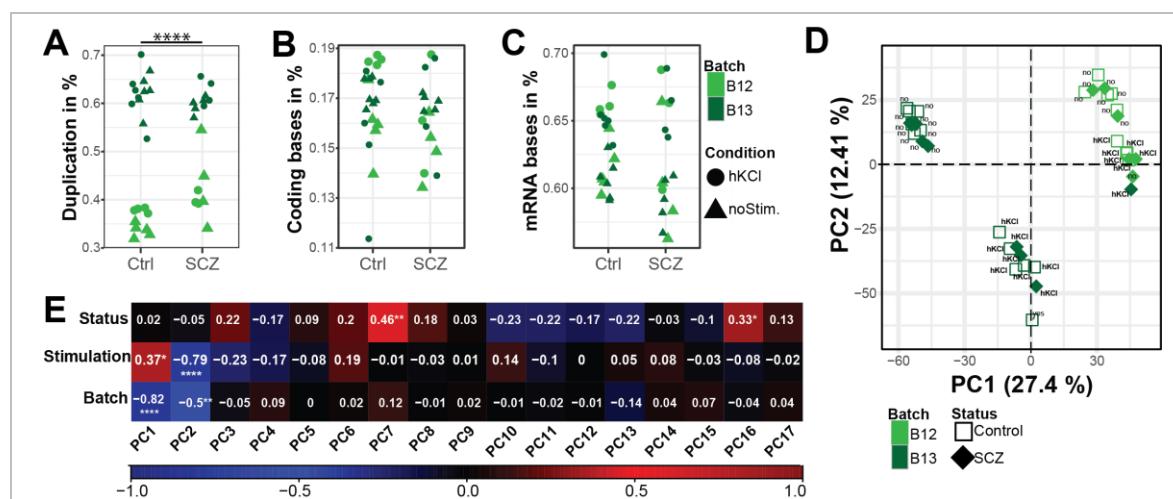
**Table 22 Overview of the bulk RNA Seq samples from day-49 iNeurons**

Overview of the here described bulk RNA sequencing samples from day-49 iNeurons from 6 healthy control and 5 patients diagnosed with schizophrenia. Samples were differentiated and sequenced in two batches. Aligned reads per sample are annotated for each library separately.

$p < 0.0001$ ) significantly (Figure 11 A) between B12 and B13. Neither batches nor disease status influenced the percentage of coding bases or mRNA bases (Figure 11 B, C). Principal component 1 (PC1) explained over 25 % of the variation in the data set (Figure 11 D) and was driven by the expression of the upper 1 % most variable genes: *RPS10*, *UBB*, *HLA.E*, *MTRNR2L8*, *RPS4Y1*, *SBF1*, *DCHS1*, *DRGX*, *CHCHD2*, *PIEZ02*, and *ARHGAP36*. Remarkably, this included the gene *HLA.E* which is highly associated with schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2014). In line with the

knowledge that the applied hKCl stimulation is a driver of huge transcriptional impact it was significantly correlated with the first two PCs (Figure 11 E). PC1 and PC2 were significantly affected by the differentiation batch.

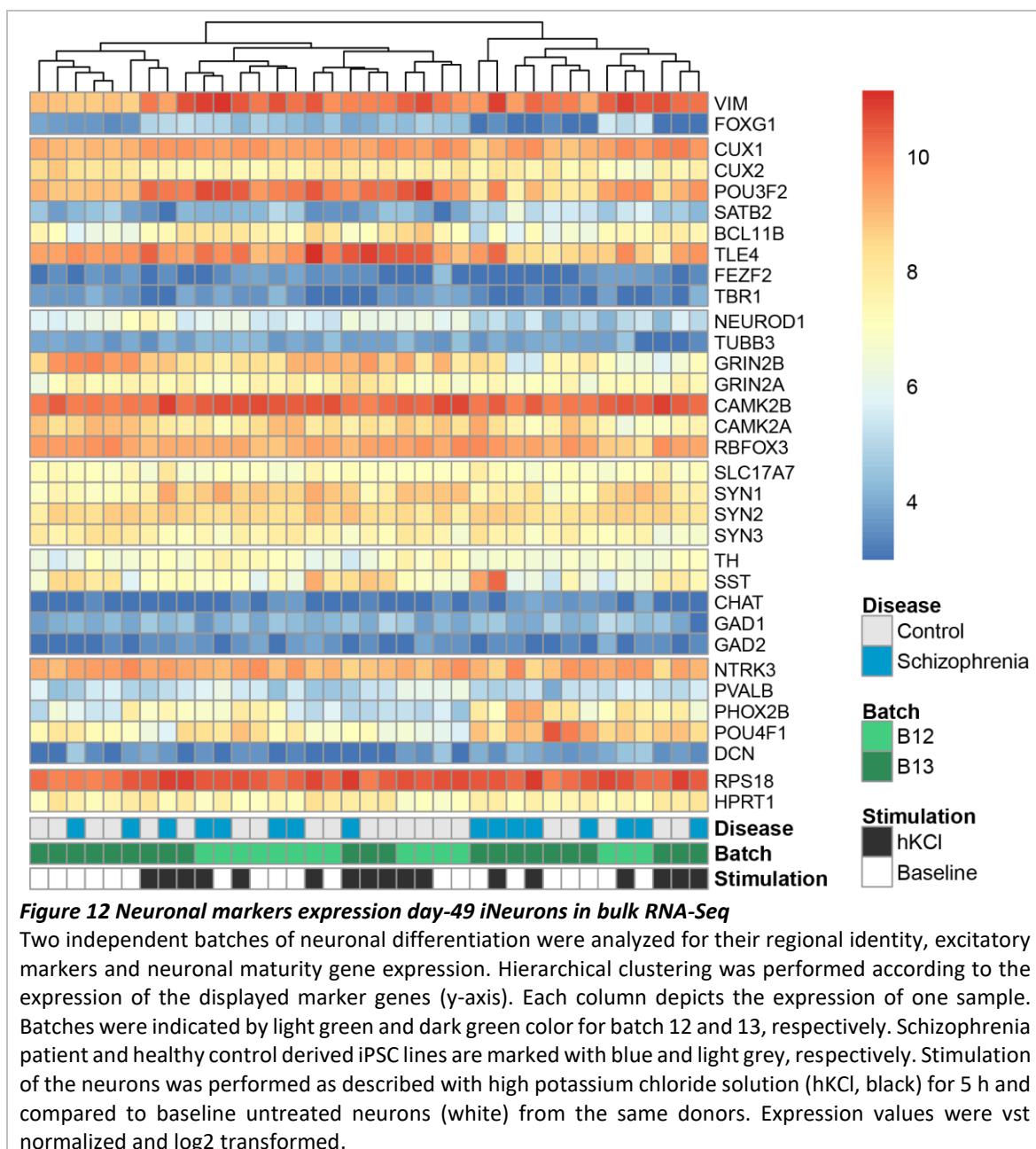
Next, hierarchical clustering of iNeurons analyzed by bulk RNA sequencing was performed on expression of several neuronal marker genes. Marker genes were selected from previously published Ngn2 overexpression studies and mousebrain.org for CNS and PNS neurons (Figure 12). In the bulk RNA-Seq approach, iNeurons are positive for forebrain cortical marker (*FOXG1*) which is mainly expressed in the telencephalon. iNeurons expressed at high level markers for cortical layers II-IV (*BRN2*, *CUX1*, *CUX2*), for mid-lower layers (*TLE4*) but also for the lower layer V (*CTIP2* / *BCL11B*). The expression of *SATB2* and *CTIP2*, however, is less homogeneous between differentiations with an opposite directionality between the two genes. The excitatory synaptic markers (vGLUT1 or *SLC17A7* and *SYN1-3*) are widely expressed in the neurons. vGLUT1 was a highly homogeneously expressed gene of the depicted synaptic markers.



**Figure 11 Quality control metrics of bulk RNA-Seq data**

**A-C)** Batch is encoded by color (B12 = light green, B13 = dark green), condition depicted by shape (hKCl=triangle, noStim. = unstimulated baseline = circle). X-axis denotes sample donor as healthy control ("Control") or patient suffering from schizophrenia ("SCZ"). Each dot represents one sample. **A)** Y-axis displays duplication rate in % of sequenced fragments. Significant difference between B12 and B13 duplication rate ( $p < 0.0001$  Mann Whitney test, non-normal distributed data according to Shapiro-Wilk test). No significant difference between duplication rate of control vs schizophrenia samples (Mann Whitney test  $p = 0.7736$ ). **B)** No significant difference between coding bases (in percentage, y-axis) of control vs schizophrenia samples (Mann Whitney test  $p = 0.2411$ ). **C)** No significant difference between mRNA bases (in percentage, y-axis) of control vs schizophrenia samples (Mann Whitney test  $p=0.3868$ ). **D)** Bi-plot of the individual differentiation samples on PC1 (x-axis) and PC2 (y-axis). Variance explained depicted as % for PC1 and PC2. Samples were color coded as in A-C) and shape encoded disease status from the cell lines: control = healthy control, SCZ = schizophrenia derived. hKCl (depolarization stimulus) and no (baseline) are indicated next to the locations of each sample. **E)** Pearson correlation coefficient (color coded, negative in blue and positive correlation in red) of disease status (schizophrenia or healthy control), stimulation (baseline or hKCl), and batch (B12 or B13) with principal components (PC1-17). **A,E)** P-value  $< 0.05 = ^*$ ,  $< 0.01 = ^{**}$ ,  $< 0.001 = ^{***}$ ,  $< 0.0001 = ^{****}$

The pan-neuronal markers *NEUROD1* and *TUBB3* (TUJ1) are expressed in all samples. *DCX*, *MAP2*, and *MAPT*, as additional pan-neuronal maturity markers could not be plotted due to the fact that they were not in the generated vst normalized output file from the DESeq2 tool. Similarly, subgroup classifiers from the single cell RNA-Seq experiment such as *PRPH*, *ISL1*, *SLC17A6*, *NTRK2*, and *GPM6A* were present in raw expression counts from bulk RNA-Seq experiment, however, were excluded from the vst normalization due to extremely high expression counts (> 99<sup>th</sup> percentile). The two plotted housekeeping genes *HPRT1* and *RPS18* (a ribosomal subunit) are expressed at similar levels between the cell lines and the batches.



The neurons were positive for the mature postmitotic marker *RBFOX3* (NEUN). Furthermore, the maturity markers *CAMK2A*, *GRIN2A*, and *GRIN2B* were expressed at similar levels, however, *GRIN2B* showed more heterogeneity in expression levels than *GRIN2A*. The expression of the *glutamate ionotropic receptor NMDA type subunit 2B* (*GRIN2B*) was high in samples with low *PHOX2B* and *POU4F1* expression. In general, *PHOX2B*, similar to *POU4F1*, expression was very heterogeneous between the samples. *GRIN2B* expression was high in samples with lower expression of the radial glia or NPC marker *vimentin* (*VIM*). Using these selected marker genes, the clustering of the cell lines was not primarily driven by differentiation batch, the disease status or the stimulation paradigm but by expression of the marker genes.

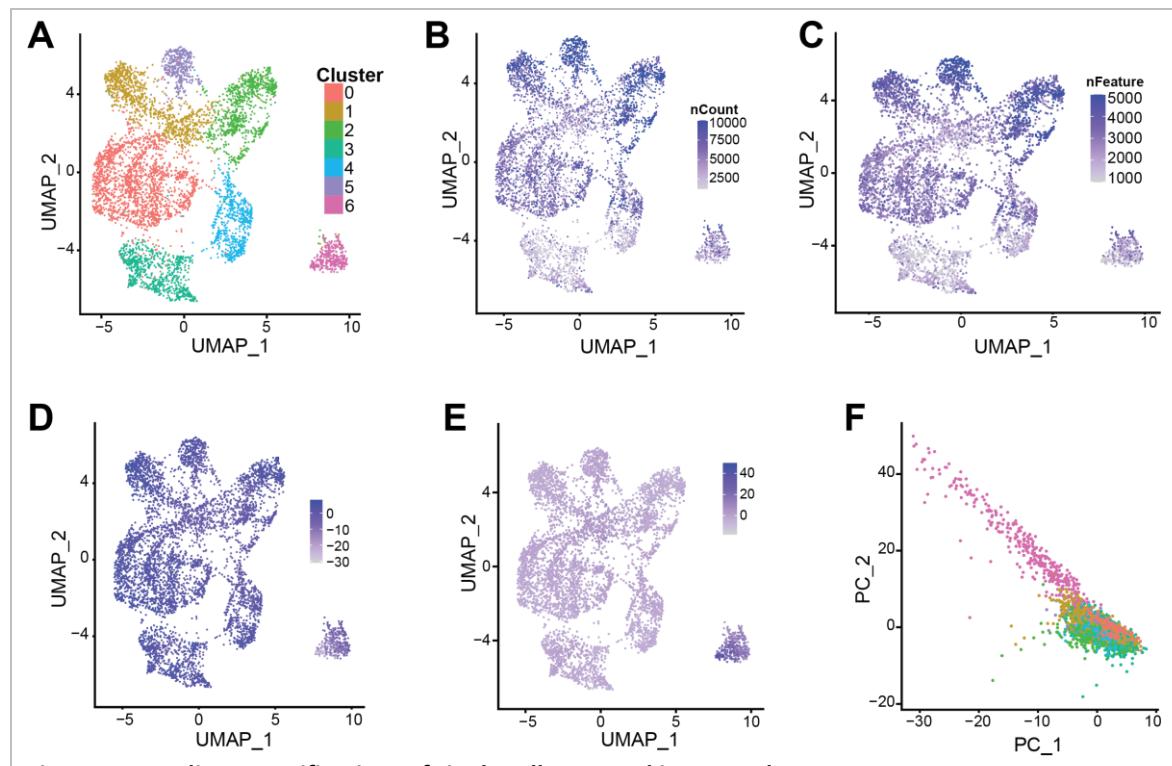
#### **4.2.4 Single cell RNA-Seq confirmed central and peripheral nervous system identity**

In a follow-up experiment to further investigate the heterogeneity of the generated neurons on a single cell level, 3' single cell RNA-Seq (scRNA-Seq) was utilized. The power of the single cell experiment in elucidating the different cell types of the Ngn2 overexpression protocol with patterning factors was corroborated, though, by the change in media composition after day-24 of this differentiation batch. Unlike all other experiments presented in this thesis, the BrainPhys media was not supplemented with FCS for the scRNA-Seq experiment. Analysis and correlation of the immunocytochemistry, bulk RNA-Seq and scRNA-Seq was still carried out since the majority of the markers were present for both differentiation paradigms. Furthermore, the cell type identity is determined for the iNeuron protocol after the first days of differentiation and, hence, should have not been affected by the late stage alteration in the differentiation protocol (Lin et al., 2020).

One single-cell RNA library of 4 cell lines was generated to describe in more detail individual cell populations which could not be resolved in the bulk RNA-Seq experiments. C1 and C6, and S1 and S2 were differentiated as healthy controls and as patient derived iPSC lines, respectively. Subgroups of neurons were clustered together and common markers for regional identity, maturation state, and routes of signaling were used to annotate the identity of the iNeurons. For this analysis, Vanessa Murek used gene expression of single iNeurons to transfer cell types from the mouse nervous system (Zeisel et al., 2018). It is important to note, however, that the used cell identities are from mouse *in vivo* tissue and

transferred on human *in vitro* differentiated neurons. Additionally, the maturity of the iNeurons is not comparable to adult (mouse) tissue (Lin et al., 2020; Nehme et al., 2018). Therefore, transferred labels were inspected manually for their transcriptional landscape. Nevertheless, this data set was used, because the labelled cell types contained neurons and supporting glia from the central and the peripheral nervous system.

#### 4.2.4.1 scRNA-Seq revealed presence of 7 distinct clusters



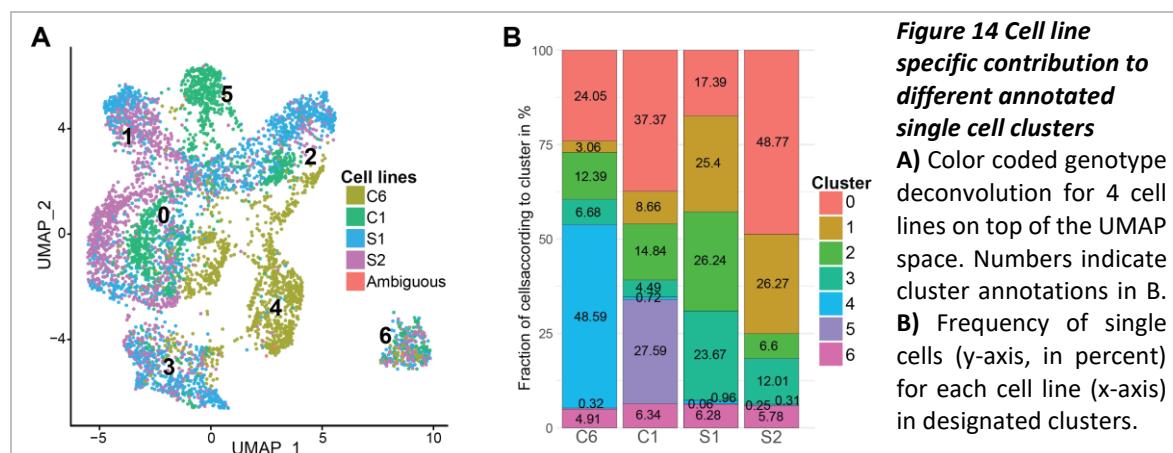
**Figure 13 Baseline quantifications of single cells grouped in seven clusters**

**A)** Transcriptomes were clustered and visualized using UMAPs. Depicted is one single cell experiment containing the differentiation of 4 independently differentiated iPSC lines. Single cells are represented with a dot and colored according to their cluster affiliation (cluster 0-6). **B-C)** UMAP display of nCount or nFeature after the cell filtration step, respectively. **D-E)** UMAP display colored by association with PC\_1 and PC\_2, respectively. **F)** Principal component score plots displaying PC\_1 (x-axis) against PC\_2 (y-axis) individually for each single cell (dots). Colors are according to cluster affiliation from **A**).

The scRNA-Seq library was sequenced with a total of 179,603,661 reads. A mean read depth of 19,841 per cell was acquired. A median of 1,992 genes were detected per cell. In total, 9,052 cells were estimated to be present by the Cell Ranger tool (10X Genomics). A total of 25,286 genes were detected when aligned against hg38 (v3.1.0), with 88.3 % of the reads mapped to the human genome. 53.7 %, 25.3 %, and 6.6 % were confidently mapped to exonic, intronic, and intergenic regions, respectively. A median of 3,622 UMI counts were detected per recovered cell. Stringent filters (see 3.5.4) were applied to the Cell Ranger

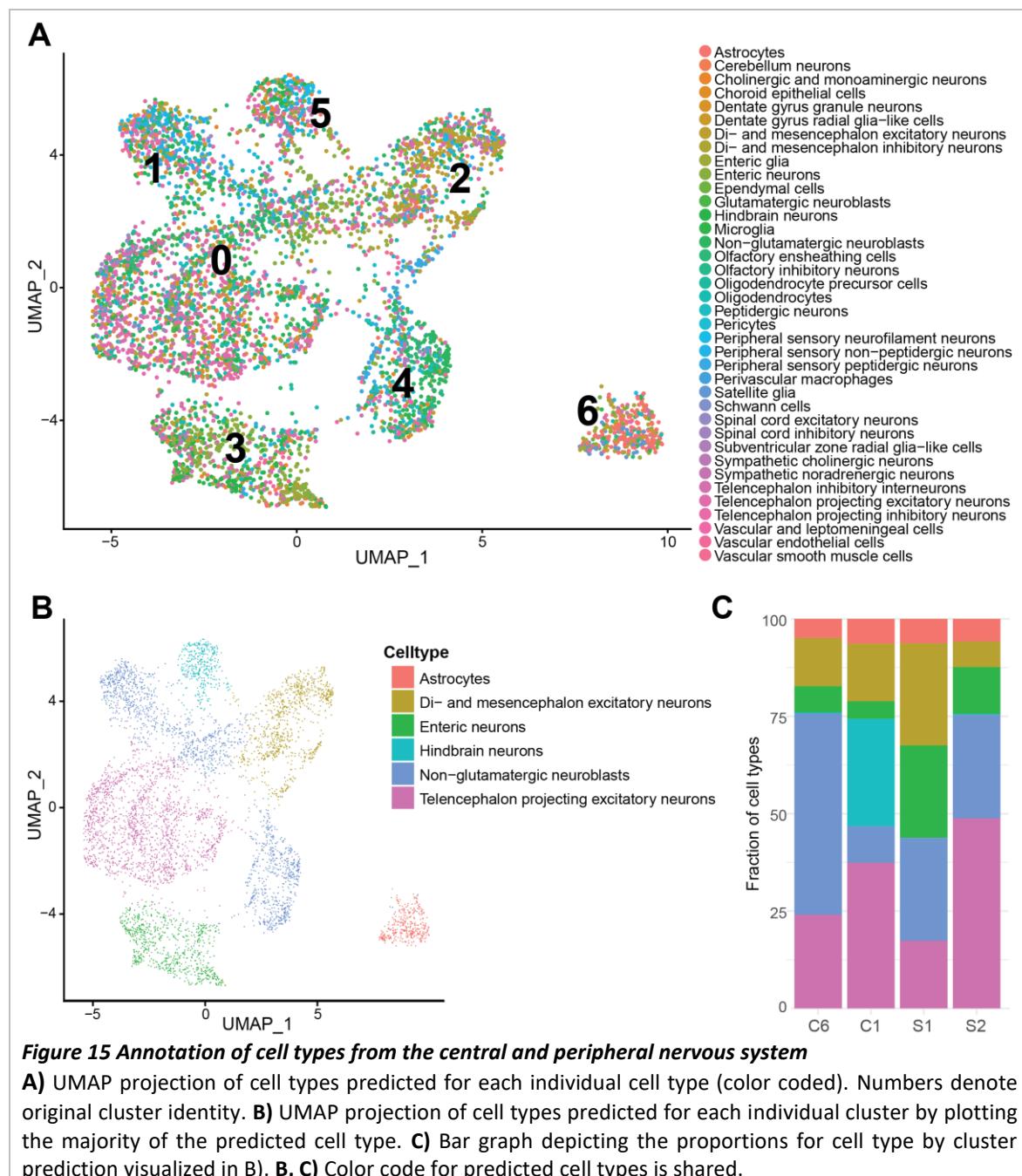
results which recovered 5,777 high quality cells. Gene expression counts were normalized by sctransform-based transformation (Hafemeister & Satija, 2019).

The remaining 5,777 cells were clustered according to their gene expression independent of the donors into 7 clusters (Figure 13 A). The uniform manifold approximation and projection (UMAP) depicts the color-coded distribution of cells in the 7 clusters: cluster 0-6. For each cell individually, the overall number of transcripts (nCount) and detected genes (nFeature) was plotted onto the UMAP distribution (Figure 13 B, C). Cluster 3 had the lowest number of feature or transcripts. Principal component analysis (PCA) indicated that the first 2 PCs explained the difference between cluster 6 and clusters 0-5. The first two PCs explained 3.98 % (2.49 % for PC\_1 and 1.49 % for PC\_2) of the total variance of the data set. The correlation between PC\_1 and PC\_2 was -0.88. (Figure 13 D-F). PC\_3, PC\_4, and PC\_5 (not depicted) explained 1.17, 1.05, and 0.77 % of the variance of the data set, respectively.



As a next step, the distribution of the four cell lines in each of the clusters was investigated. Only 0.4 % of the cells were not attributed to originate from one of the four sequenced cell lines and labelled as “ambiguous”. Clusters 4 or 5 were driven mostly by one cell line: C6 or C1, respectively (Figure 14 A). In contrast, the fraction of cells in cluster 6 was very similar between cell lines ranging from minimum 4.91 % (C6) to maximum 6.34 % (C1). For clusters 0 - 2, a separation of cell lines was more apparent. Cluster 1 mainly consisted of cells derived from S1 and S2 (25.4 and 26.27 %) compared to a total of 11.72 % originating of cells derived from C1 and C6. The majority of cells from cluster 3 were generated by S1 (Figure 14 B).

In order to annotate not only the cell line but also the cell types present in the generated *Ngn2* induced neuronal cultures, a publicly available data set of neuronal and non-neuronal cell types of the peripheral and the central nervous system was utilized (Zeisel et al., 2018). A label transfer approach predicted the cell type of the iNeurons based on the similarity of



**Figure 15 Annotation of cell types from the central and peripheral nervous system**

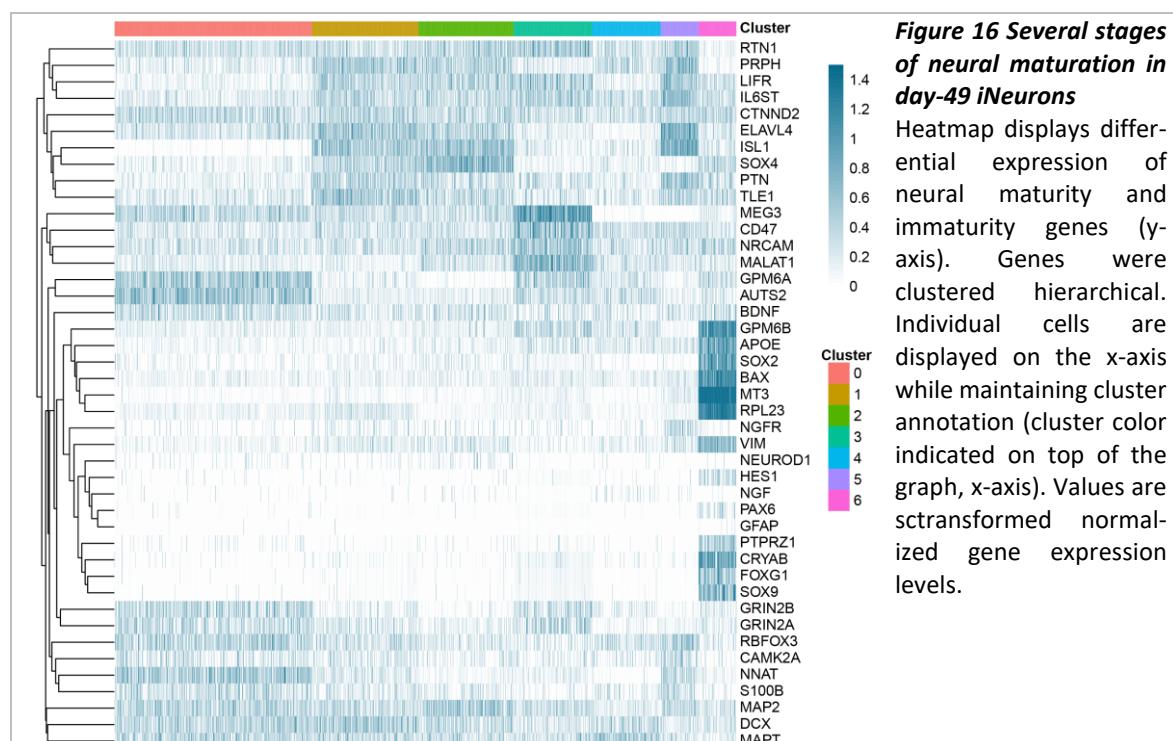
**A)** UMAP projection of cell types predicted for each individual cell type (color coded). Numbers denote original cluster identity. **B)** UMAP projection of cell types predicted for each individual cluster by plotting the majority of the predicted cell type. **C)** Bar graph depicting the proportions for cell type by cluster prediction visualized in B). **B, C)** Color code for predicted cell types is shared.

the transcriptome between the mouse database and *in vivo* differentiated iNeurons. Two approaches were followed: 1) the prediction for each cell individually (Figure 15 A) and 2) the prediction for all cells grouped in one cluster (Figure 15 B). For the first approach 38 different cell types were detected with some of them indicating regionally close related cell subtypes e.g., neurons from the spinal cord or the telencephalon. For the second approach,

label transfer-based annotation from primary adult mouse brain suggested 6 distinct cellular identities for 7 clusters (Figure 15 C). In general, though, iNeurons are comparable to fetal human neurons which is not taken into account in this approach of identity annotation (Nehme et al., 2018). This approach used a public database which consists of adult (postnatal age P12 - 30 or 6 and 8 weeks after birth) mouse single cell 3' RNA-Seq data (Zeisel et al., 2018). Therefore, the identity label of the cell clusters was investigated with a manual approach.

#### 4.2.4.2 Presence of immature and mature neuronal marker genes in day-49 iNeuron

Dispersed throughout the clusters, several cells were labelled as astrocytes of different regional identity and cluster 6 was labelled as an astrocyte cluster by the label transfer approach in Figure 15 B. The astrocytic marker gene *S100B* was detected in a high proportion of cells. However, *S100B* is not exclusively expressed in astrocytes and can also be expressed by peripheral sensory neurons. Additionally, none of the clusters expressed the astrocytic lineage genes *GFAP* and *LDL*. Cluster 6 was, though, enriched for the proapoptotic marker *BAX* which is important for astrocyte lineage induction from radial glia cells (Chang et al., 2007; Yuzwa et al., 2017).



In line with the lack of classic astrocytic marker genes in cluster 6 and the annotation of neuroblast cells for clusters 1 and 4, neural maturation genes were further investigated for

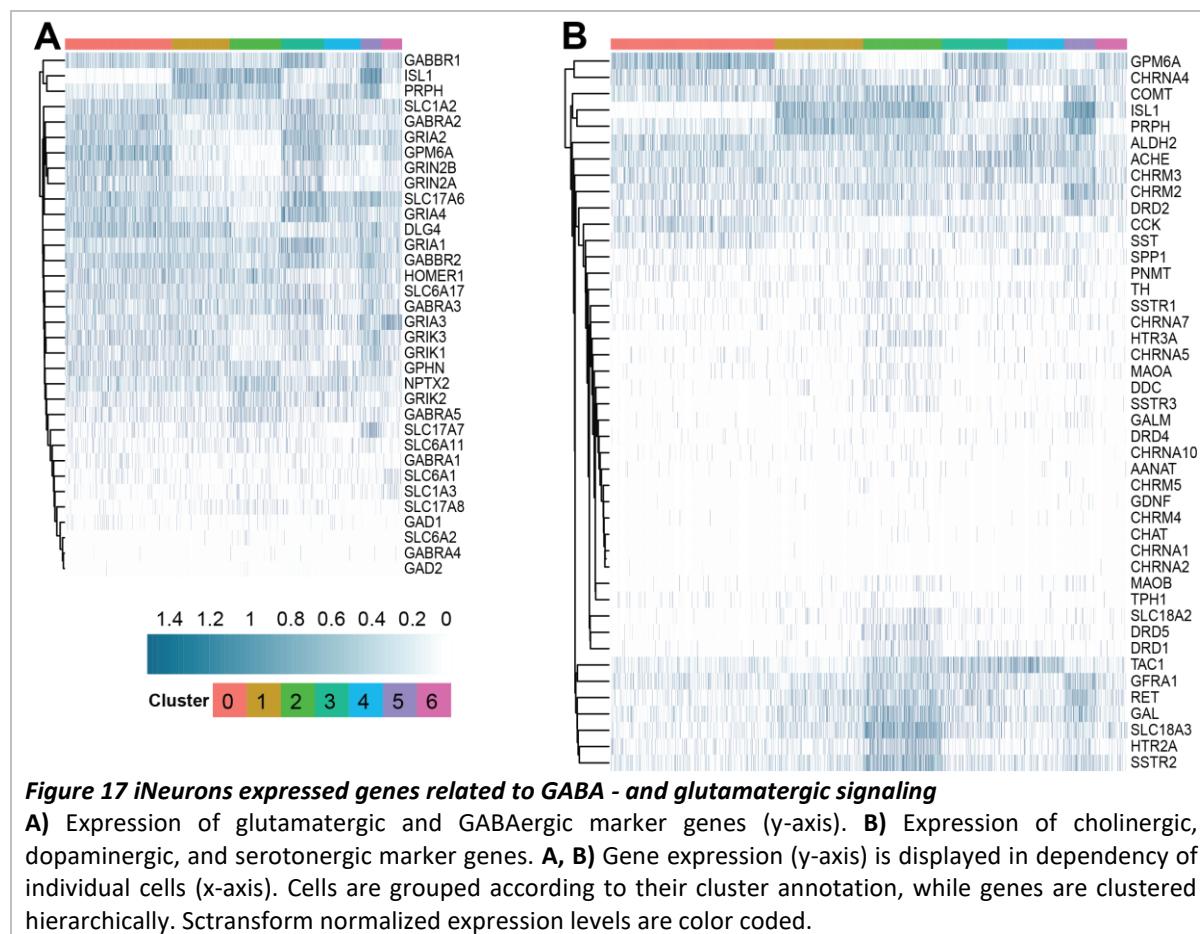
their expression in the different cluster. In general, the neurite markers *MAP2* and *DCX* were expressed in all clusters at varying intensity (Figure 16). In agreement with this, expression of the pan-neuronal markers *RBFOX1*, *MALAT1*, *CD47*, and *NRCAM* was detected. *CD47* and *NRCAM* promote the development and growth of neurites (Murata et al., 2006). Furthermore, *NRCAM* and the long non coding RNAs (lncRNA) *MALAT1* and *MEG3* (specifically high in cluster 3) are important for synapse formation during neuronal development (Xuejing Zhang et al., 2017; Zhao et al., 2020). *NNAT* is important for early neuronal development. In more mature states, *NNAT* expression overlaps with *SOX11* (Burke et al., 2020). *SOX11* is present in early-born postmitotic neurons (C. Chen et al., 2015) and can be expressed in *Ngn2* positive neurons (immature) and mature neurons in parallel with *NEUROD4* (and *MATH3*, not expressed) and *SOX4* (Bergsland et al., 2006). *SOX4* is enriched for cerebral cortex neurons and is expressed in postmitotic intermediate progenitors and differentiating neurons.

In intermediate progenitors, *Sox4* partners with the proneuronal gene *Ngn2* to activate *Eomes* (*EOMES* was not expressed) (C. Chen et al., 2015). In line with this, the maturity markers *CAMK2A*, *GRIN2A*, *GRIN2B*, and *MAPT* were expressed at varying degrees in all of the clusters, apart from cluster 6. Similarly, the mature neuronal markers *RBFOX3*, *CTNND2*, *NRCAM*, and *RTN1* were depleted specifically in cluster 6 (Pollen et al., 2015). In line with this, cluster 6 specifically expressed the neuronal progenitor markers *VIM*, *SOX2*, *SOX9*, *APOE*, *PTPRZ1*, *CRYAB*, *FOGX1*, *HES1*, and *RPL23* (Figure 16). Expression of *APOE* and *SOX9* are known neural stem cell (NSCs) marker genes and *SOX9* in combination with *SOX2* regulate adult NSCs maintenance and function (Shin et al., 2015). Additionally, the radial glia (RG) marker *CRYAB* was enriched in cluster 6. High levels of *MT3*, a neuronal growth inhibitor, is associated with neuronal progenitors or RG cells of the developing mouse cortex (J. Wang et al., 2011; Yuzwa et al., 2017). RG markers *PTPRZ1* and *APOE2* were enriched in cluster 6 while other RG markers such as *IL6ST*, *ISL1*, *LIFR*, and *PTN* were also present in clusters 0-5. *LIFR*, *IL6ST*, and *TLE1*, though, are also indicative of mature postmitotic neurons and mediate neuronal survival (clusters 1 - 3,5). Forebrain specificity of the immature neurons in cluster 6 was supported by the expression of *FOGX1*. *SLC1A3* (EAAT1) expression supported the annotation of cluster 6 as glutamatergic excitatory RG. In conclusion, the above-mentioned expressed markers supported the label “neural

progenitor cells" instead of astrocytes for cluster 6 in line with the specificity of Ngn2 overexpression for neuronal differentiation.

#### 4.2.4.3 Neurotransmitter supported subdivision of day-49 iNeurons

Apart from their developmental maturity, the generated neurons were also classified according to their released neurotransmitter (NT). The markers mainly considered for NT identity were genes coding for 1) enzymes that are necessary in the presynapse to synthesize the NT and 2) specialized transporters to load the NT into the vesicles located in the presynapse. NT specific receptors, reuptake transporters, and degrading enzymes, though, are an indirect measurement of the presence of certain NT specific synapses and do not indicate the identity of the cell itself (Figure 17 A, B).



Several marker genes implying excitatory signaling were expressed in the iNeurons indicating different maturation stages of glutamatergic neurons: *ELAVL4* (commitment and sustaining), *NPTX2*, *PLD3* and *AUTS2* (maturation) (Pollen et al., 2015). The vesicular transporters *SLC17A6* (vGLUT2) and *SLC17A7* (vGLUT1) levels were high in *GPM6A* expressing cells and cluster 5, respectively. *SLC17A8* (vGLUT3) was present only in a

minority of the cells. The expression of the excitatory glutamatergic amino acid transporter 2 (*SLC1A2*, EAAT2) is present in human CNS neurons during development and lasts until 2 months after birth. Later on, *SLC1A2* expression is restricted to astrocytes. Excitatory post synapses (*HOMER1*) contain NMDA receptors (*GRIN2A*, *GRIN2B*), kainate responsive glutamatergic receptors (*GRIK1-3*), and AMPA responsive glutamate receptors (GRIA gene family).

Cluster 2 (lowest expression of glutamatergic vesicular transporters) expressed dopamine and norepinephrine related enzymes (*TH*, *DDC*, *PNMT*), vesicular transporter *SLC18A2* (VMNT2), receptors (*DRD1-5*), and degrading enzymes (*MAOA*, *MAOB*, *COMT*). Moreover, cluster 2 expressed serotonin marker genes (*DLK1*, *VGF*, *ATP6VOA1*, *SLC3A2*), enzymes necessary for the serotonin synthesis (*TPH1* or *TPH2*, *DDH*), vesicular transport *SLC18A2* (VMNT2), degrading enzymes (*AANAT* and *ASMT*, or *ALDH2* and *MAOA*), and receptors (G-protein coupled *HTR2A*, ionotropic *HTR3A*).

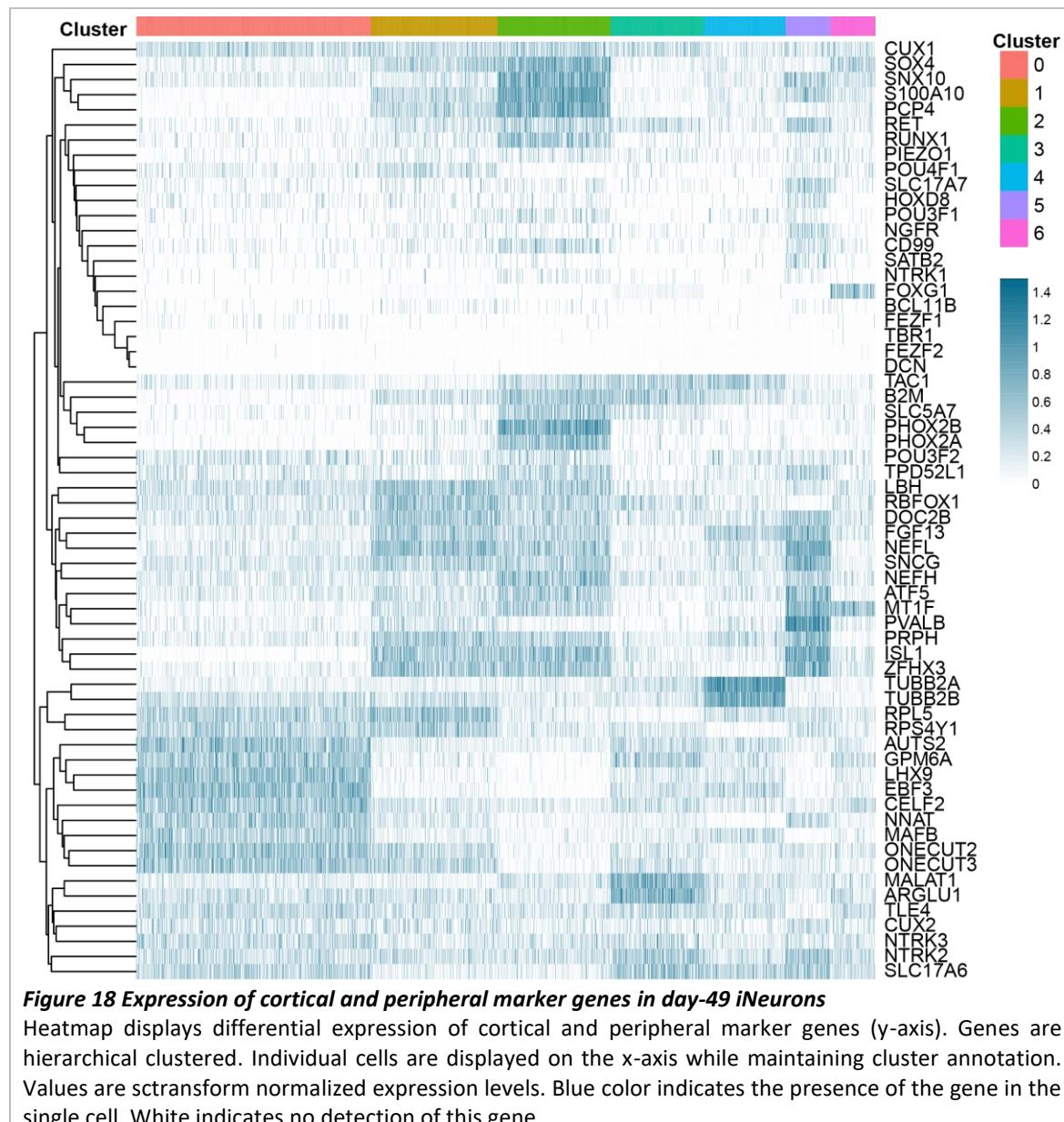
GABAergic signaling was detected in only a few cells according to marker genes (*MAFB*, *SST*, *PVALB*), the enzymes which convert glutamate to GABA (*GAD1*, *GAD2*), and the vesicular transporters (*SLC6A1*, vGAT1 and *SLC6A2*, vGAT2). Nevertheless, scRNASeq data suggested presence of gephyrin (*GPHN*) in all cluster, a marker gene for inhibitory post synapses. Additionally, GABAergic receptors (ionotropic GABA<sub>A</sub>-receptors *GABRA1-5* and G-protein coupled GABA<sub>B</sub>-receptors *GABBR1-2*) were detected in all clusters.

The NT acetylcholine (ACh) is produced by the enzyme *CHAT* (no detection), transported into the vesicle by *SLC18A3* (enriched in cluster 2), and binds to two types of ACh receptors: muscarinic (CHRM gene family) and nicotinic (CHRN gene family). *ACHE* is necessary to degrade ACh and was not detected in the cells. Peptidergic signaling with *TAC1*, *SST*, *CCK*, and *SPP1* was present in several cells.

#### 4.2.4.4 Expression of central and peripheral nervous system marker

As a next step and a follow-up from the immunocytochemistry and the bulk RNA-Seq data, the regional identity of the iNeurons was further investigated (Figure 18). Cortical layer marker genes were co-expressed in cells expressing *PRPH*, a marker nearly exclusive for peripheral neurons (Zeisel et al., 2018). Nevertheless, both major marker genes (*GPM6A* for CNS and *PRPH* for PNS) overlapped in several clusters. The presence of PNS and CNS

neurons with similar co-expressed genes in the same cluster further aggravated the annotation of cell types. Besides that, the marker genes of the previously annotated mouse neuronal identities did not indicate a conclusive cell type for each cluster.



Telencephalic glutamatergic neurons, the targeted cell type was supported by a multitude of expressed genes (e.g. *SOX4*, *GPM6A*, *AUTS2*, *CELF2*, *SNX10*, etc.) but also contradicted by expression of e.g., *POU4F1*, *GSX2*, *ISL1*, *TAC1*, and *PRPH* (Pollen et al., 2015; Y. Qi et al., 2017). Glutamatergic neurons of the hindbrain were implied by expression of *LHX9*, *MAFB*, *AUTS2*, *NNAT*, *CELF2*, *POUF41*, *FGF13*, *NEFL*, *RPL5*, and *GPM6A*. Nevertheless, some neurons were also positive for *ONECUT2/3*, *EBF3*, *NNAT*, *CELF2*, *ZFHX3*, *RBFOX1*, *PRPH*, *NEFL*, *LBH*, *DOC2B*, *RPS4Y1*, and *RPL5* indicating cholinergic hindbrain neurons. Similarly,

the dopaminergic cell type expressed *TH* but lacked the expression of *CHMR4A*, *NR4A2*, *LMX1B*, *PPP1R1B*, *GPR88*, and *RXRG*.

Nevertheless, due to the presence of *PRPH*, *PHOX2B*, *POU4F1*, and *ISL1* an additional peripheral identity of the neurons was indicated even though *DCN*, a marker for neural crest cells, was not detected in the here presented data set (Chambers et al., 2009; Lin et al., 2020). However, *COL5A1*, a marker for mesenchymal cells (annotated as such in Lin et al.), was present but could also indicate hindbrain or enteric fate. Enteric fate, also detected in the automated approach, was further supported by expression of, for instance, *DOC2B*, *TAC1* (Marmigère & Ernfors, 2007), *SNCG*, *MAFB*, *ONECUT2*, *EBF3*, *NNAT*, *ISL1* *SNX10*, *SOX4*, *S100A10*, *PHOX2A*, *TAC1*, *RET*, *PCP4*, *SLC5A7*, *PRPH*, *TPD52L1*, and especially *PIEZ01*, *ATF5*, *ZFHX3*, and *SLC5A7*. Dorsal root ganglion neurons (DRGNs) which can be classified into mechanosensitive, nociceptive, and proprioceptive neurons are marked by *NTRK1* (not expressed), *NTRK2*, and *NTRK3* expression, respectively. The presence of DRGNs is further supported by the expression of *NTRK2-3*, *POU4F1*, *NGFR*, *FGF13*, *MT1F*, *ISL1*, and *SLC17A6*, *SLC17A7*. Nevertheless, neurons lacked expression of *RUNX3*, while unexpectedly expressing *RUNX1*. Besides, NCCs differentiate into sympathetic neurons which is indicated by *MALAT1*, *ARGLU1*, *PPP1R1Z*, *ISL1*, *ZFHX3*, *RBFOX1*, *PRPH*, *NEFL*, *LBH*, *DOC2B*, *RPS4Y1*, and *B2M* expression. However, sympathetic genes such as *FOXD3*, *HOXC9*, *HOXB6-8* were not detected in the sequenced cells.

### 4.3 Transcriptional activity and regulation in human day-49 iNeurons

After successful reprogramming of over 20 primary human cells and the differentiation of those into disease relevant iNeurons, the next aim of this thesis was to investigate differential behavior in case-control studies. Hence, transcription (bulk RNA-Seq) and open chromatin (bulk ATAC-Seq) experiments were performed. In order to model  $\text{Ca}^{2+}$  induced alterations on the transcriptional and epigenetic landscape, a high potassium chloride solution (hKCl) was applied for 5 h on the day of the harvest. Furthermore, activity induced chromatin regions (ARChs) were linked to activity regulated genes (ARGs). The analysis of the activity induced regulation of gene expression was performed to follow the line of research that indicates an impacted synaptic signaling in schizophrenia.

### 4.3.1 Schizophrenia altered expression of 5 genes in day-49 iNeurons

iNeurons of 11 cell lines were differentiated in 2 batches into day-49 iNeurons (for details on the data set see 4.2.3). After stimulation with hKCl for 5 h (Figure 19 A), cells were harvested, RNA was isolated, and RNA-Seq libraries were prepared (3.2.6.2, 3.5.1, and 3.5.3). DESeq2 analysis of differential expression between case and healthy control, using stimulation as a covariate, annotated 12 differentially expressed genes (Table 23) using the significance threshold of p-value  $\leq 0.05$ , and 0.5 for log2(fold change). 3 genes were downregulated in control compared to schizophrenia (*OR8A1*, *PAX8-AS1*, *MTRNR2L8*) and 2 genes were upregulated (*GUSBP1* and *ENSG00000280255*).

ENSEMBL gene ID	HGNC symbol	Base mean	Log2(FC)	P-value	P.adjusted
<i>ENSG00000255823</i>	<i>MTRNR2L8</i>	507.99	-2.50	<i>1.53E-07</i>	<i>0.002</i>
<i>ENSG00000189223</i>	<i>PAX8-AS1</i>	41.68	-2.70	<i>6.18E-07</i>	<i>0.005</i>
<i>ENSG00000280255</i>	-	57.36	0.60	<i>1.07E-06</i>	<i>0.007</i>
<i>ENSG00000183666</i>	<i>GUSBP1</i>	79.15	1.01	<i>2.72E-06</i>	<i>0.014</i>
<i>ENSG00000196119</i>	<i>OR8A1</i>	10.16	-2.49	<i>8.82E-06</i>	<i>0.028</i>
<i>ENSG00000103534</i>	<i>TMC5</i>	41.24	-0.96	<i>2.90E-05</i>	0.055
<i>ENSG00000128335</i>	<i>APOL2</i>	57.35	-0.92	<i>4.01E-05</i>	0.067
<i>ENSG00000152117</i>	<i>SMPD4BP</i>	257.44	0.65	<i>4.51E-05</i>	0.067
<i>ENSG00000105048</i>	<i>TNNT1</i>	54.80	0.60	<i>5.74E-05</i>	0.074
<i>ENSG00000211445</i>	<i>GPX3</i>	91.37	-0.89	<i>5.82E-05</i>	0.074
<i>ENSG00000164756</i>	<i>SLC30A8</i>	10.49	-2.63	<i>8.64E-05</i>	0.098
<i>ENSG00000180543</i>	<i>TSPYL5</i>	164.38	-0.52	<i>8.86E-05</i>	0.098

**Table 23 Significantly dysregulated genes between case and control iNeurons**

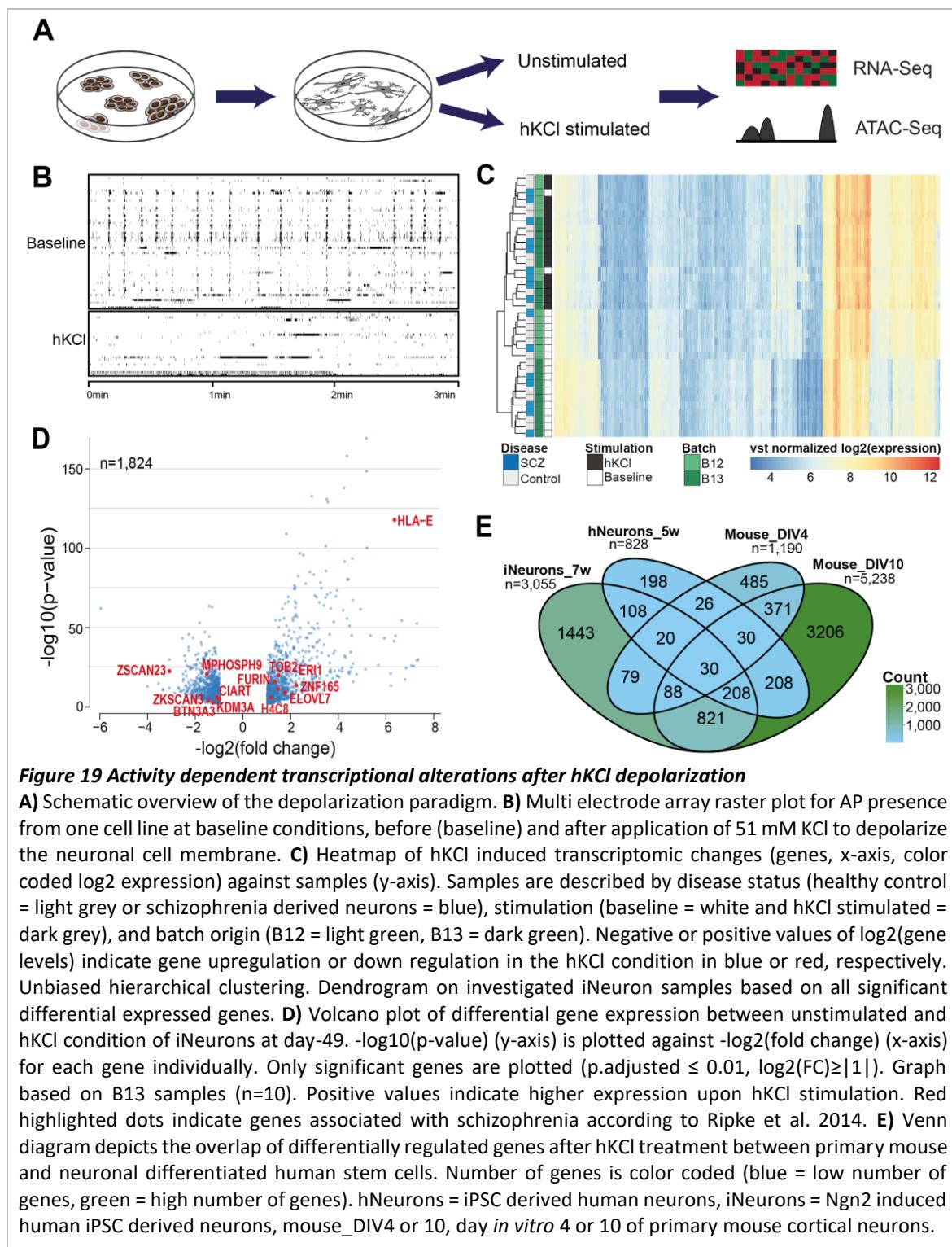
5 genes were significantly (adjusted p-value (p.adjusted)  $\leq 0.05$ ) differentially regulated between healthy control and schizophrenia patients cell lines and are denoted in italic. FC = fold change

### 4.3.2 Induction of gene expression after a sustained depolarization stimulus

High potassium chloride (hKCl) was added to the culture media for 5 h on day-49 (Figure 19 A). hKCl treatment depolarizes cell membranes and in turn induces the influx of the second messenger  $\text{Ca}^{2+}$ . Activity dependent gene regulation (bulk RNASeq data set described in 4.2.2 and 4.3.2) was compared between baseline and hKCl treated iNeurons. Next, the here described transcriptional alterations were compared to DEGs derived from depolarized primary mouse neurons and iPSC derived neurons (Tyssowski et al., 2018).

As a first step, to visualize the impact of hKCl on the neuronal activity, iNeurons were plated on MEA devices and were recorded at baseline activity before and right after the stimulation with hKCl. As described before (baseline inspections at Figure 7 A, B), network activity was present at day-49 (Figure 19 B upper track). After stimulation, the network

bursts were ablated, AP generation was uncoordinated, and AP activity was detected in fewer channels (Figure 19 B lower track).



Differential genes were detected in a pairwise manner (same batch, same cell line, two stimulation conditions) with DESeq2. The inclusion of the disease status to the linear model did not reveal significantly different genes in the stimulation condition between case and control (data not shown). In total, 3,055 genes were differentially regulated after hKCl

stimulation compared to baseline activity. The heatmap of significantly differentially regulated genes (3,055, FC  $\geq$  0.5, p.adjusted  $<$  0.05) showed clear distinct transcriptional alterations (Figure 19 C).

After hKCl treatment, well-known immediate early genes such as *NPAS4*, *FOSB*, *FOS*, *JUNB*, and *IER2*, were significantly upregulated (p.adjusted  $<$  0.00008). Several of the delayed primary regulated genes (PRGs) such as *PCSK1*, *BDNF*, *ATF3*, *TAC1*, *NR4A3*, and *NR4A2* were also significantly upregulated (p.adjusted  $<$  1 \* 10<sup>-10</sup>) after sustained hKCl stimulation. Potassium channels such as *KCNA1*, *KCNJ3*, *KCNJ14*, *KCNJ2*, *KCNAB2*, *KCNE4*, *KCNC1*, *KCNB2*, *KCNK1*, *KCNK9*, and *KCNA5* were significantly increased (p.adjusted  $<$  2.2 \* 10<sup>-4</sup>) in expression in the stimulated condition. Genes directly responsive for chromatin modifications such as histone deacetylases were significantly (p.adjusted  $<$  1.2 \* 10<sup>-4</sup>) induced (*HDAC1*, *HDAC10*, and *HDAC2-AS1*) and reduced (*HDAC5*) upon hKCl stimulation (Tyssowski et al., 2018).

hKCl depolarization induced gene expression enriched for GO terms related to polymerase II activities: sequence specific regulatory binding and transcription factor activity. Additionally, toll-like-receptor 1/2,2,5,6/2,7/8,10 cascade, MAPK, and serine/threonine kinase pathways were induced. Apart from these, GO pathways indicating cellular response to external stimuli or hormones were also enriched for genes upregulated in the hKCl condition. Besides the pathway enrichment analysis, the significant 1,824 ARGs (B12 samples only) were analyzed for their association with schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2014). For the schizophrenia associated genes, 7 were upregulated (*HLA-E*, *ERI1*, *TOB2*, *FURIN*, *H4CB*, *ELOVL7*, *ZNF165*) and 6 (*MPHOPH9*, *ZSCAN23*, *ZKSCAN3*, *BTN3A3*, *CIART*, *KDM3A*) were downregulated in the hKCl condition compared to the unstimulated condition (Figure 19 D). It is noteworthy, that *HLA-E* (a gene associated with schizophrenia) is also part of the component loading for PC1 for the bulk RNA-Seq experiments.

ARGs of iNeurons were compared to ARGs of previously published sustained depolarization data sets from iPSC differentiated human excitatory cortical forebrain (hNeurons) and primary mouse neurons. To account for the developmental maturity of the iPSC derived neurons, two developmental time points were used for the primary mouse neurons: DIV4 and DIV10. There was a substantial overlap of ARGs between all investigated neuronal subtypes. The lowest overlap of ARGs was detected between primary mouse DIV4 and the

iNeurons. iNeurons shared around 13 or 15 % ARGs of mouse DIV10 or hNeurons, respectively. Only 30 genes (1 % ARGs of iNeurons) were shared between all tested samples. The overlap of ARGs was highest between iNeurons, hNeurons, and mouse DIV10 neurons (6 % of iNeurons, 25 % hNeurons). 63 % and 50 % of ARGs detected in hNeurons and iNeurons were detected also in the mouse neurons, respectively (Qiu et al., 2016). The human specific genes *PTDSS1*, *CCNH*, *KCTD1*, and *TRAFD1* were regulated into the same direction in the human data sets (Qiu et al., 2016). On the other hand, *TRAFD1* and *ATP1B1*, were not significantly regulated in iNeurons but in hNeurons, even though *ATP1B3* and *ATP8B3* were ARGs of iNeurons (Figure 19 E).

### 4.3.3 Open chromatin upon depolarization of neurons

Gene expression is tightly regulated by the activity of gene regulatory elements (GREs) such as enhancers and promoters. Here, open chromatin which indicates the accessibility of chromatin was measured with ATAC-Seq (Figure 20). Accessibility of chromatin was compared between iNeurons (day-49) differentiated from 5 cell lines derived from patients suffering from schizophrenia and 5 healthy control derived cell lines (Table 24). iNeurons were used because enhancer and promoter activity similar to transcription are highly cell specific. Furthermore, the before mentioned hKCl stimulation paradigm was applied (Figure 19 A). On average 87,700,972 high quality reads per library were aligned to the human genome with an average duplicated read fraction of 28 %.

The first two PCs segregated the ATAC-Seq-samples according to the stimulation paradigm and not according to the disease status (Figure 20 A). Similar to the bulk RNA-Seq, in the ATAC-Seq data set, differential accessible chromatin locations between case and control occurred in only a few instances. Additionally, most of the significantly detected chromatin marks were driven by 1) a subgroup of samples of the two disease status groups or 2) mapped to loci on the genome which contained low chromatin peak signal in one or both conditions (data not shown). In total, independent of the disease status, 228,841 chromatin marks were detected combining both stimulation conditions (Figure 20 B).

The presence of the depolarization stimulus altered chromatin accessibility compared to the unstimulated condition at 5,520 (more accessible) or 2,115 (less accessible) enhancer applying a 5 % FDR cutoff and a an absolute  $\log_2(FC) \geq 1$ . Activity regulated chromatin marks

(ARCh) were analyzed in a pairwise setup independent of their disease status. The combined effect of disease and stimulation detected fewer ARCh marks (236 less accessible and 52 more accessible ARCh, 1 % FDR,  $|\log_2(\text{FC})| > 0.5$ ) than detected by the disease status alone (601 more, 300 less accessible, 5 % FDR,  $|\log_2(\text{FC})| > 1$ ) (Figure 20 C). Subsequently, the open chromatin marks from day-49 iNeurons were investigated for their enrichment in expressive quantitative expression loci (eQTL) from postmortem brain tissue (Fromer et al., 2016). eQTLs impact gene expression depending on the genotype present at the genetic variant. There is no significant difference in the percentage of the overlap with eQTL loci and ARCh against all ATAC peaks (Figure 20 D).

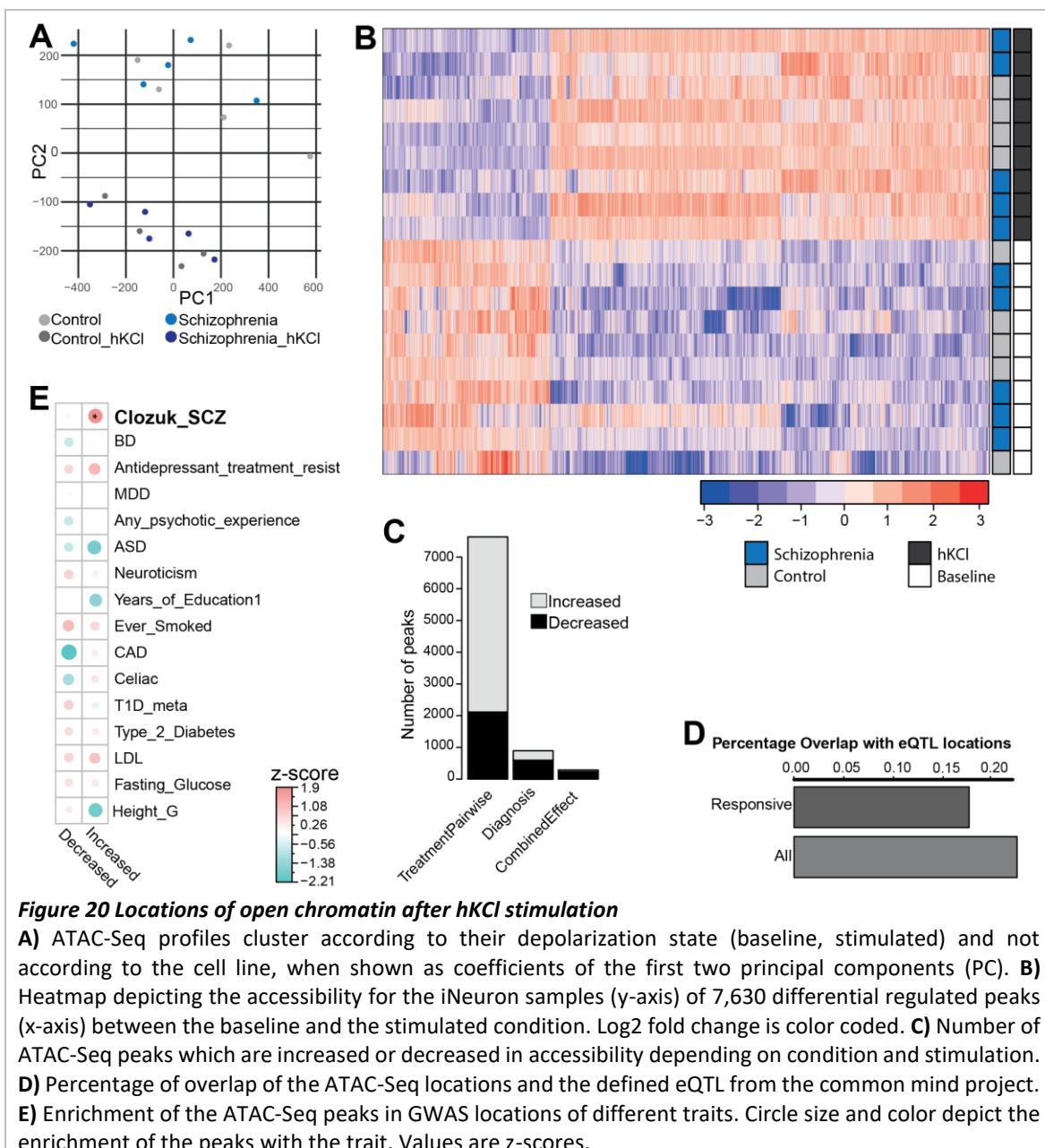
Donor	Stimulation	Aligned reads	Duplicate reads in %
C1	Unstimulated	115,940,690	30.3
C1	hKCl	70,943,769	27.0
C2	Unstimulated	67,181,098	25.6
C2	hKCl	91,435,053	25.3
C3	Unstimulated	78,887,220	27.8
C3	hKCl	117,751,147	25.2
C4	hKCl	89,894,968	26.2
C5	Unstimulated	193,083,037	22.8
C6	Unstimulated	42,698,183	26.4
S1	Unstimulated	63,873,105	27.8
S1	hKCl	130,546,432	28.3
S4	Unstimulated	48,288,646	29.4
S4	hKCl	147,410,714	28.0
S6	Unstimulated	68,571,309	25.9
S6	hKCl	77,142,369	26.9
S7	Unstimulated	34,838,035	44.7
S7	hKCl	76,157,494	30.8
S8	Unstimulated	49,980,661	26.6
S8	hKCl	101,694,542	25.3

**Table 24 Overview of the ATACSeq samples from day-49 iNeurons**

6 healthy control and 5 schizophrenia patient derived cell lines were differentiated to iNeurons. Stimulation with hKCl was performed on the day of harvest (day-49). The sequencing meta data (aligned and duplicate reads) is annotated per sample.

To further verify iNeurons as a valuable tool to investigate schizophrenia genetics, the enrichment of schizophrenia associated SNPs in iNeuron open chromatin regions was calculated. This line of research was followed, due to the fact that H3K4me3 marks (active promoter regions) of NEUN (*RBFOX3*) sorted post-mortem nuclei from hESC differentiated neurons (Tansey & Hill, 2018), and the cortex in general (Finucane et al., 2018), are enriched for schizophrenia associated variants. Hence, activity specific elements were investigated

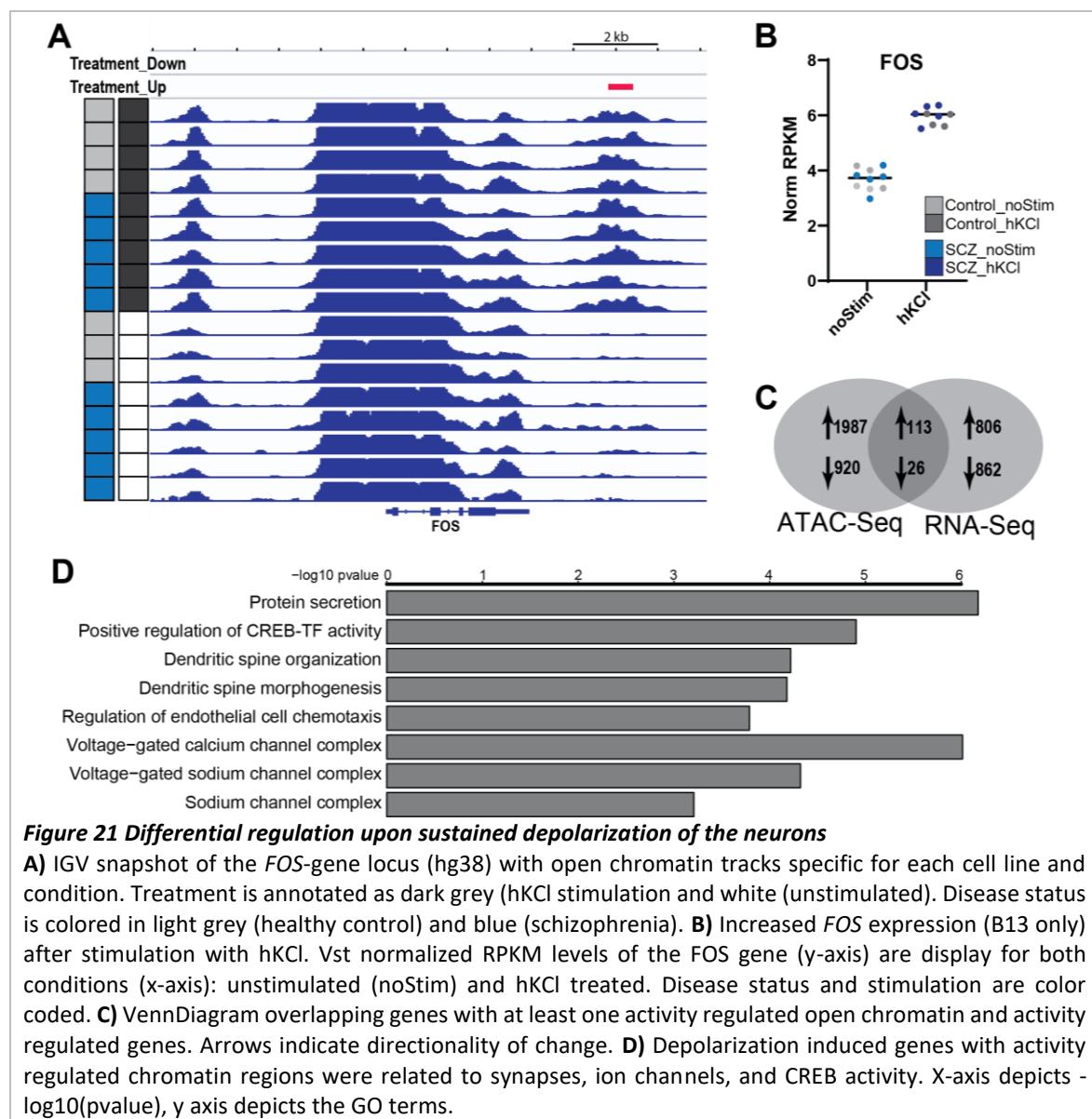
to estimate their genome-wide polygenic contribution to schizophrenia heritability. GWAS summary statistics from several polygenic traits and schizophrenia were used to estimate the partitioning heritability of open ARCh (Laura Jimenez Barron). The regions defined by ATAC-Seq were segregated into two groups, one for increased (5,520 peaks) and one for decreased accessibility (2,115 peaks) after hKCl stimulation (Figure 20 E). Variants located in more accessible regions upon stimulation explained a significant amount of schizophrenia heritability ( $p = 0.02847$ ). Variants located in decreased signal of ATAC peak regions were not significantly associated with schizophrenia ( $p = 0.41972$ ).



#### 4.3.4 Proximity of activity regulated chromatin states and transcription levels

hKCl stimulation induced alterations in open chromatin marks and gene expression. Open chromatin can be indicative for the presence of enhancer and promoter regulating gene expression. Therefore, the presence of ARCh loci was correlated with the presence of nearby ARGs.

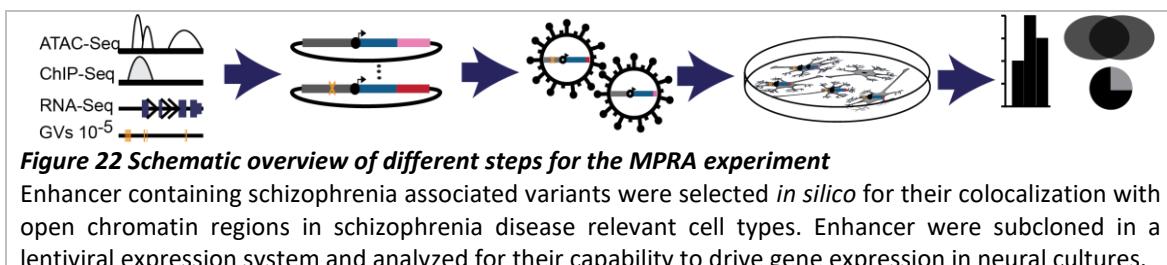
The IGV snapshot from the genomic location of the immediate early gene *FOS* illustrates the stimulation dependent appearance of an open chromatin mark (red bar) in day-49 iNeurons independent of the disease status. This activity regulated open chromatin mark is located distal of the 3'UTR of the *FOS* gene (Figure 21 A). This open chromatin is paralleled by a significant ( $p.\text{adjusted} = 8.79*10^{-24}$ ) increase in *FOS* gene expression after hKCl simulation independent of disease status (Figure 21 B, B13).



ARCh regions were associated statistically for their location within a 100 kb window in both directions of the transcription start sites (TSS) of nearby genes. In total, there were 3,046 genes (based on B13) which contained at least one ARCh loci in proximity to its TSS. The majority of the genes, though, were not significantly altered in their gene expression by hKCl depolarization (2,907 genes) despite an altered accessibility of chromatin marks nearby. 1,668 genes were differentially regulated without containing ARCh within a 100 kb distance to their TSS. The expression of 113 or 26 ARGs was up- or downregulated while containing an ARCh in close vicinity with the same directionality, respectively (Figure 21 C). The depolarization induced GO pathways of differentially regulated genes which also contained in their vicinity a differentially regulated chromatin mark were significantly enriched for genes coding for (voltage gated) ion channels, spine morphology and organization, protein secretion, and activity of the CREB-TF (Figure 21 D).

#### 4.4 Differential enhancer activity in neuronal cell models

Next, this thesis aimed to functionally annotate schizophrenia associated variants located in enhancer regions. For this purpose, schizophrenia variants in open chromatin regions were analyzed for their general and allele specific activity in different neural contexts (Figure 22). Enhancer activity was measured in a parallel fashion in several schizophrenia disease relevant cell types: human NPCs, human iNeurons, and primary mouse cortical neurons. In addition, stimulation paradigms were performed before the harvest of the samples that mimicked the gene regulatory landscape changes upon network activity.



In brief, selected schizophrenia relevant small enhancer fragments were synthesized, barcoded, and subcloned into a lentiviral expression plasmid. Enhancer fragments integrated into the genome via lentiviral particles. Several thousand barcoded putative enhancer fragments were analyzed in parallel for their individual putative enhancer activity. This enhancer activity was measured as the capability of the enhancer to drive the

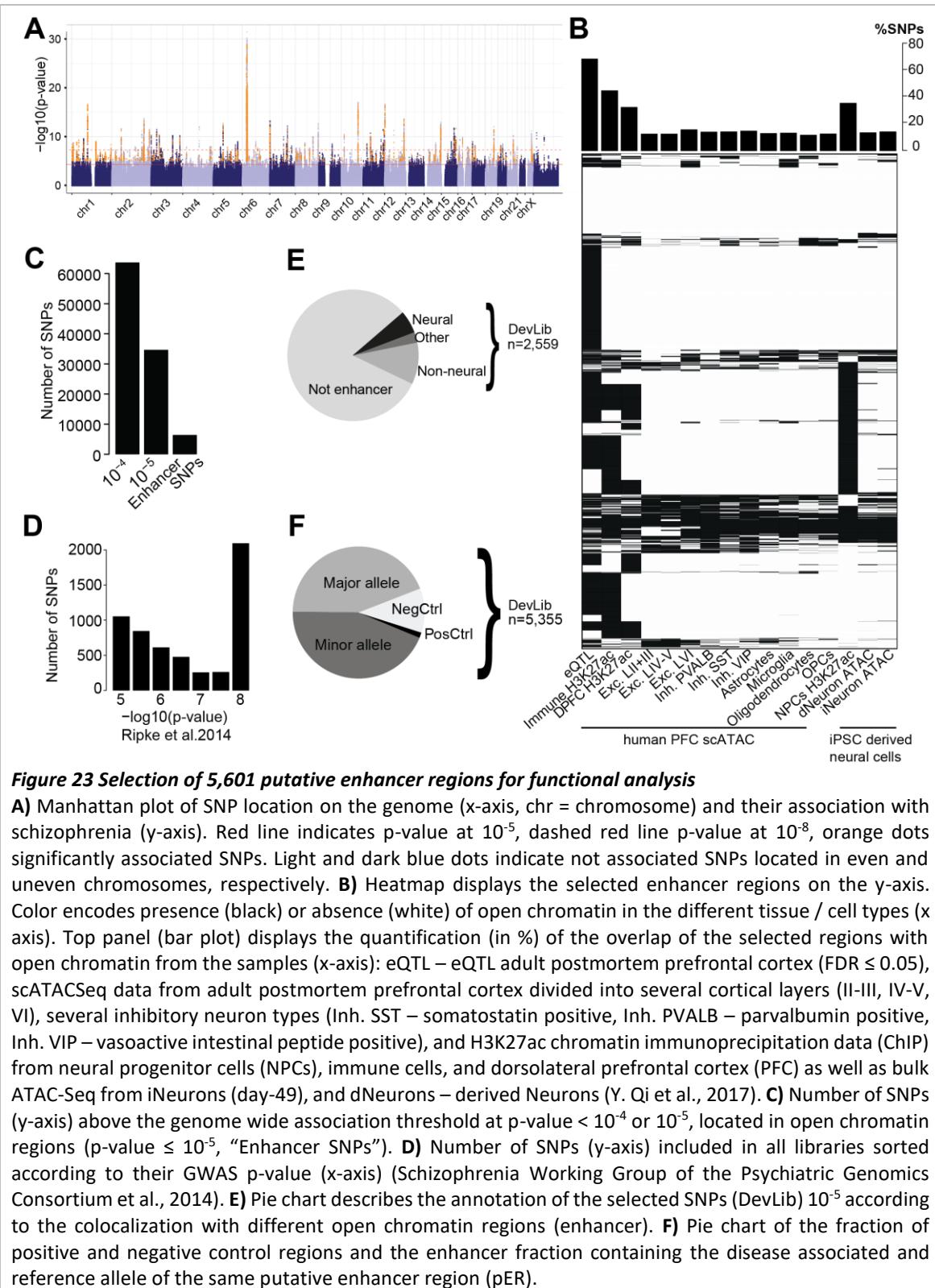
transcription of a reporter gene. Those genetic variants which affected the activity of the enhancers were termed expression modulating variants (emVars) (Tewhey et al., 2016).

#### 4.4.1 Library design of the different MPRA libraries

The first step to achieve the aim to functionally annotate schizophrenia associated enhancer regions was to select meaningful loci for further investigation. In total 34,634 SNPs were associated by GWAS with schizophrenia ( $p\text{-value} < 10^{-5}$ ) (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2014) (Figure 23 A). Subthreshold associated SNPs ( $10^{-5} \leq p\text{-value} \leq 10^{-8}$ ) were included in this study, due to the fact that they were highly functional in another study and affected the cardiac QRS duration and QT interval (X. Wang et al., 2016). Moreover, subthreshold SNPs can achieve significant association with larger cohort sizes.

The preselected schizophrenia associated SNPs were also required to colocalize with at least one open chromatin mark in at least one of the considered tissues or cell types (see 3.9.1) or to be annotated as an eQTL in human postmortem DLPFC. Around one fourth of the schizophrenia associated SNPs ( $p\text{-value} \leq 10^{-5}$ ) were located in open chromatin regions (Figure 23 B): H3K27ac ChIPSeq data was derived from human PFC, iPSC derived neural precursor cells, and immune system related cell types such as microglia, CD25<sup>+</sup> and CD45<sup>+</sup> leukocytes. Leukocytes were further sub-grouped into CD34<sup>+</sup> hematopoietic stem cells, CD8<sup>+</sup> cells, and CD4<sup>+</sup> cells. Additionally, the above-mentioned bulk ATAC-Seq data from iPSC derived iNeurons and in-house generated scATAC-Seq data from human postmortem PFC (Dr. Miriam Gagliardi) were used to annotate open chromatin regions in schizophrenia relevant cell types (Figure 23 C).

These SNP inclusion criteria identified 5,601 schizophrenia associated SNPs for testing. The chosen SNPs overlapped in 80 % with eQTL in post-mortem PFC and in around 50 % with active enhancer marks of postmortem PFC and other cell type specific open chromatin marks. The selected variants were distributed over the entire genome and covered all annotated loci containing schizophrenia associated variants (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2014). Around half of the selected SNPs are associated with schizophrenia at a  $p\text{-value}$  above the significance threshold  $p\text{-value} \leq 10^{-8}$  (Figure 23 D).



In summary, three partially overlapping libraries (Table 25) were generated to analyze the activity of 5,601 schizophrenia associated putative enhancer regions (pERs). 215 elements were shared between ValLib and DevLib, or ValLib and eQTLlib in order to evaluate cross-library reproducibility. The three libraries varied in size and focused on different background data sets: 1) DevLib: neuronal development and immune system specific loci

(non-neuronal accessible chromatin in Figure 23 E), 2) eQTLLib: genotype related tissue expression (GTEX) and postmortem specific locations, and 3) ValLib: a subset of pERs also present in DevLib and eQTLLib, locations with closed chromatin, and regions with SNPs not associated with schizophrenia and not located in neural open chromatin. Each library contained putative enhancer fragments (pEFs) with the major and the minor allele, and negative and positive control fragments (Figure 23 F, DevLib) which served as calibrators for the general activity of each enhancer or element (Figure 29 A, DevLib).

Library name	Length in bp	Number of enhancer regions	Number of control enhancer regions	Manufacturer
Development (DevLib)	160	2,559	159 negative and 78 positive	Agilent Technologies
eQTLLib	160	4,325	159 negative and 78 positive	Twist Bioscience
Validation (ValLib)	170	333	28 positive, 4 negative, 24 random sequences and 100 random SNPs	Twist Bioscience

**Table 25 Overview of the number of putative enhancer fragments in each MPRA**

Three individual MPRA libraries contained putative enhancer regions and control fragments.

#### 4.4.2 Successful generation of barcoded enhancer libraries

After the *in-silico* selection of the putative enhancer regions (pERs) of interest by Dr. Michael Ziller, the putative enhancer fragments (pEFs) were synthesized as an oligo pool by Twist Bioscience or Agilent Technologies. Those pEFs had to be cloned in several steps into a lentiviral expression vector (Figure 24 A). The quality of the DNA preparation was monitored during the cloning process.

Colony forming units (CFUs) were counted per transformation to ensure sufficient complexity of the generated plasmid pools. For step2 (Figure 24 A) of the library generation 1.3 mil, 6.7mil, or 6.6 mil CFU were transformed for the DevLib, the eQTLLib, or the ValLib, respectively. This approach accounted for a minimum of 50-fold excess of CFUs over pEFs and was followed to secure a coverage of on average 30 barcodes per pEF considering a normal distribution of the number of barcodes per enhancer. Libraries were pooled according to their concentration and their CFU count after each cloning step for the eQTLLib and the ValLib. The DevLib transformations, though, were pooled directly after the transformation without accounting for different transformation efficiencies. At step3 (Figure 24 A), a 20-fold or 30-fold increase in CFUs were generated for the eQTLLib or ValLib, respectively. The number of CFUs for the DevLib were increased by nearly a 100-

fold to 128 mil. The final lentiviral library (step4, Figure 24 A) contained 178 mil, 227 mil, or 1,500 mil CFUs in the DevLib, the eQTLLib, or the ValLib, respectively (Table 26).

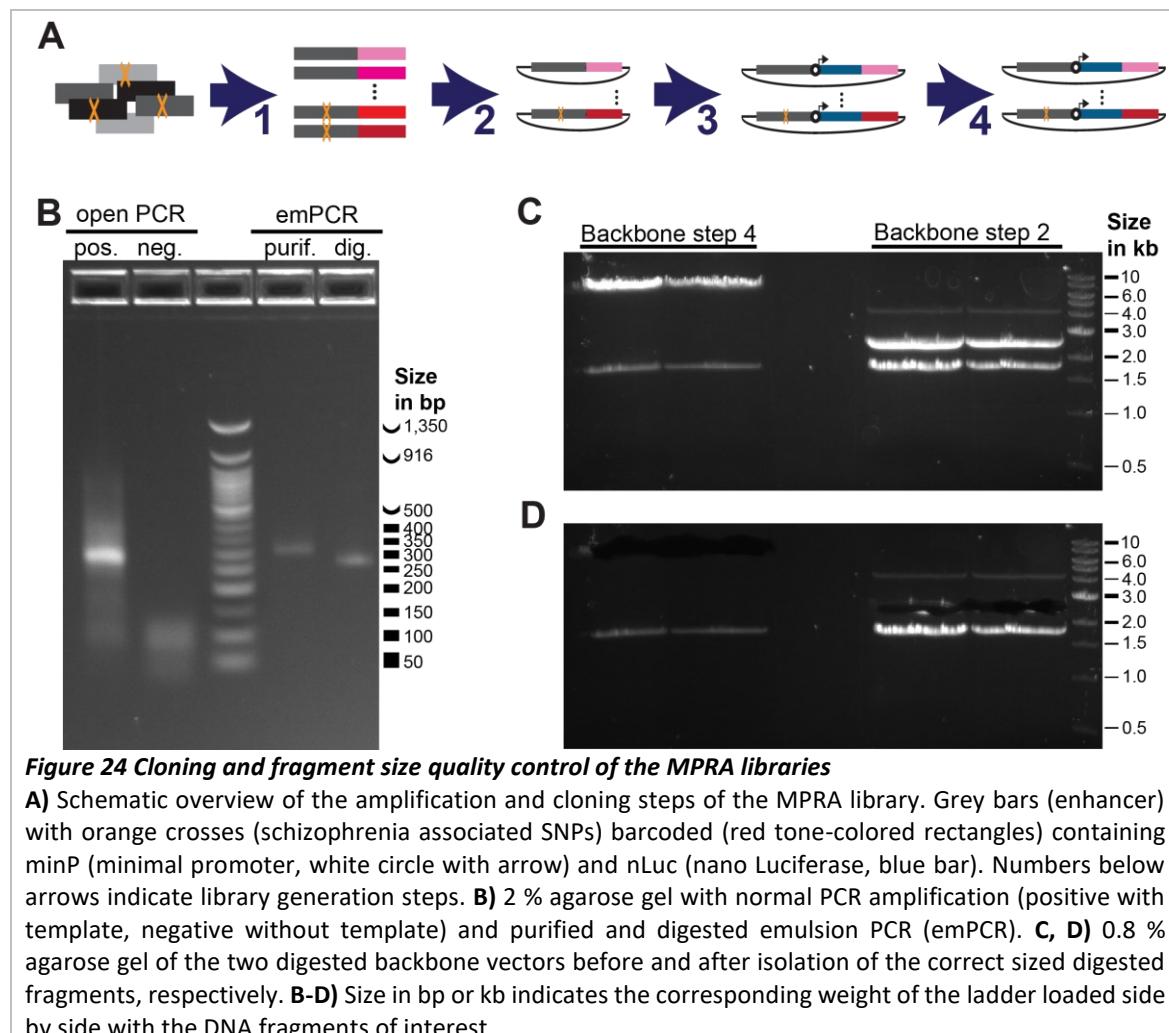
Step	DevLib	eQTLLib	ValLib
2	1.3	2.9 / 3.8	0.7 / 1.4 / 4.7
3	2.5 / 11.0 / 8.3 / 4.9 / 6.5 / 20.0 / 10.0 / 16.0 / 6.6 / 20.0	10.1 / 10.9 / 9.4 / 6.5 / 6.6 / 7.2 / 8.7 / 6.5 / 5.4 / 9.8 / 7.8 / 13.8 / 4.5 / 12.0	81.0 / 69.0 / 63.0
4	7.3 / 4.0 / 4.5 / 8.1 / 6.9 / 1.1 / 7.8 / 0.7 / 0.8 / 14.6 / 1.0 / 1.1 / 0.34 / 8.9 / 4.4 / 5.2 / 0.5 / 9.5 / 20.7 / 2.1 / 17.0 / 22.8 / 11.2 / 17.4	7.4 / 7.9 / 5.9 / 4.4 / 8.0 / 5 / 11.2 / 4.1 / 14.6 / 14.1 / 11.0 / 12.4 / 6.9 / 8.1 / 7.9 / 5.3 / 5.2 / 4.5 / 5.7 / 6.4 / 4.4 / 5.2 / 3.5 / 6.8 / 5.5 / 13.6 / 4.5 / 2.4 / 5.2 / 4.4 / 5.0 / 4.7 / 8.7	18.1 / 41.9 / 51.9 / 22.2 / 71.9 / 49.4 / 21.3 / 16.0 / 25.8 / 30.0 / 41.9 / 10.8 / 79.7 / 25.2 / 26.5 / 29.0 / 25.1

**Table 26 Number of colony forming units for the three MPRA experiments**

Steps denotes the cloning steps according to Figure 24 A. Averaged number of colony forming units (same transformation on several agar plates) for the individual transformations is rounded to one decimal digit and separated by “/”.

To ensure a complex randomly barcoded plasmid library, emulsion PCRs (emPCRs) were performed (step1, Figure 24 A). This method increased the specificity of the amplification and decreased the likelihood of an unwanted PCR bias towards a small fraction of enhancer regions and barcodes. Furthermore, the PCR based barcoding enabled the synthesis of longer enhancer regions (160 bp instead of 144 bp) (Figure 24 B). The amplified PCR product was purified and digested with restriction enzymes to ligate the enhancer into the plasmid backbone in one direction (Figure 24 A step2).

The correct fragment length and the activity of the used restriction enzymes were verified during the restriction digests. In order to minimize the occurrence of religations of digested backbones, digested plasmid backbones were 1) dephosphorylated (rSAP incubation) and 2) purified over agarose gels (0.8 %, Figure 24 C-D). Furthermore, vector religation was monitored in a separate ligation reaction for which the DNA insert was replaced with water. For the first (step2) and the third ligation (step4), a new backbone was used, therefore, an unwanted religation rate of maximum 1,000 copies per transformation was tolerated. For step3, less than 10 % of the total CFUs were accepted to be present in the religation control. Due to the fact that two restriction enzymes were necessary to digest the backbone in close proximity, the sensitivity of the second ligation (step3) was low due to the presence of single digested plasmid backbones. However, the next step transferred the 847 bp enhancer-nanoLuc-barcode fragment into a new plasmid backbone and, thus, was used as a selection step for full length enhancer-nanoLuc-barcode DNA fragments.

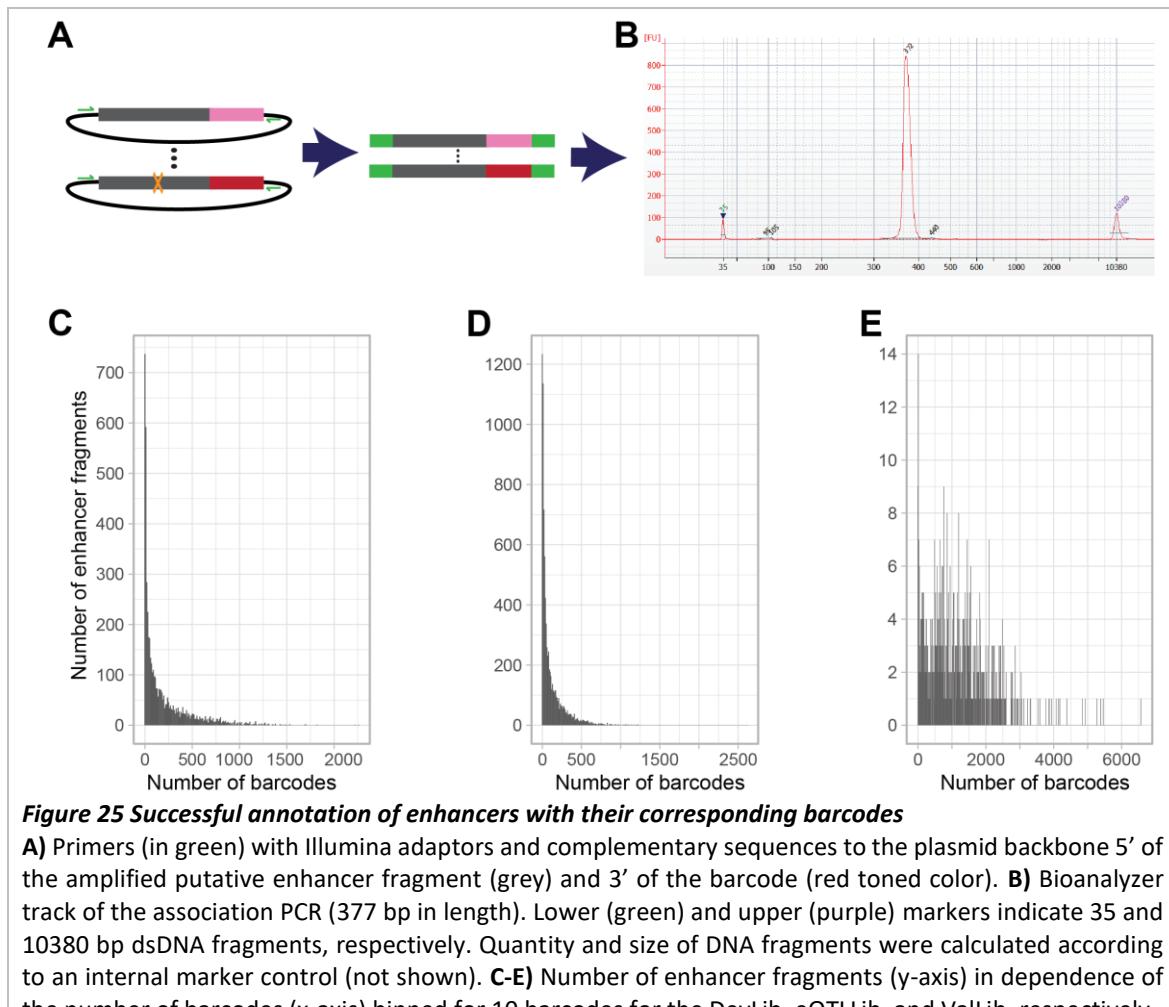


#### 4.4.3 Efficient strategy for enhancer barcoding and retrieval of barcodes

After the confirmation of sufficiently high number of CFUs in the three MPRA libraries after each cloning step, each library was screened for 1) barcode complexity and 2) the association of barcodes with the pEFs present in the library (Figure 25). Absolut barcode recovery for each initial plasmid library (after step2 of Figure 24 A) was analyzed. Enhancer-barcode associations were amplified directly from the plasmid libraries with primers containing Illumina sequencing adapters (Figure 25 A). Quality control and quantification of amplification libraries were performed on a Bioanalyzer (Figure 25 B) and PCR based (not shown), respectively.

5,964,220, or 3,127,233, or 2,033,733 valid reads were sequenced for the DevLib, eQTLlib, or ValLib, respectively. The number of considered barcodes (criteria see 3.9.7) was highest in the eQTL Lib (927,647 barcodes) and lowest in the DevLib (370,705 barcodes). A median

barcode of 29, 45, and 500 encoded for each pEF in the DevLib, eQTLLib, and ValLib, respectively (Figure 25 C-E).



Subsequently, the three MPRA libraries were infected into disease relevant cell types: iPSC derived NPCs and iNeurons, and primary mouse cortical neurons. Moreover, activity dependent enhancer activity was analyzed in stimulation (BIC, hKCl), silencing (TTX, silent) and baseline (unstimulated, noStim) conditions (Table 27).

Cell type	NPCs		iNeurons		Primary mouse neurons		
	Condition	Baseline	Baseline	hKCl	Silent	Baseline	Excited
DevLib	3	3	3	3	5	4	3
eQTLLib	-	4	-	-	4	3	4
ValLib	-	-	-	-	3	4	4

**Table 27** Number of sample replicates per cell type and MPRA

Each condition of each cell type contained between 3 – 5 technical replicates of the same MPRA library.

Barcode containing cDNA was amplified after RNA isolation and mRNA enrichment (Figure 26 A). Amplification primers contained complementary DNA sequences for the 3'end of the nLuc gene and 3'end of the barcode and Illumina sequencing adapters (i5 and i7). Valid barcode reads per library derived from the cell samples was higher in the DevLib (mean

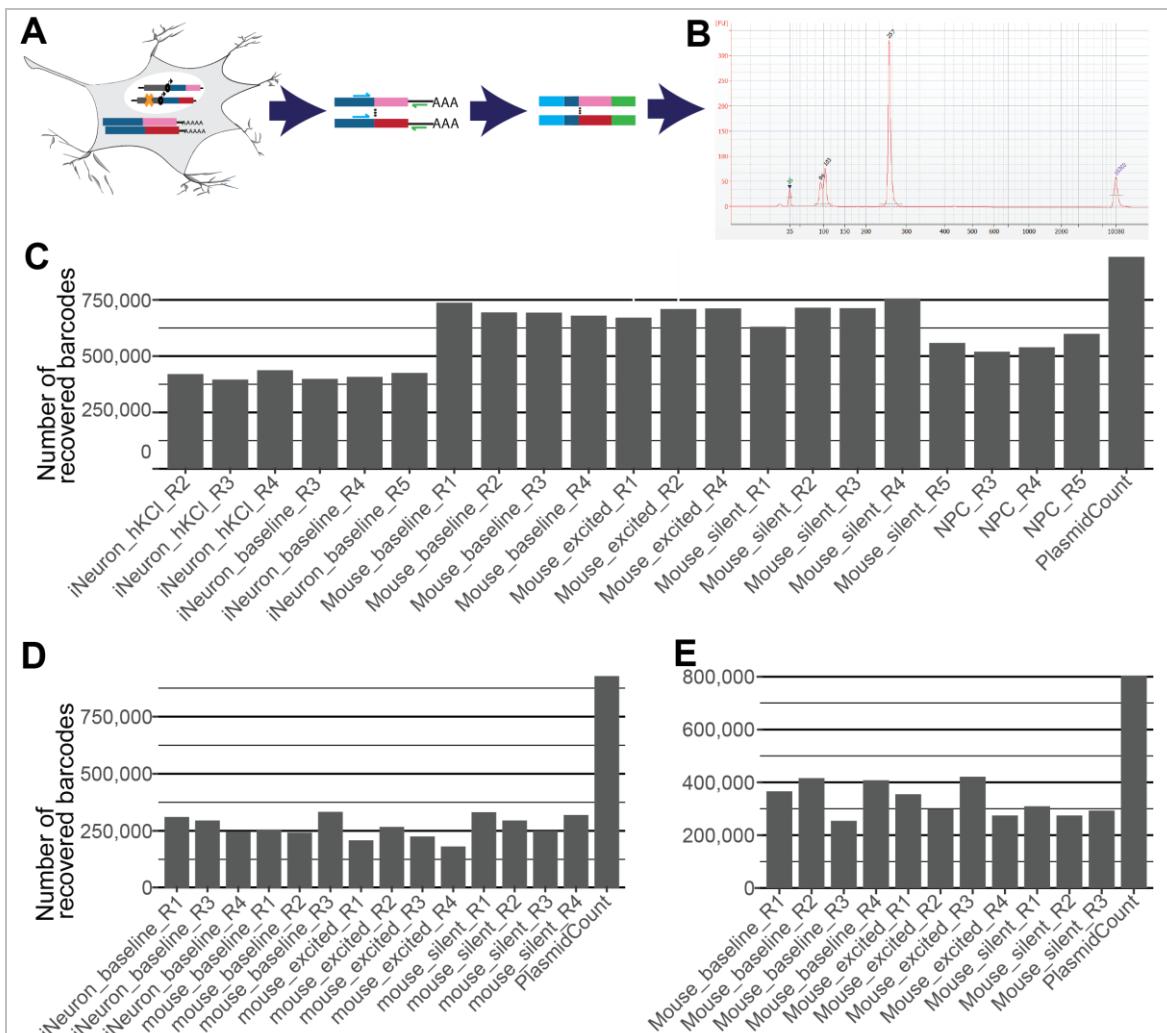
11,740,495; 7,341,132 median reads), than in the other two libraries. The average valid number of barcode reads for the eQTLLib and ValLib was 1,433,939 and 1,241,620, respectively (Table 28).

MPRA library	Replicate number	Condition	Batch	Number of valid barcode reads
DevLib	2	iNeuron_hKCl	B	4,945,388
DevLib	3	iNeuron_hKCl	B	4,396,012
DevLib	4	iNeuron_hKCl	B	5,156,810
DevLib	3	iNeuron_baseline	B	6,011,467
DevLib	4	iNeuron_baseline	B	5,160,270
DevLib	5	iNeuron_baseline	B	5,198,206
DevLib	1	Mouse_baseline	C	18,753,219
DevLib	2	Mouse_baseline	C	13,103,616
DevLib	3	Mouse_baseline	C	11,064,016
DevLib	4	Mouse_baseline	C	9,066,652
DevLib	1	Mouse_excited	C	9,577,029
DevLib	2	Mouse_excited	C	14,776,275
DevLib	4	Mouse_excited	C	15,330,176
DevLib	1	Mouse_excited	C	6,606,366
DevLib	2	Mouse_silent	C	18,829,511
DevLib	3	Mouse_silent	C	16,729,358
DevLib	4	Mouse_silent	C	62,646,918
DevLib	5	Mouse_silent	C	3,548,096
DevLib	3	NPCs	A	3,277,298
DevLib	4	NPCs	A	5,032,575
DevLib	5	NPCs	A	7,341,132
eQTLLib	1	iNeuron_baseline	D	3,432,735
eQTLLib	3	iNeuron_baseline	D	2,062,993
eQTLLib	4	iNeuron_baseline	D	2,740,128
eQTLLib	1	Mouse_baseline	E	928,604
eQTLLib	2	Mouse_baseline	E	929,810
eQTLLib	3	Mouse_baseline	E	969,339
eQTLLib	1	Mouse_excited	E	906,786
eQTLLib	2	Mouse_excited	E	1,126,577
eQTLLib	3	Mouse_excited	E	932,113
eQTLLib	4	Mouse_excited	E	1,018,660
eQTLLib	1	Mouse_silent	E	1,702,922
eQTLLib	2	Mouse_silent	E	1,011,122
eQTLLib	3	Mouse_silent	E	813,330
eQTLLib	4	Mouse_silent	E	1,500,033
ValLib	1	Mouse_baseline	F	1,437,474
ValLib	2	Mouse_baseline	F	2,017,871
ValLib	3	Mouse_baseline	F	915,747
ValLib	4	Mouse_baseline	F	1,879,772

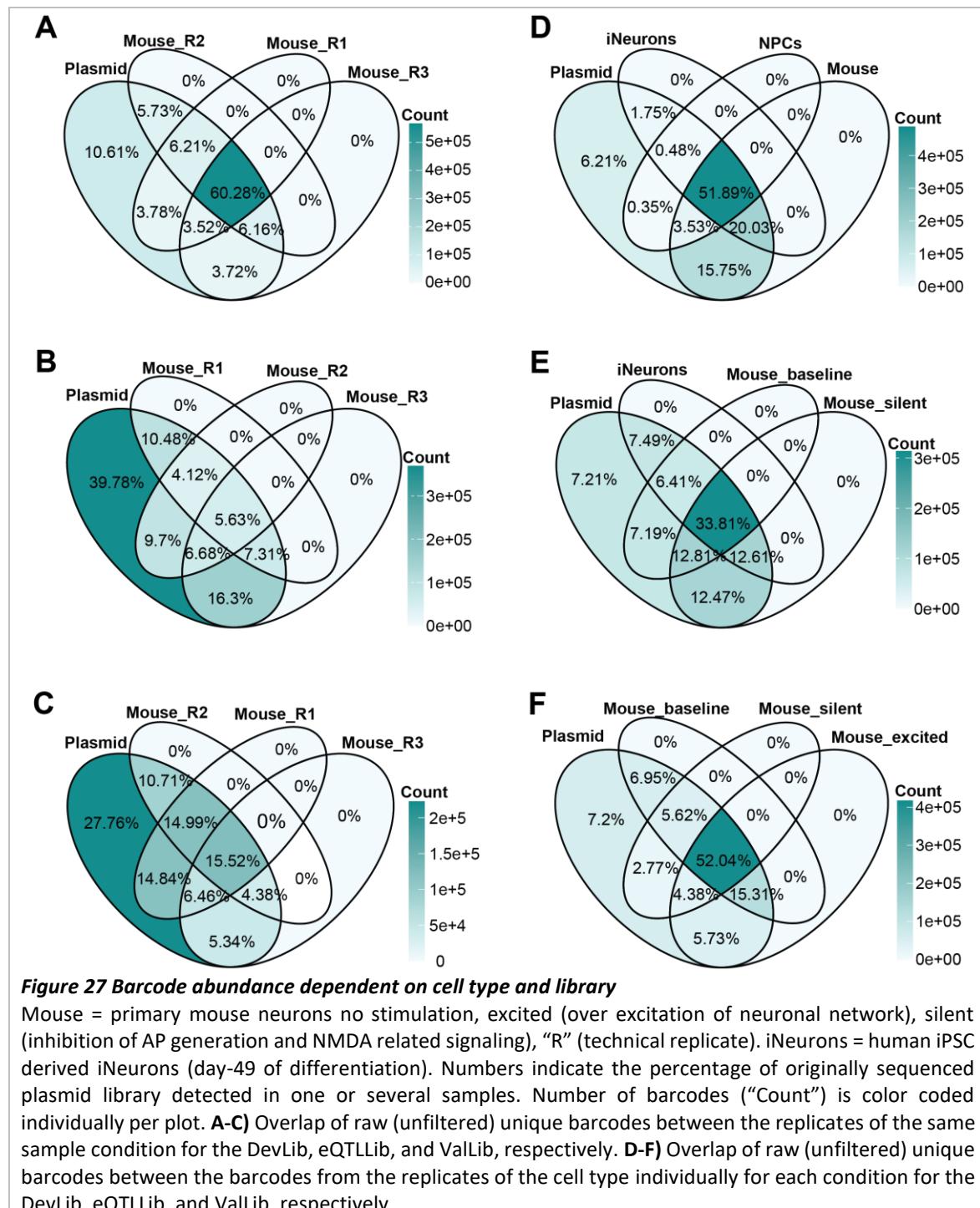
ValLib	1	Mouse_excited	F	1,287,650
ValLib	2	Mouse_excited	F	816,200
ValLib	3	Mouse_excited	F	1,837,857
ValLib	4	Mouse_excited	F	754,002
ValLib	1	Mouse_silent	F	1,044,412
ValLib	2	Mouse_silent	F	749,306
ValLib	3	Mouse_silent	F	917,527

**Table 28 Overview of batches, samples, and read depth for samples used in the MPRA**

Annotation of the different MPRA libraries, the cell culture batch, the stimulation condition and the read depth for each sample.

**Figure 26 Barcode recovery from technical replicates of the MPRA libraries**

**A)** Enhancer integrated into the gDNA drove enhancer specific expression of nLuc (dark blue) and the adjacent barcodes (reddish: pink to red). RNA was isolated from neural cultures and enriched for mRNA. Primers (light blue) specific for nLuc and the remaining of the plasmid backbone (green) amplified the barcode libraries for Illumina sequencing. **B)** Bioanalyzer High sensitivity track of the barcode PCR (length of amplicon: 259 bp). Lower (green, 35 bp) and upper (purple, 10380 bp) markers are used for calibration of each lane and indicate 35 and 10380 bp long dsDNA, respectively. Amount and size of DNA fragments were calculated according to an internal marker control (not shown). Residual primers are detected with 94 or 103 bp length. **C-E)** Number of unique barcodes (y-axis) recovered in from the investigated neural cell conditions (x-axis) for the DevLib, the eQTLlib, and the ValLib, respectively. hKCl (high potassium chloride depolarization), excited (over excitation of neuronal network), silent (inhibition of AP generation and NMDA related signaling), "R" (replicate).



After sequencing the barcode libraries, the set of raw recovered barcodes was compared between the original plasmid library and those libraries generated from cellular cDNA. Absolut barcode recovery from the initial plasmid library after step2 (Figure 24 A, before the nanoLuciferase insertion) was higher in all analyzed libraries compared to the number of barcodes recovered from the neural samples (Figure 26 C-E, Figure 27 A-F). For the DevLib, the overall barcode recovery in the iNeurons was lower than in the NPCs or the primary mouse neurons. Therefore, the number of investigated cells per iNeuron samples

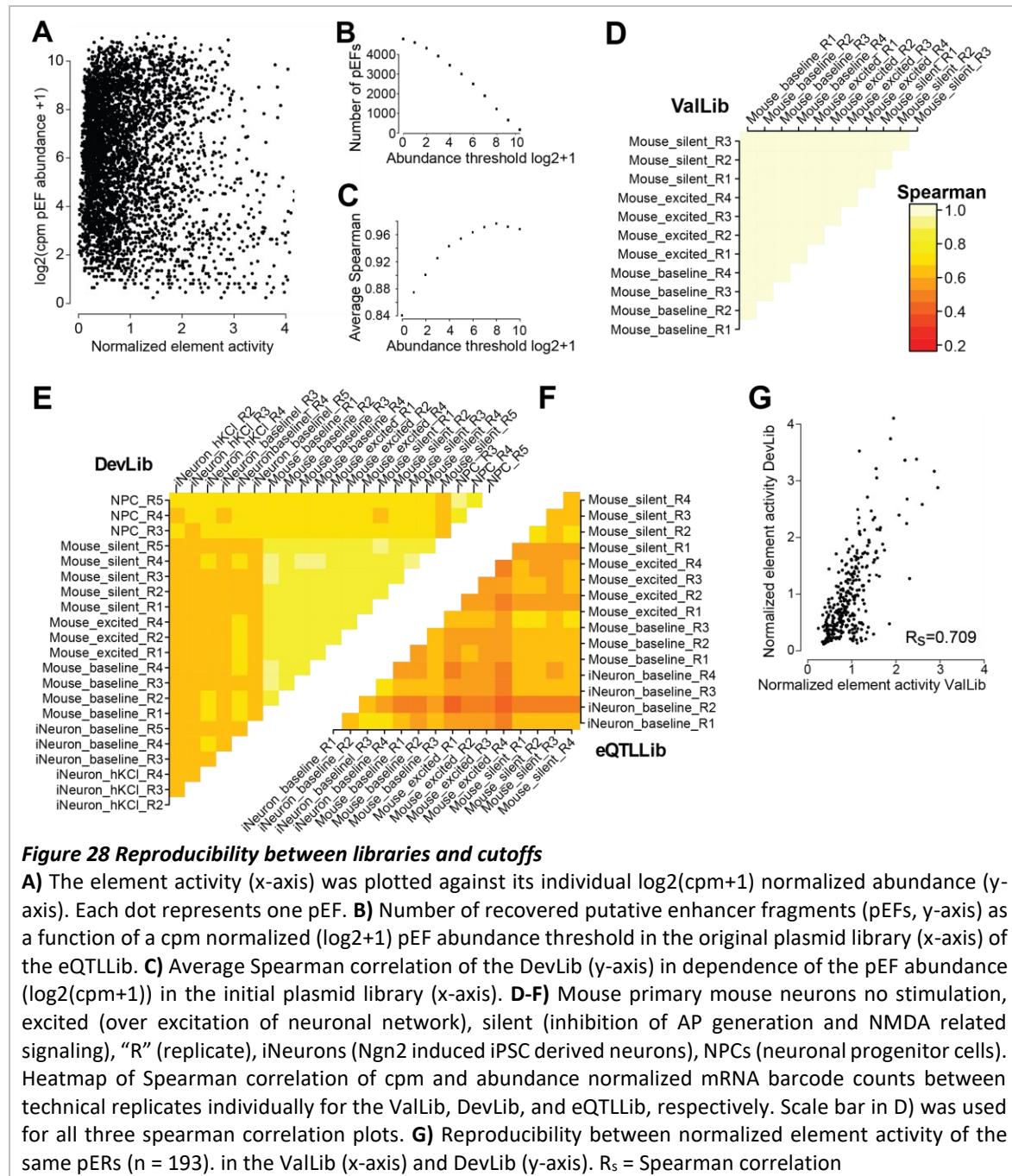
was increased for the eQTLLib. This resulted in similar numbers of recovered barcodes in iNeurons than detected in the mouse neurons for the eQTLLib.

Subsequently, the presence or absence of the barcodes in each individual cellular library was investigated. First, the overlap between the different replicates of unstimulated primary mouse cultures was investigated. The shared barcodes of the three replicates of the unstimulated mouse condition were depicted as a percentage of the overall recovered barcodes from the plasmid library (Figure 27 A-C). All three libraries showed a similar non-covered barcode percentage (6.21 %, 7.21 %, or 7.20 % for DevLib, eQTLLib, or ValLib) comparing three different cellular samples towards the plasmid library, respectively. The overlap between analyzed cellular conditions changed for each of the oligo libraries in the depicted graphs (Figure 27 D-F). It is important to keep in mind, though, that a recovery of less than 100 % of the barcodes per sample, does not necessarily result in a proportional loss of recovered pEFs nor pERs because an excess of barcodes (> 30 barcodes) was present for most of the pEFs (Figure 25 C-E). In order to analyze most of the pEFs and due to the fact that barcodes were recovered specific for each sample, barcodes were collapsed for the individual pEF for the subsequent statistical analyses. 60 - 95 % of the originally selected pERs were recovered and analyzed for condition and allele specific enhancer activity. In total 9,902 allele specific pEFs representing 4,951 pERs (88.4 % recovery of the oligo pool) were assessed for their ability to act as an enhancer.

#### **4.4.4 Good correlation of oligo pools after normalization of barcode abundance**

After the preselection of valid barcodes, the next objective was to quantify the ability of each putative enhancer to drive gene expression. Therefore, the abundance of barcodes in the mRNA/cDNA was normalized towards barcode abundance in the initial plasmid library (sequencing after step2, Figure 24 A, Figure 28 A). This approach highly increased the intra-condition and -cell type reproducibility. The normalized barcode abundance was an approximation for the activity of each enhancer to drive reporter gene (nLuc) and barcode expression (Figure 26 A).

There is a clear correlation between the pEF abundance in the library and the reproducibility of the assay. Reproducibility inside conditions increased with increasing abundance of each element in the barcode library. On the other hand, the number of

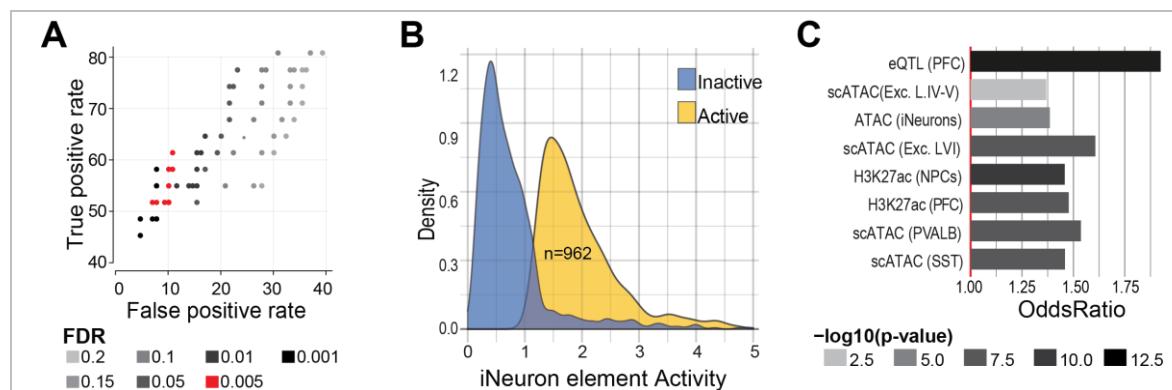


putative enhancer fragments under investigation decreased with an increased abundance threshold (Figure 28 B, C). Therefore, the pEFs were prefiltered for a minimal abundance of  $\log_2(\text{cpm} + 1) > 4$  (5 for eQTLlib). The increased complexity of the eQTLlib compared to the DevLib did not result in a higher reproducibility (measured by Spearman correlation). Overall, reproducibility between libraries of the same condition, independent of the conditions, was higher for mouse than for human cell types. Likewise, spearman correlation coefficient dropped from mouse neurons to human NPCs to human iNeurons for the DevLib (Figure 28 D-F). The correlation of overlapping enhancer between MPRA libraries (DevLib vs ValLib) was high in iNeurons (Spearman correlation  $R = 0.709$ ) despite the differences in

library complexity (Figure 28 G). The Spearman correlation between the joint analyzed enhancer fragments between DevLib and eQTLLib was 0.73.

#### 4.4.5 Classification of active and inactive enhancer regions

Next, the enhancers were investigated for their activity in the different cell types and stimulation conditions. This minimal activity threshold selected those enhancers which were active and, thus, could be affected by treatment condition or allele status. The activity threshold was set according to the baseline activity of negative control enhancer fragments. A generalized linear model was applied to determine a multiple-testing corrected p-value for the activity of each element. The minimal element activity cutoff for at least one allele was set at a p-value of less or equal to 0.005 (Figure 29 A). 152 control fragments were used in the DevLib to annotate active and inactive enhancer elements from primary mouse neurons at baseline condition. 962 or 481 were classified as active enhancer fragments (EFs) or regions (ERs), respectively (Figure 29 B). Over all libraries, 4,297 ERs were active in at least one condition. Active enhancer regions were significantly enriched in eQTL, open chromatin regions (H3K27ac ChIP, scATAC) from different cortical layers and cell types of the human postmortem brain, iNeurons (bulk ATAC), and NPCs (H3K27ac ChIP) (Figure 29 C).



**Figure 29 Activity threshold for iNeurons and mouse Neurons**

**A**) The true positive rate (TPR, y-axis) was plotted as a function of the false positive rate (FPR, x-axis) across varying thresholds of the FDR (false discovery rate, color coded) of differential allele activity. FDR = 0.005 (red) indicates chosen threshold. Graph depicts one condition (mouse baseline, DevLib) with the control fragments as ground truth. **B**) Distribution of normalized pEF activity (x-axis) in unstimulated iNeurons plotted for the DevLib. Light blue inactive and light-yellow active enhancer fragments based on comparison to the negative control fragments (DevLib). **C**) Bar plot displays in grey tones the significance (p-value, Fisher's exact test) over the threshold (odds ratio of 1, red line) for the enrichment of active ERs (x-axis) overlapping with genomic features such as open chromatin regions (y-axis): eQTL of adult postmortem prefrontal cortex (FDR  $\leq 0.05$ ), scATAC data from adult postmortem prefrontal cortex divided into several cortical layers (IV-V, VI), and into several inhibitory neuron types (SST – somatostatin positive, PVALB – parvalbumin positive), and H3K27ac chromatin immunoprecipitation data (ChIP) from neural progenitor cells (NPCs) and dorsolateral prefrontal cortex (PFC) and bulk ATAC-Seq from iNeurons (day-49).

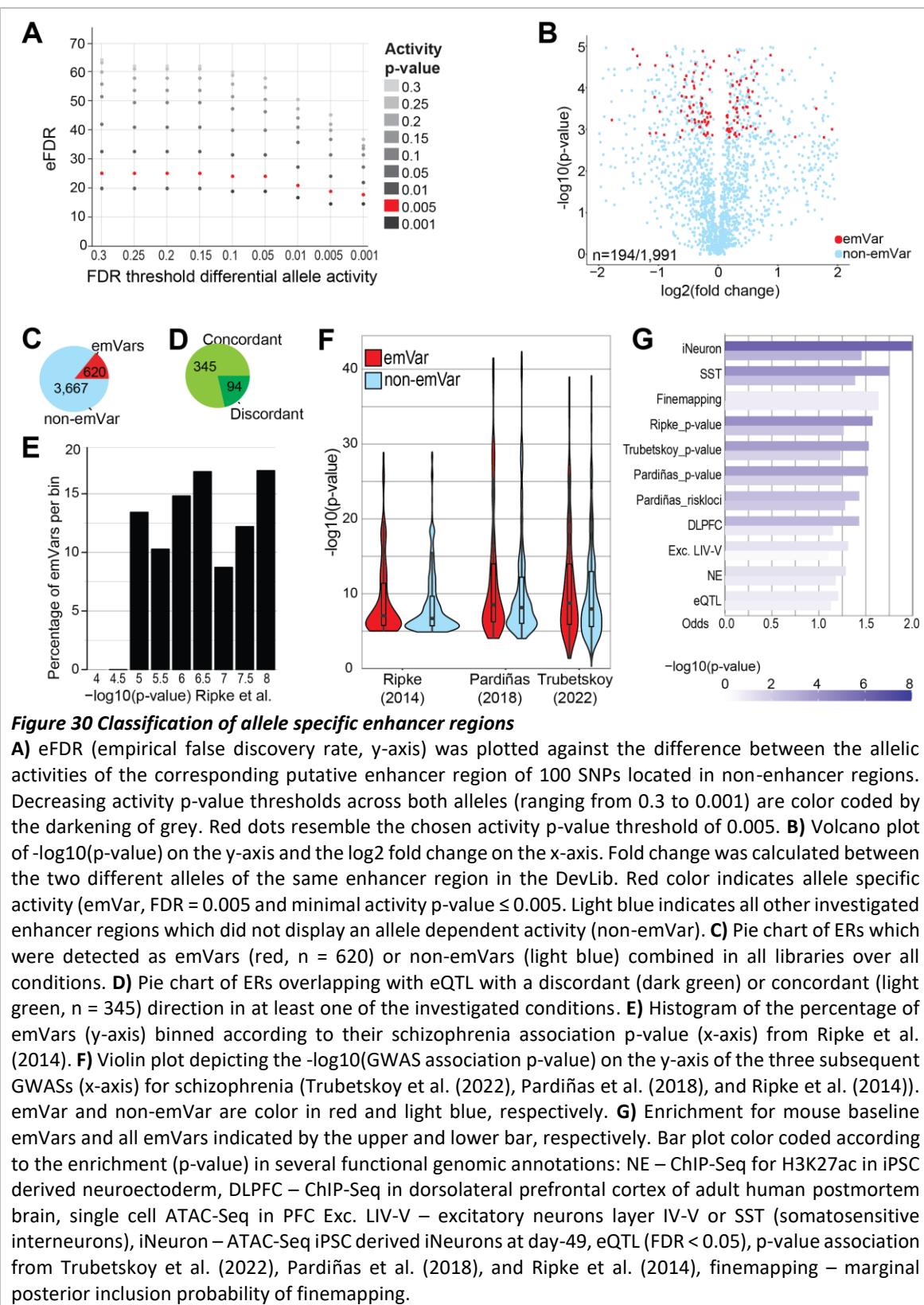
#### 4.4.6 MPRA identified 620 allele specific active enhancer regions

The second characterization of the active ERs examined whether their regulatory ability was altered by the presence of schizophrenia associated variants. Those ERs with a clear allele dependent impact on enhancer activity were classified as expression modulating variants (emVars) according to Tewhey et al (2016).

The criteria for the annotation as an emVar were dependent on specificity and sensitivity estimates based on the VallLib which contained randomly selected control (non-disease associated) SNPs (Figure 30 A). For this approach, it was assumed that the majority of the randomly selected SNPs, which were situated in non-enhancer locations of neural cell types, did not have any allelic effect on gene expression. The FDR to classify an active ER as an emVar was set to 0.005. This resulted in an overall true positive rate (TPR) of 55 % and a false positive rate (FPR) of 9 %. For all allele specific ERs, the fold-change (FC) between pEFs of the same pER was calculated separately for each condition (Ashuach et al., 2019) (Figure 30 B).

Using the above-mentioned parameters, 194 of 1,991 investigated active DevLib enhancer regions exhibited allele specific activity in baseline iNeurons. Those rather conservative activity thresholds excluded a substantial fraction of SNPs that showed evidence for allele specific activity without passing the activity threshold (675 non-enhancer emVars, Figure 30 B). In total 13 % (620 of 4,766) active emVars were detected across a combination of DevLib and eQTLLib over all tested conditions (Figure 30 C). This resembles 1.7 % of the originally considered SNPs ( $p\text{-value} \leq 10^{-5}$ ). Validation of a subset of hits ( $n = 44$ ) with the VallLib confirmed 65.9 % of the allele specific enhancers. 86 % of the 44 enhancer elements showed an allele specific effect when enhancer activity was disregarded. Moreover, over 75 % (345 ERs) of the identified emVars that overlapped with an eQTL of the (PFC) showed allelic concordance with at least one eQTL (Figure 30 D).

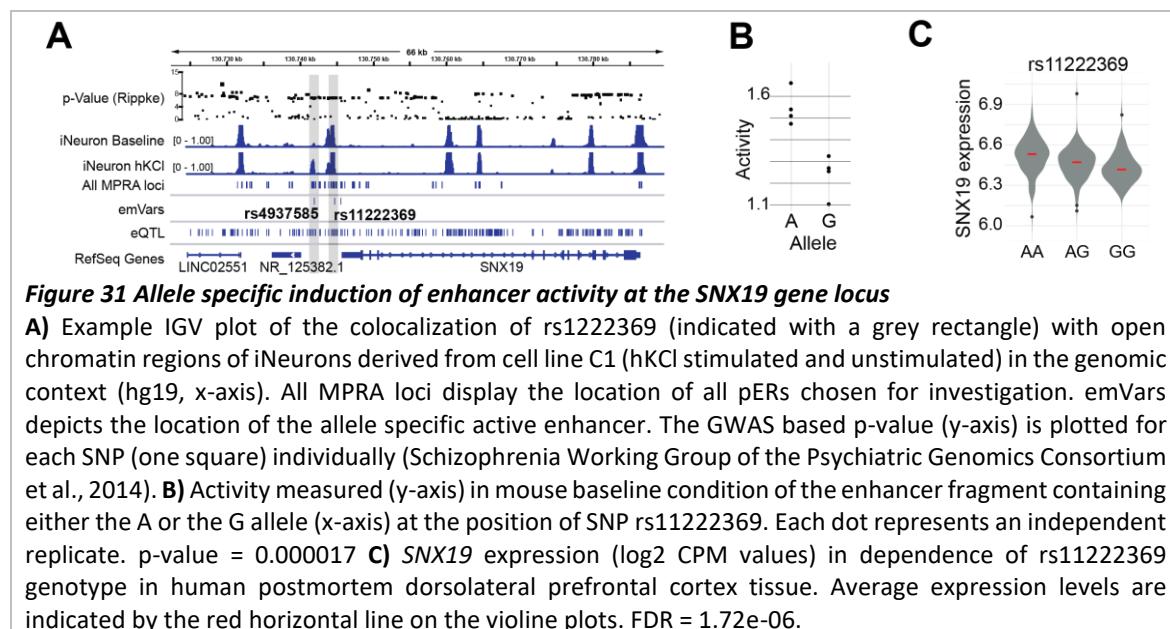
emVars were distributed across the entire considered  $p\text{-value}$  range ( $\leq 10^{-4.5}$ ) of schizophrenia association (Figure 30 E). Around 12 % of the pERs were emVars per indicated  $p\text{-value}$  association bin. The GWAS  $p\text{-value}$  of the SNPs annotated as emVars successively shifted towards lower levels with increased sample sizes in follow-up GWAS (Pardiñas et al., 2018; Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2014; Trubetskoy et al., 2022) (Figure 30 F).



Besides that, the enrichment of emVars compared to non-emVars was tested in several functional genomic annotations. Higher finemapping rates were not statistically significantly associated with emVar likelihood. In contrast, the association p-value from all three considered GWAS were predictive of emVar likelihood. Additionally, significant

enrichment in disease relevant accessible chromatin proofed to be highly predictive for emVar likelihood (Figure 30 G).

Allele specific enhancer regions were also identified nearby well-known schizophrenia associated genes such as *sorting nexin 19* (*SNX19*, minus strand) gene (Ma et al., 2020). The highest confident variant at this locus, rs11222369 (chr11:130,874,775-130,874,775), is located proximal to the 3'UTR of *SNX19* (Figure 31 A). The decrease in activity for the G (minor) versus the A (major) allele in mouse primary neurons at baseline (Figure 31 B) was in line with a decreased gene expression in postmortem PFC (Figure 31 C).

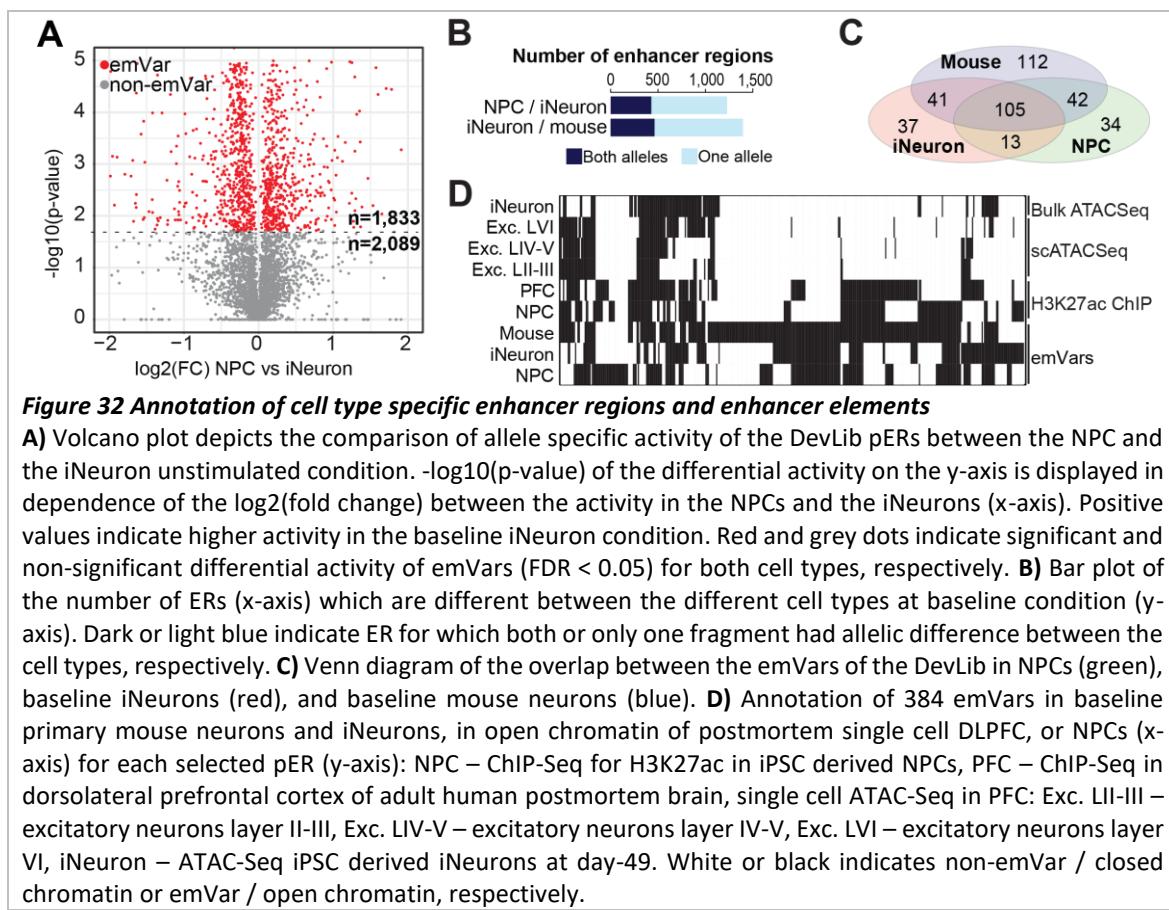


#### 4.4.7 Cell type and stimulus dependent active enhancer regions

Transcriptome and epigenome differ significantly between cell types and states. Hence, gene regulation operates highly cell type and state specific (Kim-Hellmuth et al., 2017; Nica et al., 2011). In line with this, eQTLs are annotated in a cell type or tissue specific manner. In addition to cell type specificity, gene regulation and gene expression are dynamic throughout the development. Therefore, pERs of the DevLib and the eQTLlib were analyzed in different neural cell types: primary mouse neurons, and human iPSC derived NPCs and iNeurons.

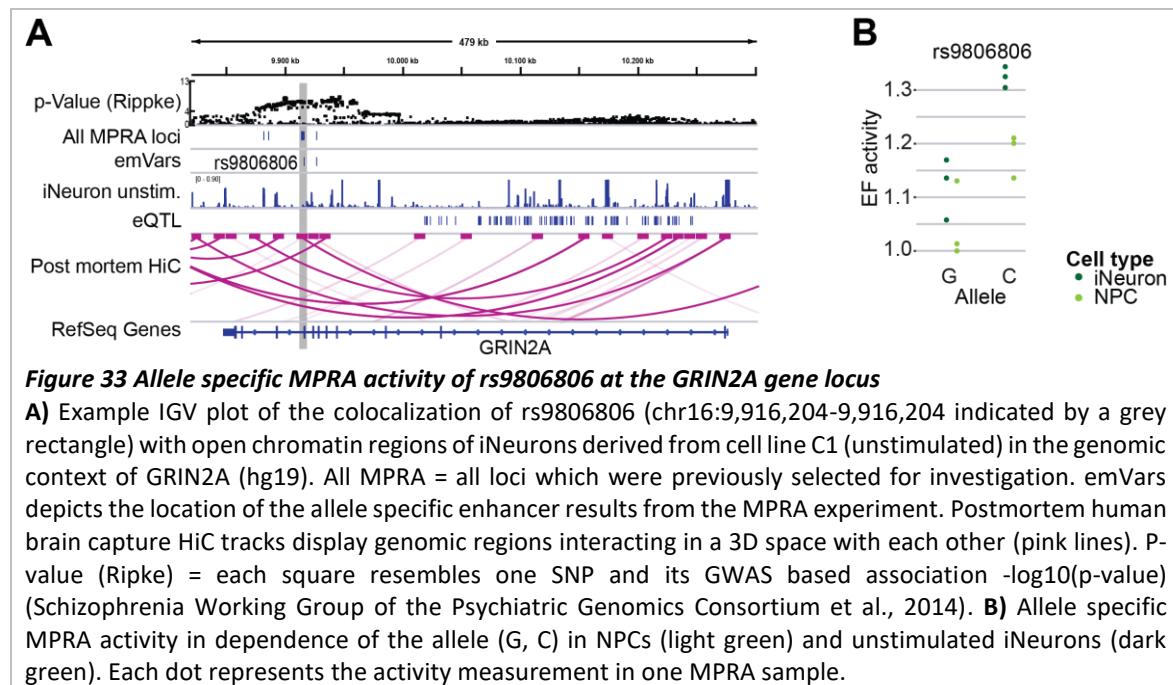
One of the hypotheses of schizophrenia is that it is a developmental disease, thus the decision to select the pERs of the DevLib due to their putative developmental specificity between neural progenitor cells (NPCs) and mature iNeurons. Almost half (47 %) of the

analyzed ERs were significantly differential active between the two neural states (Figure 32 A). Additionally, around one fourth of the ERs were differentially active for at least one of the alleles between two cell types (NPCs vs iNeurons, or iNeurons vs mouse neurons). Around one third of those allele specific EFs were differentially active for both alleles between two cell types (Figure 32 B). 48 % of all emVars in the DevLib were active in a cell type specific fashion (Figure 32 C) while 25 % of emVars were shared between only two of the three cell types. Active EFs were consistently detected by the MPRA when they were annotated as open chromatin in corresponding or similar cell types (Figure 32 D).



The cell type specific emVar rs9806806 is located at the *GRIN2A* gene locus on chromosome 16 (Figure 33 A). *GRIN2A* is a subunit of the glutamatergic binding N-methyl-d-aspartate (NMDA) receptor. rs9806806 is not annotated as an eQTL in postmortem brain studies (Fromer et al., 2016) even though it is highly associated ( $p\text{-value} = 1.847\text{e-}7$ ) with schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2014). In baseline iNeurons, the emVar capability of rs9806806 was more pronounced ( $\text{FDR} = 0.0002$ ) than in NPCs (Figure 33 B).

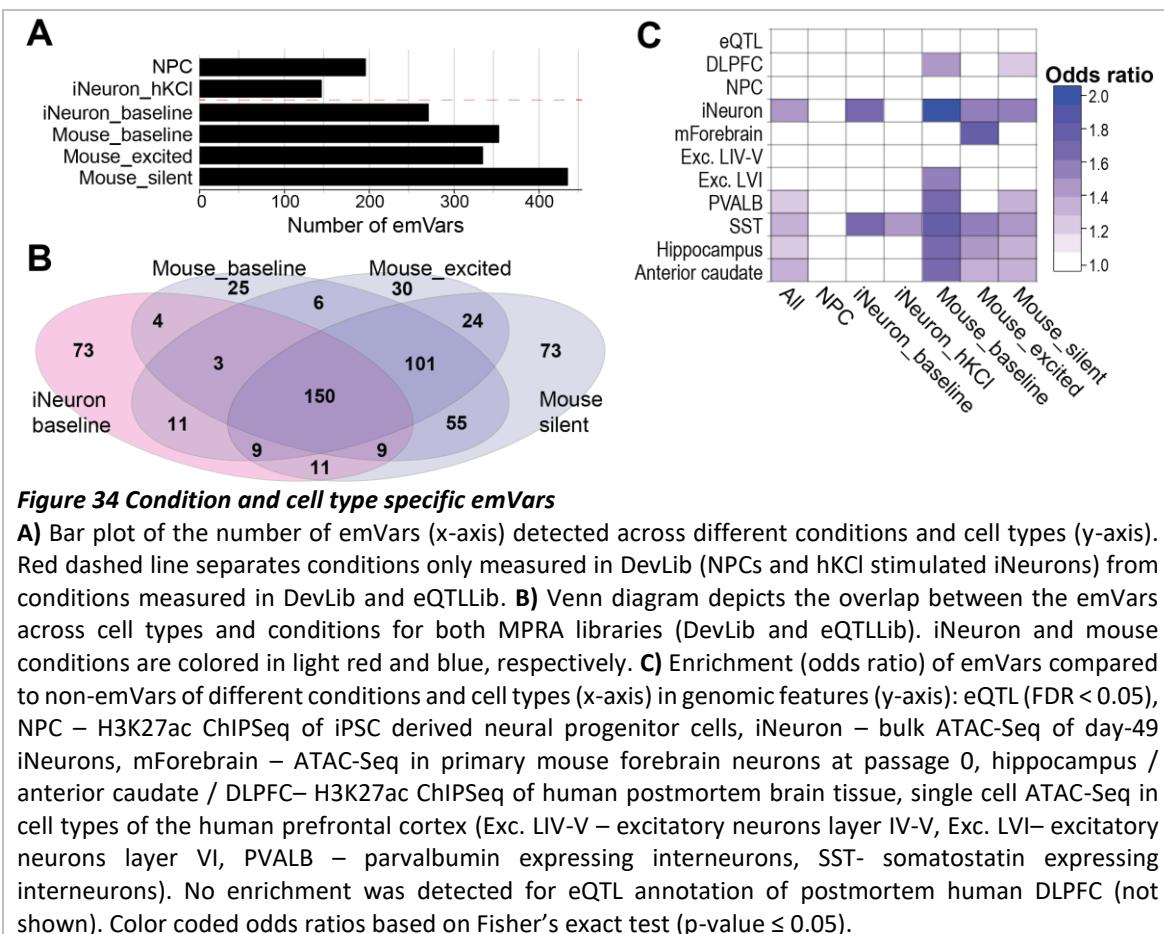
Besides the cell type specific activity of pERs, the cell state (stimulated or baseline) dependent activity of ERs and their corresponding EFs was analyzed. These experiments were based on the evidence that dramatic changes in the epigenome and transcriptome are described after neuronal activity which were recapitulated in this thesis. These changes in GRE activity and transcription are one mechanism to encode neuronal plasticity and react to neuronal inputs such as neurotransmission or hormonal stimulation. This cell altered state, after an initial stimulus, can heavily affect the response to future stimuli in, for instance, learning processes or concordant responses to environmental stimuli (Dong et al., 2020; Yap & Greenberg, 2018). Stimulation dependent variants (stimVars) were defined by a presence of an allelic effect to stimulation in respect to the silent condition.



In order to mimic different states of neuronal (network) activity, iNeurons were treated with their media containing a high potassium chloride solution (hKCl) to analyze calcium ion dependent change in EF activity. Moreover, the primary mouse neuronal cultures were treated with a combination of drugs to silence AP transmission or over - excite the neurons by preventing inhibitory neuronal transmission. The highest and lowest total number of emVars was detected in the silent condition for mouse neurons and the hKCl stimulated iNeurons, respectively (Figure 34 A). 25 % of all emVars (150 emVars) were shared between baseline iNeurons and primary mouse neurons (baseline, silent, excited) (Figure 34 B). Those emVars which were active in all neuronal conditions were annotated as pan-

neuronal emVars. In total, one fourth of all emVars were operating in a stimulation dependent fashion.

In line with their original selection criteria, the majority of the emVars were enriched in open chromatin regions of a similar cell type compared to in which cell type the MPRA activity was measured. In contrast, the emVars from the NPC condition were not enriched for open chromatin regions derived from NPCs. Nevertheless, emVars from certain conditions were also enriched for chromatin marks of other cell types, for instance, mouse baseline emVars in iNeuron open chromatin (Figure 34 C).

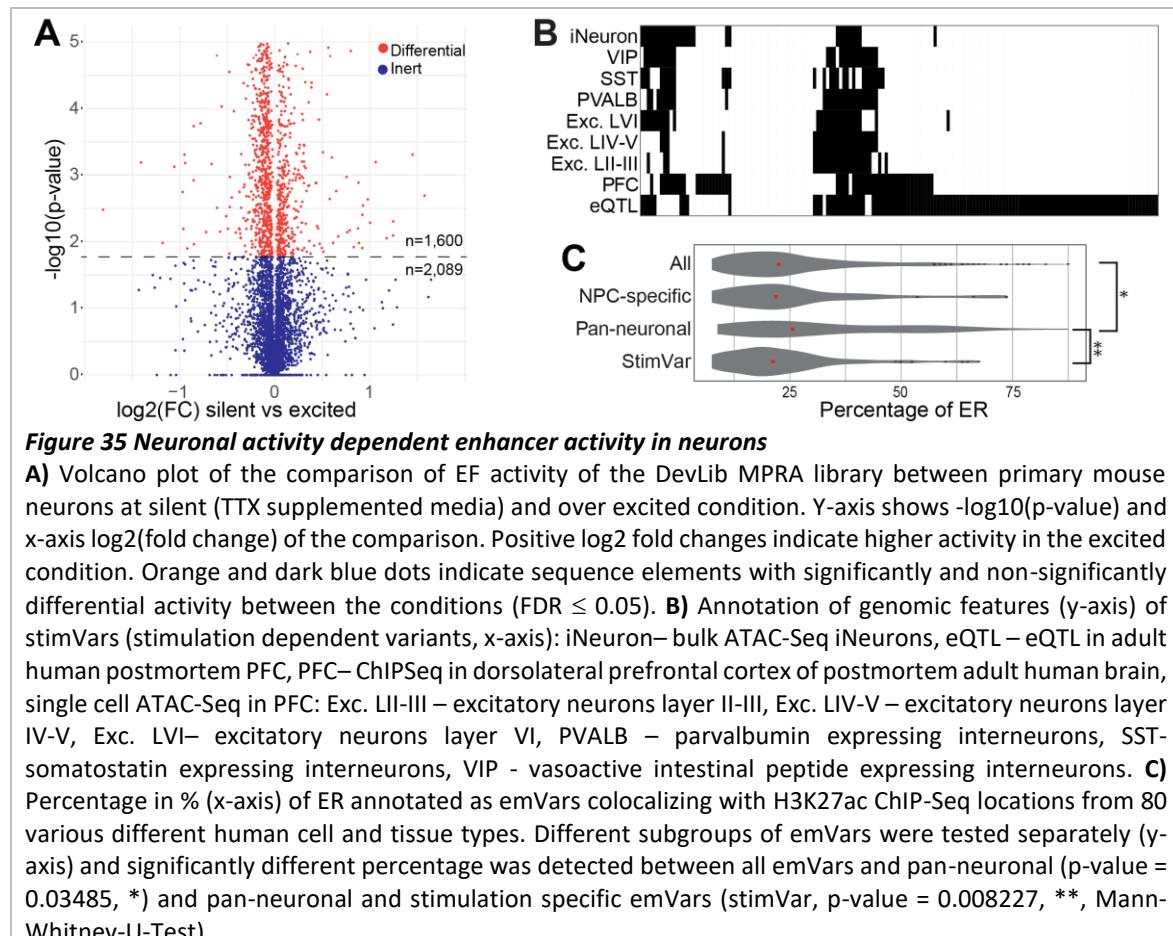


**Figure 34 Condition and cell type specific emVars**

**A**) Bar plot of the number of emVars (x-axis) detected across different conditions and cell types (y-axis). Red dashed line separates conditions only measured in DevLib (NPCs and hKCl stimulated iNeurons) from conditions measured in DevLib and eQTLlib. **B**) Venn diagram depicts the overlap between the emVars across cell types and conditions for both MPRA libraries (DevLib and eQTLlib). iNeuron and mouse conditions are colored in light red and blue, respectively. **C**) Enrichment (odds ratio) of emVars compared to non-emVars of different conditions and cell types (x-axis) in genomic features (y-axis): eQTL (FDR < 0.05), NPC – H3K27ac ChIPSeq of iPSC derived neural progenitor cells, iNeuron – bulk ATAC-Seq of day-49 iNeurons, mForebrain – ATAC-Seq in primary mouse forebrain neurons at passage 0, hippocampus / anterior caudate / DLPFC – H3K27ac ChIPSeq of human postmortem brain tissue, single cell ATAC-Seq in cell types of the human prefrontal cortex (Exc. LIV-V – excitatory neurons layer IV-V, Exc. LVI – excitatory neurons layer VI, PVALB – parvalbumin expressing interneurons, SST – somatostatin expressing interneurons). No enrichment was detected for eQTL annotation of postmortem human DLPFC (not shown). Color coded odds ratios based on Fisher's exact test (p-value  $\leq 0.05$ ).

Next, stimulation dependent ERs (stimERs) were annotated for the hKCl or excited compared to baseline condition for the iNeurons or the mouse neurons, respectively. Similar to the difference in ER activity between cell types, 43 % of the ERs in the DevLib were differentially active in the same cell type (primary mouse neurons) comparing the two most different conditions: silent and excited (Figure 35 A). In total, approximately one fourth of all ERs (4,959) were stimulation dependent ERs. StimERs of the DevLib were located in open chromatin regions of neuronal tissue or neural cell types. Most stimVars were colocalizing with annotated eQTL (Figure 35 B). emVars co-localized with H3K27ac

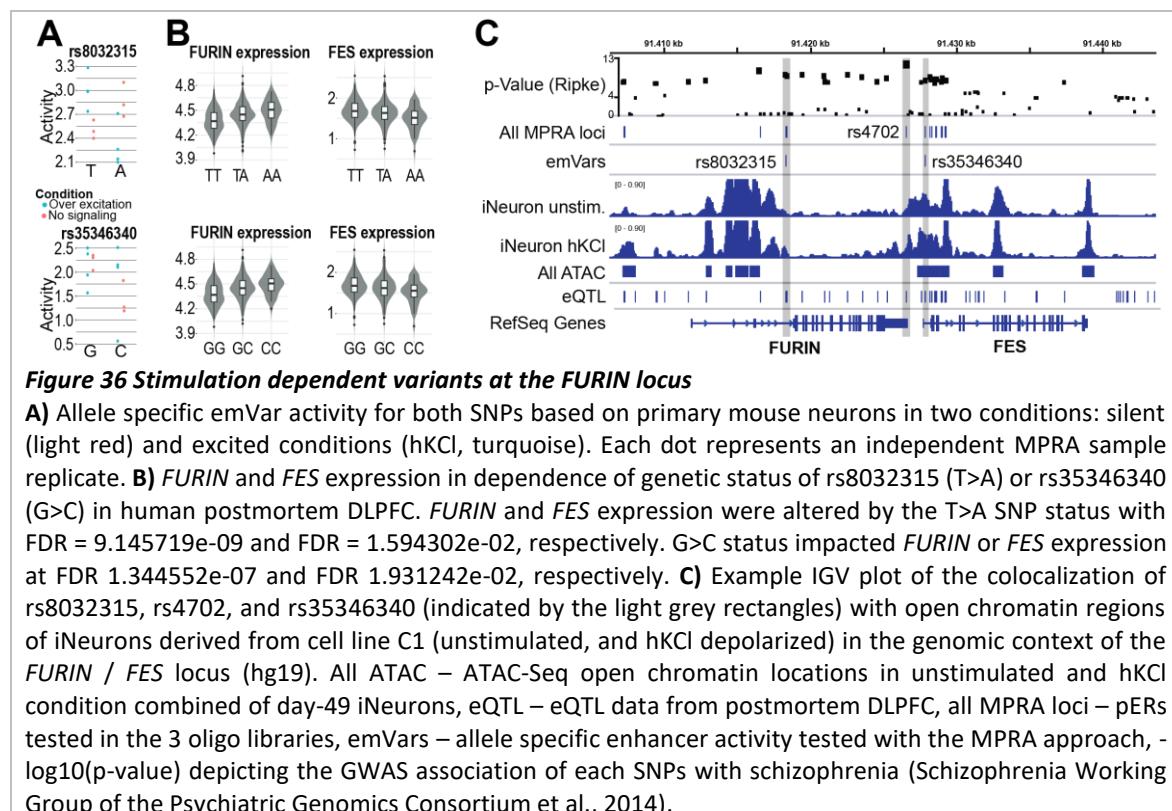
regions from over 80 neuronal and non-neuronal tissues was investigated (Figure 35 C). The pan-neuronal emVars were significantly more located in H3K27ac peaks that exhibited a more ubiquitous activity pattern across the 80 tested cell types compared to more specialized emVars such as the stimVars.



StimERs in respect to non-stimulation responsive active ERs co-localized with several stimulus dependent trans-acting transcription factors including *ATF*, *AP-1*, *ETV*, *AP2A*, *AP2C*, *ATF2*, *CREB3*, *HES1*, and *PAX8*. Similarly, stimulation dependent emVars (stimVars) overlapped with binding sites of stimulus dependent trans-acting transcription factors such as *EGR1*, *NR4A1/2*, and *REST* (Yap & Greenberg, 2018).

The set of stimVars also contained two schizophrenia SNPs at the well-known schizophrenia associated locus of the two genes *FURIN* and *FES* (Schrode et al., 2019). rs8032315 (upper panel Figure 36 A) clearly regulated MPRA activity in dependence of the stimuli applied to the primary mouse culture into opposing directions. Even though, rs35346340 (lower panel Figure 36 A) showed a similar stimulation dependent ER activity change, the alterations in MPRA activity were different to those detected from rs8032315. Rs35346340 only exhibited lower activity in the silent condition for the minor allele (C<G) consistent with the

eQTL effect on *FES* expression (lower panel right Figure 36 C). Both emVars are annotated as eQTLs (Figure 36 B) with opposing effects on gene expression of the implicated genes. Both major alleles increase *FURIN* gene expression while decreasing *FES* gene expression (Figure 36 C) in human postmortem DLPFC. The directionality of the eQTL effect for *FURIN* and *FES* expression corresponded with the activity pattern of rs8032315 upon silent and excited treatment in primary mouse neurons, respectively. Interestingly, rs35346340 was located in close proximity (1.3 kb upstream) to the previously identified eQTL (rs4702) of the *FURIN* gene. For the eQTL at rs4702, a strong effect was described on cell physiology in another iPSC derived neuronal cell model (Schrode et al., 2019). Consistent with this described eQTL effect, the MPRA assay revealed allele specificity (FDR = 0.000259) for rs4702. However, it was classified as a non-active emVar in line with the location of rs4702 outside annotated open chromatin from iNeurons and postmortem brain (Figure 36 B for open chromatin in iNeurons) while the neighboring rs35346340 was located inside the open chromatin.



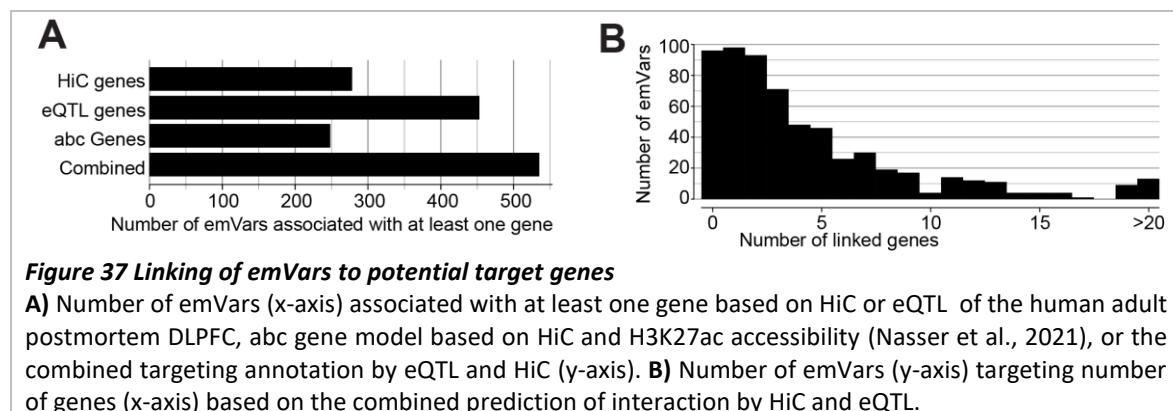
## 4.5 Identification of target genes from emVar regions

Functional annotation of GWAS results is inevitable to pinpoint causal variants for disease development and progression. However, a database of functional SNPs from e.g., MPRA

experiments, is only the starting point to identify the context specific targets of those functional variants. This knowledge will eventually help to understand and rescue schizophrenia endophenotypes.

#### 4.5.1 Linkage of 535 emVars to target genes

In order to understand the functional impact of emVars on gene expression, it is important to pinpoint their (context specific) target genes. Therefore, eQTL and chromatin conformation capture (HiC) data sets were used to link emVars with putative target genes as published previously for other complex diseases affecting e.g., the cardiac system (Rajarajan et al., 2018; X. Wang et al., 2016).



**Figure 37 Linking of emVars to potential target genes**

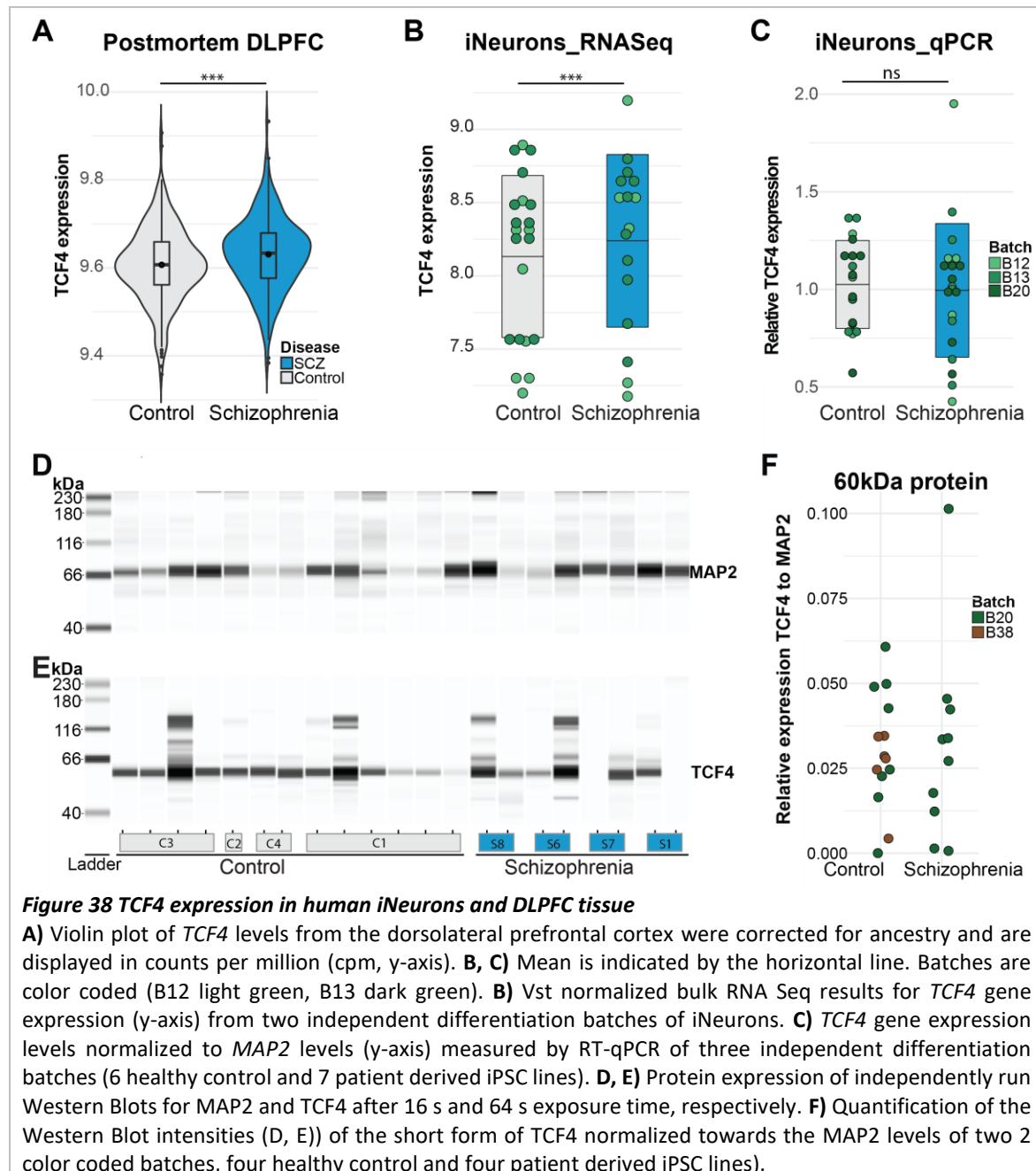
**A)** Number of emVars (x-axis) associated with at least one gene based on HiC or eQTL of the human adult postmortem DLPFC, abc gene model based on HiC and H3K27ac accessibility (Nasser et al., 2021), or the combined targeting annotation by eQTL and HiC (y-axis). **B)** Number of emVars (y-axis) targeting number of genes (x-axis) based on the combined prediction of interaction by HiC and eQTL.

278 or 453 emVars were linked to at least one putative target gene utilizing a HiC data set or eQTL (Fromer et al., 2016) from postmortem DLPFC. 248 emVars were linked based on the abc gene model that links GRE (H3K27ac ChIP) to genes based on HiC data (Nasser et al., 2021). The combination of the eQTL and HiC approach linked 535 genes (82 more than eQTL alone) to at least one target gene. Nevertheless, a substantial fraction (85) of emVars could not be linked to any target gene (Figure 37 B), while the majority of emVars were associated to several genes: around 44 % of emVars were linked to more than one gene (emVar linked gene) while around 40 % of those emVar linked genes were linked to more than one emVar (Figure 37 B).

#### 4.5.2 Decreased enhancer activity decreased RNA expression at the TCF4 locus

Linking of emVars to target genes can be improved by more granular data sets than HiC and by using methods which do not entirely rely on correlation assumptions like eQTLs. Therefore, a loci specific low-throughput approach was utilized to investigate the target

genes of emVars (Genga et al., 2019): The CRISPRi system enables to investigate enhancer activity in its natural genomic context. A krüppel associated box (KRAB) domain was linked to an inactivated Cas9 (dCas9) which is targeted by guide RNAs (gRNAs) to specific gDNA sequences (Fulco et al., 2016). When the KRAB domain is in close proximity to the gDNA, it co-binds other repressor proteins which results in the formation of heterochromatin (Xu & Qi, 2019).



In this thesis, as a proof of concept for future pooled screen approaches as described in Gasperini et al. (2019), the dCas9 fusion protein was targeted to the schizophrenia associated *TCF4* gene (Gene 6925, HGNC:11634). *TCF4* is referred to as *SEF2* (*SL3-3 enhancer factor 2*), *ITF2* (*immunoglobulin transcription factor 2*), *bHLHB19*, *E2-2*, *PTHS*,

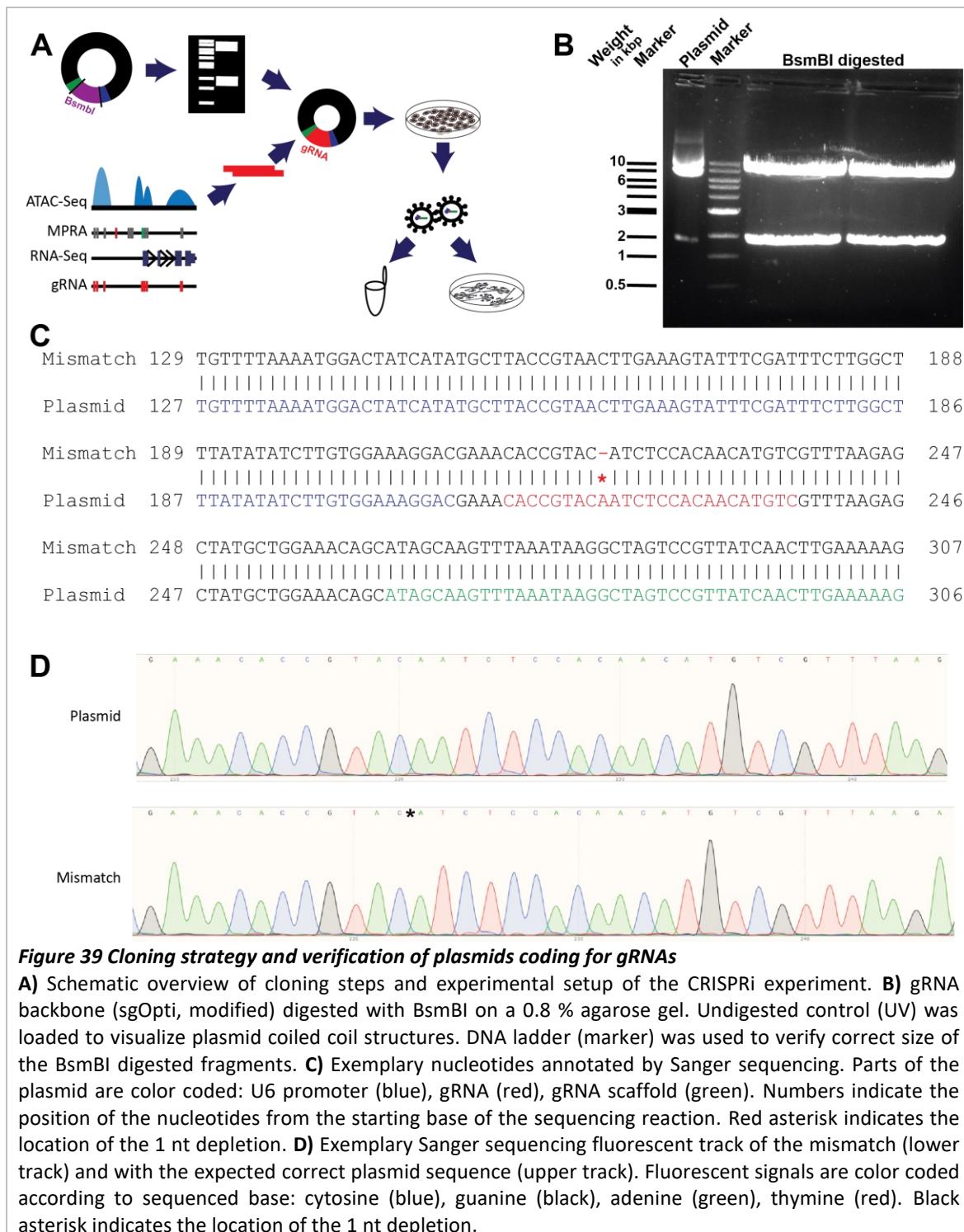
*FECD3*, or *CDG2T* and is often mistaken for the T-Cell factor 4 (*TCF7L2*, Gene 6934, chromosome 10q25.3). Several SNPs in the *TCF4* locus are associated with the age of onset of schizophrenia and the severity of the negative symptoms (Wirgenes et al., 2012). Additionally, there is evidence that rs1272363, located in the shared 3'UTR of *TCF4*, reduces the activity of the GRE (Mohamed et al., 2019). From the schizophrenia associated 181 SNPs (p-value < 10<sup>-5</sup>), the original MPRA pool contained 72 SNP loci and 51 SNPs were analyzed (Pardiñas et al., 2018; Ripke et al., 2011). 10 % (5 of 51) showed differential activity between the two tested alleles in at least one of the cell type or state conditions. The presence of the alternative allele for two SNPs, rs78281202 and rs55776938, increased the enhancer activity in the analyzed conditions in the MPRA assay.

Expression levels of *TCF4* in the dorsolateral prefrontal cortex (DLPFC) (Fromer et al., 2016) were normally distributed (Shapiro Wilkinson test,  $W = 0.9908$  and  $0.9936$ ) in the healthy control ( $n = 279$ ) and patients suffering from schizophrenia ( $n = 258$ ) group, respectively. The variance between the two groups is equal ( $F = 1040$ ,  $p = 0.7475$ ) and *TCF4* expression between the two groups is significantly different (two tailed unpaired t-test:  $t = 3.27$ ,  $p = 0.0009$ ). Median expression of *TCF4* in the DLPFC was 9.607 and 9.634 for healthy controls and patients suffering from schizophrenia, respectively (Figure 38 A).

Next, *TCF4* expression was inspected in the iNeuron expression data set to verify that iNeurons can be utilized to investigate disease and model-based alterations in *TCF4* expression. *TCF4* expression was significantly increased 1.479-fold ( $p.\text{adjusted} = 0.00086$ ) in bulk RNA-Seq of iNeurons derived from patients suffering from schizophrenia (five donors) compared to healthy control (six donors) derived iNeurons (Figure 38 B). RNA-Seq results were validated with quantitative real time polymerase chain reaction (RT-qPCR) in three independent neuronal differentiation batches with six healthy control and seven patient derived cell lines with 19 and 20 independently processed samples, respectively. *TCF4* expression was normalized towards *MAP2* expression individually for each sample. In contrast to the postmortem and bulk RNA-Seq results, no significant increase in *TCF4* expression was detected in iNeurons derived from patients suffering from schizophrenia compared to healthy control ( $p = 0.6267$ ,  $U = 172$ , two-tailed Mann - Whitney test, Figure 38 C).

Western Blot confirmed the translation of the mRNA into protein and expression of the protein in iNeurons derived from iPSCs from patient and control donors. The most

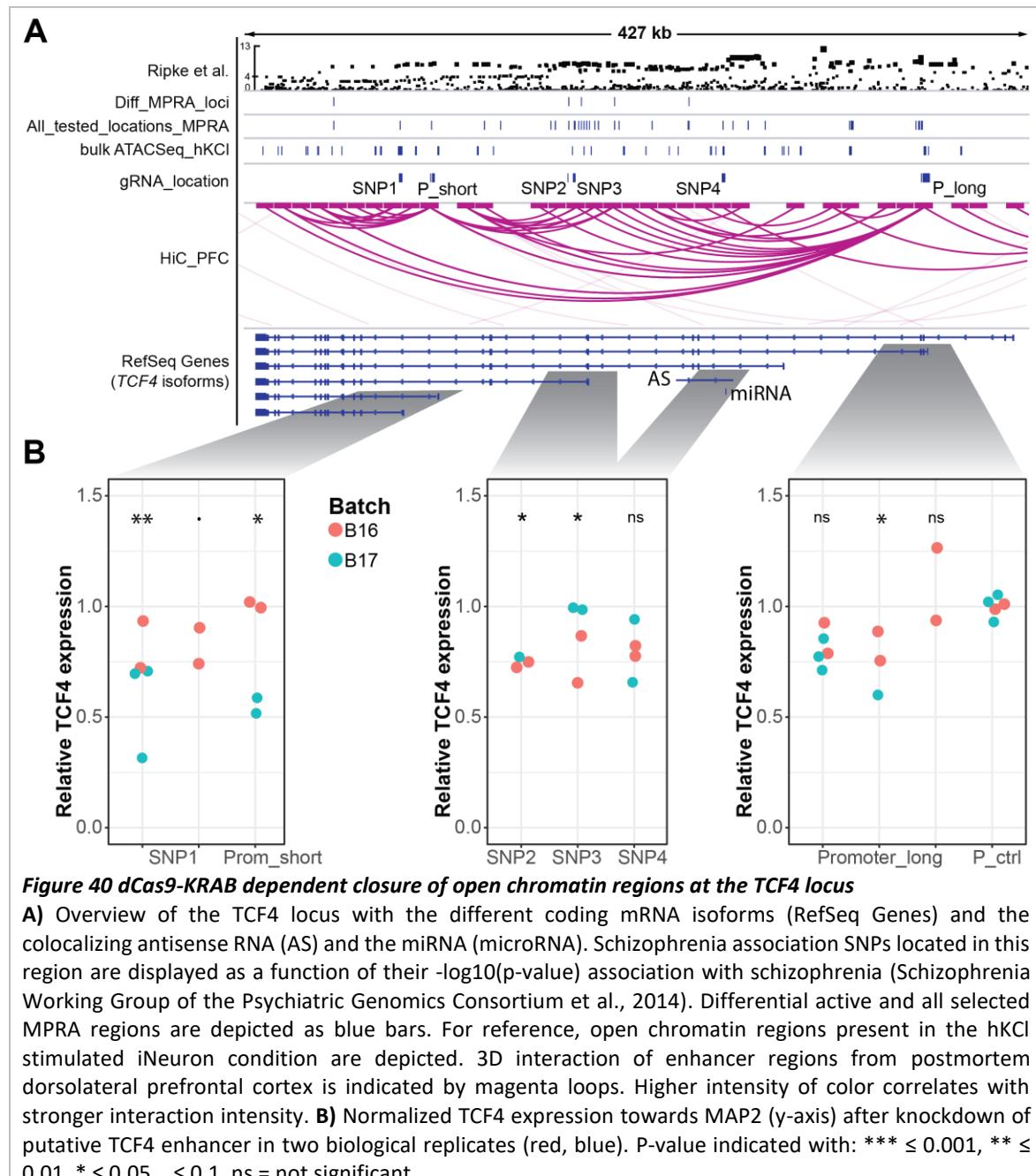
abundant fragment at 60 kDa has a similar weight as the short protein isoforms, for instance TCF4\_A or TCF4\_H (for further information on the isoforms of TCF4 see Sepp et al., 2017). MAP2 protein abundance was used to normalize TCF4 protein expression per sample and appeared at an apparent molecular weight of 70 kDa (Figure 38 D, E). There is no significant difference in the normalized TCF4 protein expression of the isoform detected at 60 kDa between neurons derived from healthy control and schizophrenia patients ( $p = 0.5224$ ,  $U = 54$ , Figure 38 F).



**Figure 39 Cloning strategy and verification of plasmids coding for gRNAs**

**A)** Schematic overview of cloning steps and experimental setup of the CRISPRi experiment. **B)** gRNA backbone (sgOpti, modified) digested with BsmBI on a 0.8 % agarose gel. Undigested control (UV) was loaded to visualize plasmid coiled coil structures. DNA ladder (marker) was used to verify correct size of the BsmBI digested fragments. **C)** Exemplary nucleotides annotated by Sanger sequencing. Parts of the plasmid are color coded: U6 promoter (blue), gRNA (red), gRNA scaffold (green). Numbers indicate the position of the nucleotides from the starting base of the sequencing reaction. Red asterisk indicates the location of the 1 nt depletion. **D)** Exemplary Sanger sequencing fluorescent track of the mismatch (lower track) and with the expected correct plasmid sequence (upper track). Fluorescent signals are color coded according to sequenced base: cytosine (blue), guanine (black), adenine (green), thymine (red). Black asterisk indicates the location of the 1 nt depletion.

After verification of TCF4 expression at gene and protein level, six target loci (Figure 40 A) for the CRISPRi experiment in the *TCF4* gene locus (chr18:52,877,000 – 53,360,000) were chosen from open chromatin regions (bulk ATAC-Seq of iNeurons) and regions with MPRA activity (Figure 40 B). TCF4 is located on the reverse strand of chromosome 18q21.2 and is encoded by 437 kb containing a total of 41 exons (Sepp et al., 2011).



The CRISPRi system's specificity is dependent on sequence specific guide RNAs (gRNAs). Hence, gRNAs, specific for the six target loci, were individually cloned into a lentiviral expression vector: modified sgOpti. The modified sgOpti plasmid contained an ampicillin resistance for bacterial selection, a blasticidin S resistance for eucaryotic selection, two

BsmBI restriction sites, and the U6 promoter driving the gRNA expression with its sgRNA scaffold. The modified sgOpti plasmid was 8,955 bp in length (Figure 39 A). Upon BsmBI restriction digest it was cleaved into two fragments of 1,885 bp and 7,070 bp. The desired backbone (7,070 bp) was isolated from the agarose gel via a column purification step (QIAGEN). The annealed DNA oligos coding for the gRNA were ligated into the vector resulting in a 7,095 bp containing DNA plasmid (Figure 39 A, B). Sanger sequencing confirmed the presence of the correct gRNA sequence individually for each cloned gRNA.

The promoter regulating the long *TCF4* form was targeted by several overlapping gRNA pools (Figure 40 A, prom\_long pools). One gRNA was present in two pools targeting the promoter\_long region (pool b and c, see Table 7 for the number of gRNAs in each pool and the targeted region). In general, HiC interaction maps of the postmortem DLPFC indicated close vicinity interactions between regions within the *TCF4* locus but also with the neighboring genomic DNA (Figure 40 A). The locus annotated as SNP4 is situated between the region coding for an antisense RNA (AS-RNA), a miRNA (MIR4529), and a short AS-RNA on the 5' site of this locus. The locus of SNP4 targeted rs72930740 which is associated with schizophrenia but was not differentially active in the MPRA enhancer screen. The region of SNP1 contained an emVar classified as a non-active enhancer. SNP2 was annotated as an emVar overlapping with a region with open chromatin in iNeurons and human DLPFC.

Two differentiation batches were independently infected with the gRNAs. Most of the targeted locations of each batch were performed in duplicates of the same infected gRNA pool. *TCF4* expression levels for individual samples (technical quadruplicates) were normalized towards their corresponding *MAP2* expression levels (Figure 40 C). To compare *TCF4* levels between batches, *MAP2* normalized *TCF4* expression levels were normalized to the mean of the *TCF4* expression levels from the Prom\_long\_d locus loci. This Prom\_long\_d pool showed the highest expression of *TCF4*. Additional control viruses in this experiment were not considered as a normalization control (data not shown) because they 1) contained puromycin instead of blasticidin S resistance which resulted in a higher density of the neuronal culture with increased *MAP2* levels, 2) contained a gRNA against the *OCT4* gene which might have resulted in an altered differentiation process, and 3) contained gRNAs against the CMV promoter which most likely affected several TSS of several genes. The effect of the gRNAs on the *TCF4* expression was highly depending on the batch of iNeuron differentiation (Figure 40 C) and ranged between a decrease of more than 75 % and an

increase to 125 %. Normalized fold change TCF4 gene expression levels were normally distributed (Shapiro-Wilk test  $W=0.8715$ ,  $p$ -value = 0.2724). Five out of nine tested gRNA pools significantly downregulated TCF4 expression. However, only one out of the two gRNA pools targeting the SNP1 location significantly impacted TCF4 expression ( $p$  – value = 0.00227 and 0.08514), similar to the gRNA pools targeting the long promoter (1 ( $p$  – value = 0.01923) out of 3 ( $p$  – value = 0.05892 and 0.74521). Decreasing the enhancer activity at SNP2 also significantly decreased TCF4 expression levels ( $p$  – value = 0.02016). There were 26 degrees of freedom and a residual standard error of 0.152 with a multiple R-squared of 0.4763.

In conclusion, I showed that iPSCs were successfully reprogrammed from primary human cells and differentiated into excitatory iNeurons. These iNeurons depicted electrical properties and responses to depolarization similar to primary (mouse) neurons. emVars annotated by MPRA were cell and condition specific and could be linked to potential target genes by eQTL, HiC, and one CRISPRi experiment. Hence, the setup to annotate the functional effect of functional disease associated variants on gene expression was successful and can now be implemented for other complex (psychiatric) diseases.

## 5 Discussion

In summary, the results presented in this thesis underline the high importance and relevance of induced pluripotent stem cells as a model organism to investigate schizophrenia. Here, the usability of published protocols to successfully reprogram primary human cells from fibroblasts to iPSCs is shown. Additionally, the Ngn2 induced generation of neurons from iPSCs was performed to enable the analysis of the genetic context of schizophrenia in disease specific cell types. In order to pinpoint the genetic variants responsible for altered gene regulation, the MPRA workflow was implemented. This approach offered the unique possibility to investigate selected disease associated variants in several disease relevant cell types. The reported set of functional variants and their context specific activity in disease relevant cell types serves as a unique resource for future mechanistic studies. One of those important follow up studies is the here reported linking of functional variant loci to gene(s). As described in this thesis, this can be performed by, for instance, chromatin conformation capture, eQTL, and CRISPRi. The combination of those high-throughput methods directly enables focused studies to summarize the variant effects in a condition and cell type specific manner on genes relevant to schizophrenia. Hence, this knowledge provides a valuable starting point for future in depth studies on the molecular, cellular, and circuit level role of schizophrenia associated non-coding genetic variants.

### 5.1 Successful generation of karyotypical normal iPSC lines

Primary human fibroblasts (and PBMCs) were reprogrammed into induced pluripotent stem cells (iPSCs) to investigate schizophrenia in a human cellular model system. Fibroblasts from five individuals suffering from schizophrenia were reprogrammed successfully into iPSCs (Figure 2) according to immunocytochemistry (Figure 6) for the pluripotency markers: OCT4 and NANOG (Takahashi & Yamanaka, 2006). Morphological hallmarks e.g., prominent nucleoli further affirmed the successful generation of iPSCs (Figure 4, Figure 5). Furthermore, the iPSCs were capable to form embryoid bodies (Figure 6) (Sheridan et al., 2012).

As a next quality control of the iPSCs, the presence of CNVs in the reprogrammed pools and single colonies was investigated (Table 21). Most of the detected reprogramming induced

deletions were smaller than 1 Mbp but larger than 200 kbp, in line with published CNV sizes (maximum 1.2 Mbp) for a similar non-integrating reprogramming approach of fibroblasts (X. Kang et al., 2015). For the here reprogrammed fibroblasts, 9 out of 14 cell lines presented CNVs. The length of the detected CNVs was below the detection threshold of other techniques such as Giemsa banding (detection threshold 6 Mbp), showcasing the value of the here applied microarrays.

In general, the efficiency of the reprogramming procedure compared to similar protocols is up to 40-fold lower than the reported efficiency of 0.011 % (alkaline phosphatase staining (Drozd et al., 2015)) or 0.013 % (estimates from bright field images (Schlaeger et al., 2015)). One reason for the lower efficiency might have been the source of the fibroblasts used for the reprogramming experiments. Previously reported reprogramming experiments used fetal or freshly isolated fibroblasts from primary human tissue. Therefore, fibroblasts were used at passage 5 - 6 for reprogramming (Rasmussen et al., 2014). The fibroblasts used in this thesis for the reprogramming, though, arrived from the vendor (Coriell) already at passage number 4, 5, or even 8 and had to be further expanded to passage 8 - 10 to generate a sufficient number of cells for nucleofection. Despite the fact that this massive expansion was necessary, it harbored the complication that primary fibroblasts from skin tissue have a limited proliferative capacity. Between 80 - 90 % of the fibroblast in the culture are senescent after 10 to 12 passages inhibiting the reprogramming process (Goldstein, 1990; Huschtscha et al., 2012). Nevertheless, the here described adaption of the reprogramming process (Diecke et al., 2015) was able to robustly induce a stem cell fate even in more often passaged primary human fibroblasts.

## 5.2 Generation of Ngn2 induced excitatory neurons from iPSCs

The differentiation of human iPSCs into excitatory neurons by overexpression of mouse Ngn2 and a short forebrain patterning period (Figure 3) was successfully performed several times for a total of 13 different cell lines. Differentiated iNeurons were characterized by electrophysiological measurements (Figure 7), immunocytochemistry (Figure 9, Figure 10), bulk-RNASeq (Figure 12), and single cell RNASeq (Figure 14 - Figure 18).

### 5.2.1 Electrophysiological active, excitatory, and mature iNeurons

The applied neuronal differentiation protocol (Figure 3) reproducibly generated electrophysiological active neurons with similar properties to those generated by comparable differentiation protocols (M. Chen et al., 2020; Nehme et al., 2018; Y. Zhang et al., 2013). Glutamatergic transmission was capable of generating synchronized network activity with bursting behavior (Figure 7 A). Glutamatergic transmission via NMDA and AMPA receptors was verified by application of specific inhibitors (Figure 7 B). Signaling through kainate receptors, not blocked by either NBQX or APV, might be responsible for the remaining action potentials in those silenced cultures. Additionally, there might be neurons present in the culture that display pace-making activity and generate AP without external stimuli. Those pace-maker neurons are important for burst firing and synchronized activity in neuronal networks (Ramirez et al., 2004). In the diencephalon, for instance, those pace-maker neurons are dopaminergic neurons and are able to corelease glutamate. This is in concordance with the expression of vesicular glutamate transporter and transcripts coding for the dopaminergic machinery in cluster 2 of the scRNASeq experiment (4.2.4, 5.2.2). The occurrence of those burst firing events and the induction of several spikes after depolarization of the neurons further supported the presence of functionally mature neurons (Ben-Yosef et al., 2014; Mongiat et al., 2009; Nehme et al., 2018).

Electric properties of the iNeurons were used to further assess the maturity of the iNeurons: the inner resistance ( $R_{in}$ ) is negatively correlated with the maturity of the neurons (Figure 7 D). In line with this, the  $R_{in}$  is negatively influenced by the size of the individual neuron: an increase in neuronal size increases the surface area. It is reported that during the maturation process, a high  $R_{in}$  ( $2\text{ G}\Omega$ ) enhances the excitability in newly born neurons in hippocampal slice recordings. After several weeks of maturation, the  $R_{in}$  of these hippocampal mouse slice cultures reaches levels also measured in iNeurons indicating the presence of mature electrophysiological properties of the iNeurons (Mongiat et al., 2009; Sun et al., 2018). Likewise, for unpatterned Ngn2 neurons, the  $R_{in}$  decreases with increased CAMK2A promoter activity, the latter indicative for neuronal maturation (Nehme et al., 2018). In contrast, a higher  $R_{in}$  was reported for unpatterned iNeurons after 12 weeks of cultivation (M. Chen et al., 2020). Similarly, the  $R_{in}$  of a 2D differentiation (dual SMAD inhibition, no Wnt inhibition, no Ngn2) was stable at high levels (5-fold of the  $R_{in}$  reported here) during a 5 week differentiation (Ben-Yosef et al., 2014).

Maturity of the iNeurons is additionally validated by the presence of a low resting membrane potential for both schizophrenia patient and healthy control cell lines (Figure 7 C). This low resting membrane potential is in line with reports from unpatterned Ngn2 neurons that decrease the membrane potential over the course of the differentiation (Nehme et al., 2018). Similar to the above-described properties of neurons, the miniature EPSC (mEPSC) frequency is maturity dependent and increases with the age of mouse neurons in the dentate gyrus from 0.75 to 1.75 Hz, respectively (Mongiat et al., 2009). The here measured mEPSC frequency (mean  $\sim$  0.75 Hz, Figure 7 D) is in accordance with the average frequency for unpatterned Ngn2 neurons (Y. Zhang et al., 2013). The average amplitude of 20 pA for the excitatory miniEPSC is within the range of unpatterned iNeurons (mean amplitude of 9 - 20 pA depending on the donor) (Y. Zhang et al., 2013) and similar to differentiated neurons cultured on astrocytes (Meijer et al., 2019).

In conclusion, the here generated neurons displayed properties of mature excitatory neurons also detected in primary mouse 2D culture systems and in freshly cut murine brain slices. Moreover, the iNeurons transmitted their activity with glutamate and were capable of generating network bursts.

### **5.2.2 Mature excitatory iNeurons of the peripheral and central nervous system**

After the classification as excitatory neurons according to the recorded electrophysiological behavior, the next objective of this thesis was to further sub-categorize the generated iNeurons. iNeurons were, thus, analyzed by immunocytochemistry (Figure 9, Figure 10), bulk (Figure 12) and single cell RNASeq (Figure 15 - Figure 18) for the expression of marker proteins and genes, respectively. In order to account for the lack of one specific neuronal marker for one specific subtype of neuron, a more generalized approach was used to annotate the cellular identity. This label transfer annotates cell types according to the most similar transcriptomic profile between the annotated data set and the data set of interest. An annotated subtype specific data set from the mouse central and peripheral nervous system was used to transfer subtype specific cell identity labels (Zeisel et al., 2018). Neuronal and non-neuronal marker genes are predicted to be sufficiently similar between mouse and human cells (Johnson et al., 2020). Nevertheless, overexpression of *POU4F1* and *Ngn2* in human iPSC generates a mechano- and cold sensitive peripheral neuronal type that is only present in human but not mouse tissue (Nickolls et al., 2020). In general, all of

the annotated neuronal subtypes were supported or discouraged by the presence or absence of varying additional marker genes. Moreover, a lot of the marker genes are expressed in several clusters and are indicative of several neuronal cell types at the same time. Furthermore, it has to be taken into consideration, that the database relies heavily on differential expressed genes and estimates enrichment of genes in a given cell population accordingly (Zeisel et al., 2018). This can explain why the mitochondrial genes MT-CO3, MT-ND3, MT-ND4L were associated only with enteric neurons and with none of the other cell types even though they encode central subunits of the respiratory chain. Hence, the lack of expression of those subunits in other neurons would result in a heavily impaired energy metabolism (Díaz-García et al., 2017).

Electrophysiologic measurements verified the presence of mature neurons capable of action potential generation and burst firing. Therefore, as a next step to characterize the iNeurons generated with the here described protocol, maturity markers on the transcriptome level were investigated. High levels of *GRIN2A* and *GRIN2B* were specifically expressed in *GPM6A* positive neurons indicating a maturity similar to previously described *Ngn2* iNeurons (Nehme et al., 2018). *GRIN2A*<sup>+</sup> cells in the single cell experiment accounted for only a minority of the sequenced cells similar to other *Ngn2* based differentiation protocols (Nehme et al., 2018). The increase in *GRIN2B* and the decrease in *PRPH* expression is associated with maturity and is reported after 5 days of dual SMAD-Wnt inhibition during iNeuron differentiation (Nehme et al., 2018). Thus, according to their *GRIN2B* and *GRIN2A* gene expression levels (Figure 16), the here described iNeurons resembled neurons that occur in the human brain before birth until early after birth (Nehme et al., 2018). *CAMK2A*, another marker for mature neurons, was primarily expressed in cluster 0, while it was depleted in the neuroblast clusters 1 and 4.

Neuroblasts are immature neuronal cells that inherit the potential to differentiate towards mature neurons as part of adult neurogenesis (Zeisel et al., 2018). Adult neurogenesis takes place for instance in the olfactory system, the subventricular zone, and in the dentate gyrus. Most of the transcriptomes for the neuroblast cell identity originate from cells from P20-P30 mouse brains in the label transfer data set (Zeisel et al., 2018). Hence, this database lacks the power to annotate developmental specific neural subtypes. This might explain the labelling of two clusters as neuroblasts instead of indicating immature, in this case non-adult neurons.

Another maturity and at the same time regional identity marker is *FOXG1*, a transcription factor, which is highly expressed in progenitor cells and at lower levels in mature excitatory neurons of the forebrain (Y. Li et al., 2017; Nehme et al., 2018; Shin et al., 2015). In line with the latter, cells in cluster 3 of the scRNA-Seq data expressed *FOXG1* at low levels while expressing maturity markers and the cortical neuron marker *GPM6A*. Similarly, *FOXG1* expression was high in cluster 6 cells. Cluster 6 was originally annotated as an astrocyte containing clusters. However, the well-known astrocyte markers *GFAP* and *S100B* were not detected in cluster 6 discouraging the annotation of astrocytes (Figure 16). The labelling of cluster 6 as a neuronal progenitor cluster is further supported by the expression of the astrocyte markers *SOX2* and *VIM* which are also expressed in radial glia cells. Radial glia cells are present during development and give rise to neurons, thus, were not part of the post-natal mouse database. Even the inclusion of mouse radial glia cells in the database might not suffice, because the human neuronal development is based on radial glia cell types not present in the mouse brain (Pollan et al., 2015). Furthermore, in the bulk RNASeq data, the levels of the additional neural progenitor markers *VIM* and *APOE* correlated well with the varying degree of *FOXG1* levels. Moreover, the neuronal progenitor fate of cluster 6 is in line with the presence of neuronal progenitor cells in previously published Ngn2 protocols (Lin et al., 2020; Nehme et al., 2018).

In general, *FOXG1* expression was low in bulk RNA-Seq (Figure 12) and only a minority of cells in the single cell experiment (Figure 16) were positive for *FOXG1* in contrast to the protein-based measurements (Figure 9, Figure 10). In the immunocytochemistry experiment a high proportion of neurons were positive for *FOXG1* in line with the previously published high proportion of *FOXG1*<sup>+</sup> cells in unpatterned day-21 iNeurons (Y. Zhang et al., 2013). This difference in the *FOXG1*<sup>+</sup> cell proportion can be a result of :1) *FOXG1* protein persisted in immature neurons while *FOXG1* mRNA has been already degraded, 2) the different media composition for the scRNA-Seq compared to the bulk RNA-Seq experiment altered proportions of progenitors, immature, and mature neurons, 3) *FOXG1* mRNA expression is too low to be reliably quantified especially in the single cell experiment (A. R. Wu et al., 2014) or, 4) the thresholds of the automatic annotation for the immunocytochemistry experiments were not strict enough (compare black to grey percentages in Figure 10) and false positive neurons were included.

In order to further decipher the type of neurons generated with the here applied Ngn2 protocol, the expression of additional marker genes and proteins was investigated. Annotating the subtypes of neurons, though, is not straight-forward. The previously mentioned maturity marker *FOXG1*, is primarily used to ensure forebrain characteristics of the neurons. Nevertheless, *FOXG1* is also expressed in both excitatory and inhibitory neurons of the telencephalon, thalamus, and striatum. In line with this, several of the here and in general utilized cortical layer markers such as SATB2, CTIP2, and CUX1, discriminate neurons only between specific cortical layers. In return, the presence of those cortical markers, though, does not indicate whether the investigated neuron is part of a specific cortical layer: CUX1 expression is specific for the upper layer in the cortex while it is also reported in most other neuronal types of the central and peripheral nervous system. *CUX1* was detected with immunocytochemistry and bulk- and scRNA-Seq in the majority of the iNeurons. This is in line with the published high percentage of CUX1<sup>+</sup> iNeurons in patterned iNeurons (Nehme et al., 2018).

The occurrence of SATB2 positive neurons is in line with other Ngn2 iNeuron differentiation protocols. Although, in the here reported 16 samples, the percentage of SATB2 positive cells is more variable than in the published counts for three different cell lines (Nehme et al., 2018). SATB2 expression is detected in upper cortical layer neurons and in serotonergic hindbrain and enteric neurons (Zeisel et al., 2018). A small minority of SATB2 positive neurons are also located in cortical layer V and VI. In layer V and VI of the mouse cortex, a small number of neurons appear to be *Satb2* and *Ctip2* positive (Alcamo et al., 2008; Cánovas et al., 2015). This double positivity of the neurons for SATB2 and CTIP2 cannot be investigated with the here described experiments due to the fact that the protein expression was investigated on different cultures of the same batch.

*Ctip2*<sup>+</sup> neurons are mainly present in cortical layer V and VI (Nehme et al., 2018; Srinivasan et al., 2012), in the spinal cord, inhibitory hindbrain, glutamatergic neurons of the cortex and midbrain, and peptidergic sensory neurons. The fraction of *SATB2* and *CTIP2* expressing neurons in the scRNA-Seq experiment resembled differentiations with a low fraction of SATB2<sup>+</sup> / CTIP2<sup>+</sup> neurons detected with immunocytochemistry. However, published results for Ngn2 induced neurons did not describe expression of CTIP2 (Nehme et al., 2018). This discrepancy might be explained by 1) the different antibodies used to stain for CTIP2 or 2) the differences in patterning duration. For those differentiations with low *CTIP2*

abundance, high abundance of *TBR1* (lower cortical layer) and *SATB2/CUX1* (upper cortical layer) was detected in contrast to other Ngn2 differentiations (Nehme et al., 2018; Y. Zhang et al., 2013).

The contrasting evidence for upper- and lower- cortical layer identity of the neurons can be explained by the fact that 1) each individual differentiation sample (even from the same differentiation batch) generated different proportions of neuronal subtypes from the cortical layers and the central or peripheral nervous system. 2) peripheral markers were not specifically investigated in the immunocytochemistry experiments; therefore, protein data is lacking for specific peripheral markers such as *PRPH*. The assignment of most of the cells to individual cortical layer markers can be explained by the fact the peripheral cells were 1) not present in the immunocytochemistry batches, 2) situated in neuronal clumps which were excluded from the automatic counting process, 3) outside of the size range of selected nuclei, or 4) the cortical markers indicate not solely neurons of the cortical layer but also the presence of neurons from the peripheral nervous system. Indeed, *PRPH* was present in every cluster of the single cell experiment and *PRPH* expression was high in bulk RNA-Seq.

The occurrence of peripheral cells is a strong indication that the length and or the concentration of the patterning was not successful for a substantial fraction of the cells. The generation of neural crest cells (NCCs) was unintentionally further supported by a sparse seeding density during dual SMAD inhibition (Chambers et al., 2009). Furthermore, several Ngn2 iNeuron differentiation protocols are published to specifically generate peripheral subtypes: *POU4F1* and *Ngn2* overexpression can be used to differentiate iPSC into mechano- and cold-sensitive peripheral neurons of the dorsal root ganglion (Hulme et al., 2020; Nickolls et al., 2020). This peripheral differentiation potential is further supported by *NEUROG2* / *Ngn2* expression in several areas of the peripheral nervous system during development. In line with this, high levels of *PRPH* are detected in bulk RNA-Seq and the majority of clusters of similarly differentiated neurons of other Ngn2 iNeuron scRNA-Seq experiments (M. Chen et al., 2020; Lin et al., 2020). Moreover, unpatterned iNeurons and the here described iNeurons express additional peripheral marker genes such as *ISL1*, *PRPH*, *PHOX2B*, and *POU4F1* (M. Chen et al., 2020; Lin et al., 2020). Independently, expression of *ISL1* and *POU4F1* is reported after dual SMAD inhibition (Chambers et al., 2009). *PHOX2B* and *POU4F1* expression indicate three distinct neuronal *PRPH*<sup>+</sup> groups of

non-CNS neuronal subtypes: sympathetic, enteric, and sensory neurons. Furthermore, in the spinal cord, Ngn2 interacts with *ISL1* and *LHX3* to generate motor neurons (Velasco et al., 2017). Nevertheless, Wnt inhibition and elongated Ngn2 expression in the beginning of the differentiation is supposed to hinder *POU4F1* expression (Lin et al., 2020). In the here applied protocol, Ngn2 was only induced by doxycycline treatment until day-24 to deplete the early conversion TF *Ngn2* (Schuurmans et al., 2004) for the remaining duration of differentiation. In line with this, endogenous *NEUROG2* expression which is induced by mouse Ngn2 overexpression was not detected in the iNeurons on day-49 (Y. Zhang et al., 2013). The reduction in doxycycline treatment length might have been sufficient to mimic the short Ngn2 induction protocols: 5 instead of 14 days of doxycycline positively correlates with an increase in CNS neurons (Lin et al., 2020).

In peripheral neurons, *Ngn2* acts in combination with *POU4F1* to generate sensory neurons. Moreover, *POU4F1* and *Ngn2* overexpression can be used to differentiate iPSC into mechano- and cold-sensitive peripheral neurons of the dorsal root ganglion (Hulme et al., 2020; Nickolls et al., 2020). *POU4F1* expression in DRGNs starts early during the development in combination with *NRP1* (not detected in cluster 4). DRGNs can be generated by overexpression of *POU4F1* and Ngn2 in neural crest progenitors. This double factor overexpression generates neurons positive for *ISL1*, *TAC1*, *NTRK2*, and *NTRK3*, but negative for *NTRK1*. In line with the absence of *NTRK1*, dual SMAD inhibition, present here, is reported to hinder *NTRK1* expression. PVALB was also highly expressed in this cluster indicating mechanosensitive DRG neurons in line with *NTRK3* expression (C. L. Li et al., 2016).

The published hindbrain markers *MAP1A*, *GABBR2*, and *SLC1A6* were highly expressed in the generated iNeurons in contradiction to published results (Nehme et al., 2018). This is in line with results from other Ngn2 iNeuron differentiation protocols that postulate the need of 6 days patterning before the induction of Ngn2 (M. Chen et al., 2020) or 4 days of patterning during the initial doxycycline induction phase (Nehme et al., 2018). This process is in stark contrast to the 2 days of patterning in the here described differentiation protocol.

Similar cluster markers were detected between the clusters described here and those published for unpatterned and patterned iNeurons (M. Chen et al., 2020; Lin et al., 2020; Nehme et al., 2018; Y. Zhang et al., 2013). However, some cluster types were not detected separately (*CD99<sup>+</sup>*) or presented altered co-gene expression: *CD99<sup>+</sup>* cluster contained more

*PRPH<sup>+</sup>/SNCG<sup>+</sup>/ISL1<sup>+</sup>* than *GPM6A<sup>+</sup>* cells (Lin et al., 2020). There is a substantial fraction of cells located in clusters 2,5 positive for *CD99*, however, cluster separation was performed according to other genes, e.g., *PHOX2B/RUNX1* and *PVALB* for clusters 2 and 5, respectively. *CD99* expression was predominantly found in the postmitotic cluster and not in the *VIM* and *GNG5* expressing cluster 6 (Lin et al., 2020) in contrast to its published role as an inhibitor of differentiation of NPCs (Lee et al., 2003). A more detailed analysis of subclusters, e.g., the *SLC6A17* negative cells of cluster 4, should be applied in the future to further pinpoint the specific subtypes of the cells. For this purpose, a refinement of the clustering could improve cell type annotation.

It must be noted that a lot of marker combinations pointed towards the presence of solely excitatory neurons despite the fact that enzymes to generate the NT ACh and GABA were detected in very few cells. *SLC18A3* or *SLC6A1* and *SLC6A2*, the transporters to load synaptic vesicles with ACh or GABA, were found in very few cells of cluster 2 or a subcluster in cluster 6 and cluster 3, respectively. However, degrading enzymes and receptors were present in the sequenced neurons. Likewise, *GAD<sup>+</sup>* neurons can be detected in unpatterned *Ngn2* driven differentiation (Rosa et al., 2020). The expression of those cholinergic and GABAergic signaling related genes could also be a result of the activity of certain pathways and TF without cholinergic or GABAergic synaptic input. Similarly, there are dopaminergic related transporters and receptors but no *TH* to generate dopamine or a related NT. Dopaminergic fate of neurons is supported by experiments which combined overexpression of *Ngn2* with *Nurr1 (NR4A2)* overexpression (Andersson et al., 2007).

In conclusion, the overlap of marker expression detected between the different *Ngn2* iNeuron protocols suggests that *Ngn2* overexpression indeed can generate neurons with varying subtypes of regional identity. In the here developed and employed protocol, the patterning duration was only sufficient to increase the telencephalic excitatory fate over the *PRPH<sup>+</sup>* fate without excluding the generation of peripheral and hindbrain neurons. It might be, therefore, helpful in the future, to consider cell identity abundances in order to improve the comparability between cell lines for bulk approaches and increase the power and meaningfulness of the results. For other differentiation methods, such as the dual SMAD inhibition, the normalization towards the abundance of e.g. astrocytes, oligodendrocytes, neuronal, and other non-neuronal markers or cells was used successfully (Hoffman et al., 2017). However, this is not a straightforward process considering the

identified overlap of cell identities for specific markers and the lack of a conclusive human single cell annotation database.

### 5.3 Effects of schizophrenia on the transcriptome and epigenome

Induced pluripotent stem cells from eleven individuals were differentiated into iNeurons for 49 days. Their transcriptome (Table 22, Table 23) and open chromatin regions (Table 24) were investigated for disease status (schizophrenia, healthy control) dependent differences. Only five genes were significantly differentially regulated between schizophrenia and healthy control derived iNeurons. These genes were examined for their tissue specificity (GTEX, Human Protein Atlas - proteinatlas.org) and association with schizophrenia. *OR8A1* is associated in two GWASs with schizophrenia due to its close vicinity to the associated variant rs7127399 (Goes et al., 2015; Lam et al., 2019). *OR8A1* is expressed in excitatory neurons of the brain and influences receptor activity. The pseudogene *MTRNR2L8* (HGNC: 37165) is also significantly upregulated in postmortem hippocampus from patients compared to healthy controls (Hwang et al., 2013). ENSG00000280255 (*RP5-1007F24.1*, AC004947.2) is an uncategorized gene which is associated with uterine corpus endometrial carcinoma. *PAX8-AS1* and *PAX8* are expressed at low levels in the brain and expressed at medium levels throughout the body. *PAX8-AS1* is also differentially expressed in a microarray data set of postmortem brain tissue investigating schizophrenia (Sabaie et al., 2021). *GUSBP1* (Glucuronidase Beta Pseudogene 1) is expressed in the brain and in the body at low levels and at medium levels in the cerebral hemisphere and the cerebellum.

Samples impacting the differential expression differed for each of the reported genes. On one hand, this small number of differentially expressed genes is in stark contrast to other studies. Some studies detect over 1,600 or 500 differentially regulated genes using an EB based differentiation protocol and applying similar thresholding (Brennand et al., 2011; Roussos et al., 2016). In the study from Brennand et al., three of the in this thesis used fibroblast derived iPSC lines originate from the same patients. The reprogramming process, though, differed between the studies: lentiviral (Brennand et al., 2011) and plasmid-based (this thesis) overexpression of the reprogramming factors.

On the other hand, in a follow up study of the same laboratory with 36 and 40 independent cell lines from ten COS and nine healthy control individuals, only one significantly differential expressed gene was detected between the healthy and the diseased neurons after applying corrections for the heterogeneity in cell types after neuronal differentiation (Hoffman et al., 2017). In general, previously published studies, similar to the here described differentiation experiment, are underpowered to detect small effects on gene expression of noncoding SNPs. Postmortem studies suggest that changes in gene expression are in the range of 1.03 - 1.33 fold (inverting downregulated expression ratios) (Fromer et al., 2016). A scRNA-Seq experiment with sensory neurons indicates that at least 20 - 80 individual donors are required to sufficiently detect effects of regulatory variants with moderately large effect sizes (Schwartzentruber et al., 2018). This is supported by estimates that with the heterogeneity present in postmortem brain tissue, several thousand samples are needed for differential gene expression analysis (Fromer et al., 2016). Therefore, the presented data set likely lacks the power to investigate differential gene expression. An additional unrelated drawback of this data set is, that the applied sequencing method (3' polyA tail) cannot differentiate between isoforms of the same gene. Therefore, this data cannot be used to replicate altered splicing previously reported in DLPFC samples (Fromer et al., 2016).

Next, open chromatin regions were analyzed for their differential presence or accessibility in schizophrenia patient and healthy derived neurons (Table 24, Figure 20). Similar to the RNA-Seq experiment, differences originated from subgroups of all cell lines rather than disease status prompted association with schizophrenia. This differential open chromatin is e.g. present in the 5' UTR of the FOS gene depicted in Figure 21 of several cell lines. One interpretation of this phenomenon is that the accessibility of these loci is impacted by noncoding schizophrenia associated SNPs independently of disease status. This SNP specific alteration in accessible chromatin can also be investigated with robust allele specific quantification and quality control (RASQUAL tool) (Kumasaka et al., 2016). However, in order to cover common SNPs, sufficiently larger cohorts of individuals need to be investigated. The importance for such caQTL on gene expression is highlighted for instance by an iPSC based neuronal model for the schizophrenia associated genomic region of MIR137 (Forrest et al., 2017). The total number of accessible chromatin peaks from iPSC derived neurons is approximately one fourth of the number of peaks detected with the

here reported data set (Forrest et al., 2017). One reason for this difference in number of accessible chromatin loci might be that the before mentioned study is based on two instead of 11 (this study) different cell lines. In conclusion, the here detected differences in gene regulation and transcription between iNeurons differentiated from healthy control and schizophrenia patient derived iPSC were marginal.

## 5.4 Depolarization induced changes in day-49 iNeurons

There is strong evidence that schizophrenia affects synaptic transmission, therefore, in this thesis one objective was to analyze the effect of schizophrenia on epigenomic and translational changes upon neuronal activity (Brennand et al., 2011, 2015; Grunwald et al., 2019; Kathuria et al., 2019). In general, neuronal excitatory activity results in a depolarization of the postsynaptic membrane and an influx of the second messenger ionic calcium. Hence, I analyzed the effects of membrane depolarization and indirect  $\text{Ca}^{2+}$  entry on the here generated iNeurons at the 1) electrophysiological, 2) gene expression, and 3) gene regulatory level.

### 5.4.1 Disruption of network bursting after high KCl depolarization

First, electrophysiological measurements were performed which confirmed by field recording of synaptically active iNeurons the presence of network burst activity (Figure 19). APs were detected automatically for all measured electrodes. Unexpectedly, though, APs were still detected with several field electrodes in the depolarized state. AP generation after depolarization is inhibited, due to the fact that voltage gated sodium channels are necessary to generate APs in neurons. These voltage gated sodium channels, however, are inactivated after opening and cannot reopen until the cell membrane is repolarized. In neurons which are highly depolarized due to the hKCl stimulation for 5 h, voltage gated sodium channels are blocked and are not able to unblock to generate APs (De Lera Ruiz & Kraus, 2015). Thus, no AP or bursts of APs were expected to be present in the highly depolarized neuronal cultures during hKCl treatment.

The recordings were performed right after the application of the hKCl solution. Hence, the here recorded APs might resemble the initial response to the depolarization stimulus of the neuronal membrane. Additionally, the hKCl solution in the recording chamber was not

constantly perfused but changed manually once before the recording with a pipette. This manual media exchange might have created microenvironments around the neuronal cell membrane which might have led to only a subtle increase in membrane potential at the very beginning of the stimulation paradigm. Such a subtle repolarization of the iNeurons could be responsible for the burst detection in some channels. The detection of APs in the depolarized state might have also resulted from aberrations in membrane properties due to the unphysiological depolarization conditions. Manual inspection of the recordings from two field electrode, however, could not entirely exclude the presence of APs. A substantial fraction of the network was silent after the hKCl treatment and network burst firing was ablated.

#### **5.4.2 Transcriptomic and epigenomic alterations after depolarization of iNeurons**

The depolarization of the iNeurons before the harvest strongly impacted gene expression and chromatin accessibility (Figure 19 - Figure 21, Table 23, Table 24). This was verified by the first two dimensions of the PCA for both experiments. As indicated (4.3, 5.3), however, the disease status did not affect the differential expressed genes between the stimulation conditions (Figure 19 C). In contrast, a previously published study of a similar hKCl stimulation reported a substantial reduction (~90 %) of number of DEGs in schizophrenia compared to healthy control iPSC neurons (Roussos et al., 2016). In this study, EB based protocols were used to generate NPCs and subsequently neurons (Brennand et al., 2011, 2015). Therefore, the described schizophrenia specific decreased number of DEGs after hKCl stimulation might have primarily been influenced by the cell type composition in the neuronal differentiation cultures. Additionally, as indicated beforehand for the case-control studies, a 4vs4 (Roussos et al., 2016) or 5vs6 cell line (this thesis) approach is under-powered to investigate differential gene expression in a polygenic complex disease scenario with small effect sizes.

Differentially regulated genes of iNeurons overlapped at a similar proportion with DEGs of mouse neurons and neurons from a different iPSC neuronal differentiation protocol (Figure 19 E). This overlap between the samples is especially intriguing, because the published data set was stimulated for only 4 h in the presence of a L-type  $\text{Ca}^{2+}$  channel agonist and a NMDA receptor antagonist to prevent excitotoxicity (Qiu et al., 2016). The here applied hKCl solution, though, did not contain such small molecules. Nevertheless, only 3 of the

significantly associated GO terms (117 in total) indicated induction of apoptotic processes: regulation of neuron apoptotic process (GO:0043523), intrinsic apoptotic signaling pathway in response to DNA damage by p53 class mediator (GO:0042771), and regulation of neuron death (GO:1901214). This might be indicative of a small subgroup of cells undergoing apoptosis and subsequently releasing and or receiving corticosteroids. The majority of the GO terms significantly associated with hKCl stimulation were based around encoding stimuli indicating an active regulation of transcriptional events: transcription (11 GO terms), transcription factor (7 GO terms), polymerase II (9 GO terms), and MAP (for the  $\text{Ca}^{2+}$  induced MAPK pathway, 4 GO terms). Normal synaptic activity strengthens synapses and induces maturation of developing neurons, hence, a substantial fraction of the GO terms was related to neuronal development. Of note, previously reported human specific genes were also differentially regulated in the human iNeurons (Qiu et al., 2016).

As expected, several immediate early genes, for instance *EGR3*, were induced by hKCl stimulation and sustained high expression levels until harvest. Other gene levels such as those for *EGR4* were either already decreased like reported for primary mouse neurons or were not altered in their expression. It was not possible to conclude the dynamics of gene expression during hKCl stimulation due to the fact that the temporal expression changes were not investigated in this study (Beagan et al., 2020; Tyssowski et al., 2018). Secondary RGs such as the potassium channel *KCNA1* were also significantly upregulated in line with previously published reports (Tyssowski et al., 2018). *KCNA1* is, together with other potassium channels, responsible for the decrease in neuronal excitability by lowering the resting membrane potential. Therefore, overexpression of such channels as a response to massive depolarization came to no surprise.

Histone deacetylases such as *HDAC1*, and *HDAC5* but not the previously reported *HDAC9*, were also induced upon stimulation (Tyssowski et al., 2018). Histone deacetylases are important for chromatin remodeling. HDACs deplete acetylation marks on lysine residues with a focus on the N-terminal end of target histones. Reduction of acetylation induces either the replacement, establishment, or erasing of other posttranslational modifications on lysine such as methylation, ubiquitination, and sumoylation (Seto & Yoshida, 2014). In general, deacetylation leads to a reduction in accessibility of chromatin regions and in turn less transcriptional activity. In line with this, alterations in histone acetylation around the

immediate early *FOS* gene (Figure 21 A), in response to stimulation, directly influences the bursting behavior of mouse primary neurons (L. F. Chen et al., 2019).

Therefore, genome wide changes in the accessibility of chromatin were analyzed in the context of hKCl stimulation of iNeurons. In total, 7,630 ARCh were detected. Only those ARCh that were more accessible upon hKCl stimulation were enriched for schizophrenia variants. This enrichment for schizophrenia variants in accessible regions of ARCh supports the hypothesis that schizophrenia impacts on the level of every single neuron how stimuli are encoded in the chromatin and at transcription levels.

Only a small subset of those ARCh, however, were located within 100 kb from an ARG start site (Ensemble annotation). The restriction to a 100 kb window around the gene start site is specifically problematic for long genes, genes with several TSS, and in gene rich genomic loci (Mora et al., 2016). In addition, the change in accessibility of an open chromatin region within a certain distance to the gene start site is not necessarily proportional or even correlated with an increase in gene expression: Chromatin regions can be primed or poised so that those regions are accessible without actively regulating gene expression. Primed chromatin is not transcriptionally active (yet) due to, for instance, lack of expression of TFs such as immediate early TF, inactive TF such as dephosphorylated CREB, or regulation by repressing TFs (Spicuglia & Vanhille, 2012). Therefore, techniques to investigate active chromatin like chromatin immunoprecipitation (ChIP) or Cut&Tag for active chromatin marks (Kaya-Okur et al., 2019; S. J. Wu et al., 2021) can improve the annotation of active chromatin regions.

Furthermore, two independent differentiation samples were used to link ARCh with ARGs without considering the heterogeneity of the iNeuron culture system which might have affected the accessibility of chromatin and gene expression independent of the hKCl stimulation. Multi-omic approaches such as scRNA-Seq and scATAC-Seq of the same cell, provided for instance by the scMultiome kit from 10xGenomics, offer the unique possibility to link chromatin accessibility and gene expression of the same cell. In general, linking approaches are purely based on correlation without considering the underlying 3D conformation of the chromosome. Similar to the accessibility of chromatin, those 3D interactions are regulated by stimulation: those which are pre-formed but not active and those which need to be formed and activated. Preformed loops can be also decommissioned by stimulation (Beagan et al., 2020). In conclusion, hKCl treatment altered

the epigenetic and transcriptomic landscape of iNeurons in a similar fashion than reported for other iPSC derived and primary mouse neurons independent of disease state.

## 5.5 Gene regulatory activity is influenced by cell type, stimulation, and genomic variants

The case control study reported in this thesis was performed with a small number of cell lines ( $n = 11$ ) and, therefore, lacked the power to detect the small changes in gene expression or accessibility of chromatin regions expected to be caused by genetic variants. This power estimate is based on calculations that hundreds of postmortem brain samples, a more heterogeneous tissue than iNeurons, are needed to analyze schizophrenia related variation (Fromer et al., 2016). Additionally, the here reported number of donors is insufficient to investigate in a genome - wide manner the impact of less common SNPs.

Although, MPRA are an artificial system in which the pEF regulates the activity of a minimal promoter (minP) and, hence, alters gene expression of a reporter gene and a barcode, it enables to investigate the activity of gene regulatory elements (GREs) containing preselected SNPs. The MPRA offers the unique opportunity to specifically investigate the activity of putative enhancer regions (pERs) with a sufficient power in disease relevant cell types , cell states, and allele status. Additionally, MPRA is able to resolve functional causal variants within one LD block which is not the case in other approaches such as eQTL studies or fine-mapping algorithms (Kreimer et al., 2017). Furthermore, the concordance rate between MPRA and SNP specific accessible chromatin (DNAse hypersensitivity sites) is very high (> 90 %) supporting MPRA as a meaningful approach (Tewhey et al., 2016) to investigate the impact from alleles of interest on GRE activity. Those pERs with an allelic skew in regulatory activity are referred to as expression modulating variants (emVars).

### 5.5.1 Applied strategies to improve sensitivity and specificity of the MPRA

In total, 5,601 pERs were synthesized for three separate oligo pools (sequences and genomic locations in 8.1 - 8.4). 4,951 of the synthesized pERs were recovered in the libraries with both alleles (Figure 26, Figure 27). The recovery of 88 % of all pERs is lower than in published protocols which also used the amplification strategy to barcode pERs (98.4 % recovery of > 39,000 pERs, 100 % recovery of > 2,000 pERs) (Klein et al., 2020; Tewhey et

al., 2016). It is also lower than with the pre-barcoded strategy for which e.g. 95 % of the original constructs were recovered for subsequent analyses (Ulirsch et al., 2016).

In contrast to published studies (Melnikov et al., 2012, 2014; Tewhey et al., 2016; Ulirsch et al., 2016), the threshold to consider enhancer fragments for the final analysis was set independent of the number of barcodes but based on abundance in the association library. For a sufficient barcoding strategy, a minimum number of 20 barcodes per pEF was targeted. The eQTLLib and DevLib were close to previously reported average of 73 or median of 100 barcodes per pEF (Klein et al., 2020; Tewhey et al., 2016), respectively (Figure 25 C - E). Hence, a 50 - 100 times more complex plasmid library was required after each cloning step (Melnikov et al., 2012, 2014; Ulirsch et al., 2016). The complexity of the library is directly translatable into colony forming units (Table 26) assuming that each bacteria colony on the agar plate contains only one plasmid. Due to limited cloning capacity, the eQTLLib did not achieve this multiplication factor. The ValLib was cloned according to those guidelines and had an excellent reproducibility ( $R_s \sim 1$ ) between technical replicates (Figure 28 D - F).

Strikingly, the reproducibility between replicates was higher in mouse neurons than iNeurons, even though a homogeneous cell population increases the reproducibility between replicates (Melnikov et al., 2012, 2014; Ulirsch et al., 2016). Compared to the iNeuron culture with mostly excitatory neurons, the primary mouse neuronal cultures contained excitatory as well as inhibitory neurons. This high reproducibility between the primary mouse samples compared to the iNeurons might be partly explained by the similarity of cell type abundances in cultures of primary mouse neurons compared to iNeurons. In line with this, the same MPRA library or different libraries show a high reproducibility of 0.99 or 0.95 in immortalized cell lines between replicates, respectively (Tewhey et al., 2016), while the correlation between the iNeuron samples for the eQTLLib was similar to the correlation between lymphoblastoid cell lines with HepG2 cells (Tewhey et al., 2016). Furthermore, an increase in number of quantifications for the same enhancer – barcode fragment increases the reproducibility between samples. This increase in analyzed fragments can be either achieved by an increased number of 1) infected cells and/or 2) enhancer – barcode fragments per cell (Myint et al., 2020). Only 8 million primary mouse neurons compared to up to 24 million iNeurons were plated for each sample (Table 27). Even though the number of harvested cells (compared to the before mentioned plated

cell numbers) were not counted, the increased number of recovered barcode – enhancer fragments per cell was likely to be the cause for the increased reproducibility in the mouse neurons compared to the iNeurons.

The DevLib was sequenced with ~4-fold higher read depth per sample compared to the eQTLLib and ValLib (Table 28). Nevertheless, the recovery of barcodes passing the cutoffs per sample was similar between the eQTL and ValLib independent of the cell type. In the case of the DevLib, barcode recovery in the iNeurons was around half of the one detected in mouse neurons (Figure 26). This low recovery can be explained by the fact that: 1) different viruses were used for the iNeurons, the NPCs, and the mouse neurons for the DevLib. The first viral isolation used for the iNeuron experiment was prepared with less complex DNA transfected in a lower number of HEK293T cells generating a lower complex virus (data not shown). 2) iNeurons were infected with less viral particles per cell than mouse neurons since they already contained several integrations of lentiviruses coding for the doxycycline inducible Ngn2 overexpression machinery necessary to generate the iNeurons. The increase in number of investigated iNeurons per sample partly could ameliorate this difference in recovered barcodes per samples for the eQTLLib (Table 14).

A sufficient barcode length is important to minimize the occurrence of the same barcode for two different pEFs. For most of the published MPRAAs a barcode length of 20 bp was reported accounting for more than  $10^{12}$  different barcodes (Inoue & Ahituv, 2015; Patwardhan et al., 2012). The here applied barcoding strategy used a 16 bp barcode, 1 bp longer than in the first MPRA paper investigating SNP effects (Tewhey et al., 2016). The utilization of 16 bp long barcodes resulted in a barcode overlap between the three libraries of 1.6 %. Those barcodes which were associated with several different pEFs were excluded from the analysis. To ensure a high sensitivity of the assay, barcodes were only considered until a hamming distance of 1. This barcoding strategy decreased the reported arbitrary impact of single outlier barcodes onto GRE activity. Such an outlier activity is reported to occur in 0.4 % - 3.2 % of the barcodes. The exact outlier rate depends on the cell type (Myint et al., 2020) and was neither described previously nor tested for the here described cell types.

Expression levels of barcodes can also be affected by template switching. Template switching occurs preferably during the generation of lenti- or retroviruses between sequence elements with a high similarity; in this case the minP-nLuc DNA fragment. The

proportion of template switching barcodes was not analyzed for the here described MPRA libraries even though in an unrelated library type 5.5 % of the elements were reported to be affected (Sack et al., 2016). Due to time constraints and the overall high average number of barcodes per pEF, template switching was not further investigated for the here described library. Amplification bias of the barcode due to GC content is reported not to impact count estimates (Ulirsch et al., 2016). Only one unassociated barcode was discovered per MPRA library in samples but not in the association library indicating, in general, a specific and error-proof barcoding system (Figure 27).

To call meaningful variants, barcode abundance in the cDNA library was normalized towards the barcode abundance in the enhancer – barcode association plasmid library (Figure 24 A, after step 2 described in Figure 28 A). This approach accounted for the over- and underrepresentation of barcodes due to the cloning steps. Normalization of barcode abundance increases the reproducibility between replicates and increases the sensitivity of the assay to detect small changes in gene regulation in dependence of allele status (Tewhey et al., 2016). The widespread range in barcode abundance per element is in line with previous studies. An abundance threshold guaranteed a sufficient baseline detection rate of each pEF. As expected, a higher abundance threshold reduced the number of investigated pEFs and vice versa (Figure 28 B, C).

Negative and positive control fragments were used (Table 25, 8.1), as described previously, to set the activity threshold in the here described cell types separately (Klein et al., 2020; Tewhey et al., 2016). For the downstream analysis, only active enhancers were investigated for an impact of the variant on gene regulation (Figure 29 A). This activity threshold was set to increase the biological meaningfulness of the approach, since the altered activity of an inactive GRE will most likely not be biologically relevant. A set of randomly selected control regions with non-disease associated SNPs and well-known regions with a variant dependent impact on gene regulation were utilized to set the thresholds for emVar detection. As expected, higher p-value thresholds increased monotonically with FPR and FDR. Similarly, a higher p-value threshold increased the proportion of emVars in relation to all active enhancers (Figure 30A - D). Empirical statistical thresholding was performed to decrease false positive rates and increase the reproducibility of the method. These empirical thresholds ensured, together with the barcoding strategy, the sensitivity of the assay to detect small effects in gene regulation driven by allelic variants. Nevertheless, the

conservative thresholds applied in this study impeded the sensitivity of the assay, and a fraction of likely meaningful SNPs might not be detected. Such a non-active emVar with significantly different activities between the two alleles was detected at the location of the pER containing rs4702 at the FURIN gene locus (Figure 36). It was classified as a non-active ER in line with an inaccessible chromatin region in iNeurons and postmortem brain tissue. This is, however, in contrast to the described effect of rs4702 on gene expression of FURIN in unpatterned iNeurons (Schrode et al., 2019). In the MPRA in patterned iNeurons, though, rs35346340 affected the activity of the GRE and was located in an open chromatin region (ATAC-Seq) of iNeurons.

### 5.5.2 Context dependent allele specific gene regulatory activity

Similar to eQTLs, also the here annotated emVars act in a cell type specific manner (Table 27, Table 28). 42 or 43 % of emVars were shared between iNeurons and primary mouse neurons or NPCs in the baseline condition, respectively (Figure 32 C). This is in line with results from other MPRA in which 52 % of ERs are shared between HepG2 cells and a lymphoblastoid cell line (Tewhey et al., 2016). The biological significance of cell type specific activity of ERs in the MPRA (Musunuru et al., 2010) is supported by their biological equivalent: cell type specific accessible chromatin regions. Likewise, TF expression and activity is cell type specific (Dunham et al., 2012; Sankaran et al., 2012). TF binding and recruiting of Polymerase II is measured by the MPRA. The allelic effect on enhancer activity measured by MPRA correlates well with sequence based predictions of TF binding strength (Avsec et al., 2021; Tewhey et al., 2016). Therefore, the cell type specific activity of a tested ER at the GRIN2A locus correlates well with low or high expression of GRIN2A in NPCs or neurons, respectively (Figure 33). As mentioned beforehand, GRIN2A / GRIN2B expression in neural cells is an indicator for neural maturation (Nehme et al., 2018). This cell type specific activity of ERs was, as expected, detected between cell types at different developmental stages (NPCs and iNeurons) and species (primary mouse and human iNeurons).

Surprisingly, NPC emVars were not specifically enriched for open chromatin loci of neural precursor cells or any of the other tissue or cell types (Figure 34 C). This might be explained by the fact that the NPC differentiation was performed for 8 days, and cells were seeded sparsely after the infection with the MPRA library and low density NPC cultures induce

neuronal maturation (Chambers et al., 2009). Therefore, the analyzed NPCs might neither resemble the open chromatin state of early NPCs nor mature neurons. Moreover, the utilized open chromatin data did not originate from NPCs generated with the same differentiation protocol. In contrast to NPCs, iNeurons and mouse neurons at baseline condition were enriched for emVars located in open chromatin regions specific to their respective cell types. iNeuron emVars were also enriched in somatostatin expressing interneurons in line with their high expression of *SST* described in the scRNA-Seq experiment (Figure 18). Overlap between mouse and iNeurons for pan-neuronal emVars was validated with this enrichment analysis underscoring the utility of primary mouse neurons as a model system for the analysis of human genetic variants (Figure 35). This is supported by the conserved TFBS between human and mice which resembles the machinery driving ER and emVar (Cheng et al., 2014).

Open chromatin regions are affected by the neuronal activity (Malik et al., 2014; Tyssowski et al., 2018) and schizophrenia associated loci are enriched in ARCh regions. Therefore, the effect of activity was also tested on the pERs tested in the here described MPRA assay (Table 28). This approach was based on reports that allelic impact on the accessibility of ARCh regions is reported in an eQTL study on immune stimulated monocytes (Kim-Hellmuth et al., 2017). Several thousand stimulation dependent open chromatin regions were detected in day-49 iNeurons and the MPRA analysis identified 25 % of all ERs as stimulation specific emVars. Hence, this observation suggests an impact of schizophrenia associated genetic variants on baseline and also on network activity induced gene expression. This altered response of neurons to stimuli can be a potential mechanism that contributes to alterations in neuronal plasticity.

In line with previous studies in K562 cells, around two thirds of the emVars associated with an eQTL were annotated as functional emVars in the MPRA (Tewhey et al., 2016). In the here described schizophrenia MPRA, the GWAS derived association p-value was not a predictor for functionality. The here described study did not repeat that 34 - 41 % of the top associated allele per eQTL were the functional SNPs in MPRA (Tewhey et al., 2016). Additionally, association p-values are not stable, because of the small number of people used to calculate the GWAS summary statistics for the schizophrenia trait and LD disequilibrium (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2014).

Indeed, as described here with the example of the FES/FURIN loci, emVars can have opposite effects on gene regulation in dependence of the cell state. Similarly, at the FES/FURIN loci, eQTL studies imply the opposite effect of the same SNP on the two nearby located genes indicating a higher complexity of cis gene regulation architecture. Therefore, all emVars which overlapped with an eQTL were tested for discordance or concordance in effect direction in comparisons to eQTLs (Figure 30). This analysis assumed that a positive fold change in gene expression linked to an eQTL is associated with an increased activity of the corresponding enhancer. In total 71 % of emVars were located in an eQTL. 73 % of those colocalized emVars regulated the enhancer activity in the same direction as the gene expression in postmortem human brain in at least one of the investigated conditions (cell type and stimulation).

Discordance between the direction of the eQTL and the emVars can indicate 1) a cell type specific emVar effect only detected in the reference human postmortem brain tissue (. Brain tissue contains several neuronal subtypes but also a high proportion of glia cells such as oligodendrocytes and astrocytes (Figure 30 D, E). Therefore, bulk RNA expression is biased towards pan-tissue eQTLs or eQTLs originating from the dominating cell population and cannot discriminate between cell type specific expression. The MPRA, however, focuses on neurons and therefore is less affected by masking effects from sample heterogeneity. 2) The gene expression measured in the postmortem brain is not only regulated by enhancer activity but also by mechanisms altering the stability of the mRNA. miRNAs and antisense transcripts are known to affect mRNA stability. Besides, miRNAs themselves can also be regulated by schizophrenia associated SNPs (Forrest et al., 2017). Therefore, altered gene expression does not immediately imply an altered activity of the enhancer in reality. In concordance with this, only a subset of all variants (23 – 64 %) are predicted to affect enhancer or promoter activity (Farh et al., 2015; Tewhey et al., 2016). 3) Antisense transcript expression can also be altered by emVars and therefore, only indirectly modulate gene expression. The eQTL algorithm, predicts in such a case a decreasing effect of the SNP, even though the enhancer of the SNP actually increases antisense RNA expression (van Arensbergen et al., 2019). 4) On average, the effect sizes of emVars are moderate at  $\log_2(\text{FC}) = 1.28$  (Ulirsch et al., 2016). In another study, only 22 % or 3-5 % of emVars impacted enhancer activity by more than  $\log_2(\text{FC})$  1.2 or 2, respectively (Patwardhan et al., 2012; Tewhey et al., 2016). Similarly, in the here described MPRA,

$\log_2(\text{FC})$  between allelic activities did not reach 3. Therefore, very weak effects may fall below the limit of detection from the MPRA. This detection limit can also explain that the cell type with higher reproducibility (mouse neurons) annotated ~100 emVars more at baseline conditions than the iPSC derived cell types. 5) Furthermore, as described beforehand, the tested enhancer fragments were only 160 - 170 bp long. Therefore, this MPRA cannot model regulatory processes that require additional sequence context such as nearby chromatin structure or presence of further DNA-binding co-factor(s) (Klein et al., 2020). The eQTL approach, however, is based on the full genome and transcriptome. Nevertheless, MPRA is a scalable method to investigate methodologically disease associated SNPs in different cellular states and conditions.

## 5.6 Functional link between emVar loci and altered target gene expression

The results of the MPRA, though, represent only the first step to decipher the genetic cause of schizophrenia. MPRA analyzes the potential of pERs to be active in certain cell types and conditions, and whether the activity of the ER is affected by allele status. Therefore, as a next step, it is important to link emVars to genes which are regulated in their expression and, hence, affected in their expression by the presence or absence of the schizophrenia associated SNP in the emVar. For this purpose, chromatin conformation links of human postmortem brain tissue were used to annotate target genes for the emVar containing loci. However, the low resolution of the publicly available HiC data set only sufficed to associate several genes or no gene to individual emVars. The rate of emVars linked to genes was increased to nearly 50 % when utilizing eQTL information (Figure 37 B), however, the problem of several linked genes to one emVar persisted (Figure 37 A). Therefore, CRISPRi was used to increase the percentage of links between emVars and genes. CRISPRi uses the specificity of the Cas9 system. The inactivated Cas9 protein (dCas9) is, in this case, linked to KRAB. KRAB activity decreases the activity of the targeted enhancer.

The schizophrenia associated *TCF4* locus was used to showcase the dCas9-KRAB system in iPSC derive iNeurons. *TCF4* is associated with schizophrenia by several SNPs located in the close vicinity to and in intronic regions of *TCF4* (Figure 40 A). Additionally, one SNP (rs1272363) in the 3' UTR of *TCF4* is known to affect gene expression in a luciferase based reporter assay (Mohamed et al., 2019). However, rs1272363 is ~250 bp away from an

MPRA pER which did not significantly affect gene regulation in the here investigated neural cell systems.

*TCF4* expression measured by RNA microarray or RT-PCR is upregulated 1.3-fold or 2.6-fold in schizophrenia patient compared to healthy control derived neurons, respectively (Brennand et al., 2011). *TCF4* is significantly upregulated in postmortem DPLFC of patients suffering from schizophrenia compared to healthy control (Figure 38 A). In line with this *TCF4* expression measured in bulk RNA-Seq was increased ~1.5-fold in iNeurons derived from schizophrenia patients compared to healthy controls (Figure 38 B). Strikingly, this upregulation in schizophrenia patient derived compared to healthy control derived iNeurons was not validated by relative quantification of *TCF4*, although cell lines partially overlapped between the relative quantification and the bulk RNA-Seq experiment (Figure 38 C). To verify the *TCF4* loci as a meaningful candidate for the CRISPRi experiment, *TCF4* protein expression was measured by Western Blot (Figure 38 D). A case-control specific comparison of *TCF4* levels was not performed, because the *TCF4* antibody detected several isoforms of *TCF4*. Additionally, MAP2 protein levels could only be measured in independent quantitative experiments from the same protein lysate. Unfortunately, these methodological problems for the protein detection could not be solved during the course of this thesis.

The stable dCas9 cell line (C1) was infected (Figure 39 A) with gRNAs targeting the short and long promoter of *TCF4* and several schizophrenia associated loci within the *TCF4* locus (Table 7, Table 9, Figure 40 A). In B16 and B17, *TCF4* was downregulated between 0 to 50 %. scRNA-Seq, bulk RNA-Seq, and qPCR experiments in other cell systems indicate that gRNA mediated knockdown reduces the expression by at most 90 % with a, here not utilized, improved version of the CRISPRi system when targeted to the promoter of the gene of interest (Alerasool et al., 2020; Duke et al., 2020; Tian et al., 2019; S. Zhang et al., 2020). The effect of the gRNA mediated knockdown is variable between replicates and between the two differentiation batches (Figure 40 B). B16 and B17 were differentiated subsequently from the same cryopreserved iPSC stock and received the same batches of astrocytes. The differentiations for the first batch (B16) were started after 2 weeks of iPSC cultivation in doxycycline containing media to induce stable expression of dCas9 (data generated by Dr. Miriam Gagliardi, data not shown). iPSCs from B17 were treated 7 days longer with doxycycline than those from B16 before starting the differentiation. The gRNA

combination with the maximal *TCF4* downregulation was not congruent between the two batches. One reason might be that *TCF4* expression level quantification was variable because of its low baseline expression levels and fold changes between the control. Another reason might be that quantitative PCR technology might not be suitable to reliably detect the subtle effect of CRISPRi gene regulation. Nevertheless, the variance of the *MAP2* normalized *TCF4* expression levels within and between the two batches is within the range of expectation (Ho et al., 2016). It was not tested whether any guide combination affected expression of additional proximal and distal genes, or any genes regulated by *TCF4* itself.

In general, the here described CRISPRi experiment was not ideal to verify the impact of the emVar at SNP2 and SNP3 loci on *TCF4* expression even though dCas9 and gRNA expression was verified for both loci. The control gRNA vectors (not shown) for this experiment could not be used for normalization, because 1) the non-targeting vector (gRNA against the *OCT4* promoter) contained the same resistance (puromycin) as the Ngn2 vector and the blasticidin S selection could not be performed. Hence, the iNeurons grew at different densities affecting the *MAP2* expression similar to 2) the no-gRNA control. Therefore, the strategy was followed to normalize the CRISPRi experiments to the highest *TCF4* expressing gRNA pool (long\_prom\_d). Besides, *TCF4* detection was not analyzed isoform specific, due to the high homology of human *TCF4* (iNeurons) and mouse *Tcf4* expressed by the mouse derived astrocytes. The here utilized qPCR primer detected specifically the human 3'UTR present in all isoforms and, therefore, could not resolve isoforms specific up- or downregulation.

Nevertheless, the selection of functional SNPs based on their activity in MPRA leads to a decreased number of disease associated loci specific for context (cell type and condition) for further low-throughput screens. This strongly decreases the number of loci in need of follow-up experiments to link emVars to target genes in e.g., pooled CRISPRi screens. This is especially meaningful, since MPRA can test several thousand SNPs at the same time while CRISPRi screens are limited in number of targeted loci. Therefore, the MPRA approach reduces the number of necessary follow-up studies to decipher the functional impact of altered gene expression on complex endophenotypes.

## 5.7 Conclusion

In conclusion, the experiments performed here describe a roadmap to investigate complex traits in human disease specific cell types. In this thesis, this experimental pipeline was applied to the devastating disease schizophrenia which affects 1 % of the general population. Schizophrenia and healthy control derived primary fibroblasts or PBMCs, were successfully reprogrammed into iPSCs in order to analyze the effect of schizophrenia variants in human neural cells. Genetic variants can alter gene regulation and, hence, gene expression levels (bulk and scRNA-Seq) as well as (bulk) accessible chromatin were investigated in a case-control setup. In order to investigate the genetic effects in disease relevant cell types, the previously published Ngn2 overexpression protocol was adapted for the various applied methods. The results of the bulk RNA-Seq and open chromatin experiments are consistent with power analyses that smaller sized ( $n=10$ ) case-control studies are not capable to reveal meaningful differential results between schizophrenia patient and healthy controls (Fromer et al., 2016). Therefore, I tested the effect of schizophrenia associated variants on gene regulatory elements with a massively parallel reporter assay (MPRA). Here, the feasibility of a MPRA at SNP resolution in iPSCs derived NPCs and iNeurons was validated for the first time. For the interrogated variants associated with schizophrenia, this thesis implicated 620 emVars affecting allele specific gene regulation in disease relevant neural cell types. Additionally, it was validated that emVars, similar to open chromatin regions and gene expression, are cell type and cell state specific. Furthermore, the dCas9-KRAB system was validated to decrease TCF4 expression when targeted to TSSs and intergenic SNP loci in iPSC derived iNeurons.

In summary, this work is highly relevant for the analysis of complex diseases. The growing need for functional annotation of GWAS associated variants clearly underlines the importance of high-throughput assays such as MPRA. MPRA results can be utilized to train machine learning models (Avsec et al., 2021; Movva et al., 2019). In follow-up studies, the dCas9 system can then be used in a pooled screen approach with single cell readout (Gasperini et al., 2019) to increase the throughput to link emVar effects to target genes. These emVar linked genes can, subsequently, serve as potential new therapeutic targets.

## 6 Bibliography

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications*, 9(1), 1911. <https://doi.org/10.1038/s41467-018-04252-2>

Ahmad, R., Sportelli, V., Ziller, M., Spengler, D., & Hoffmann, A. (2018). Tracing early neurodevelopment in schizophrenia with induced pluripotent stem cells. *Cells*, 7(9), 140. <https://doi.org/10.3390/cells7090140>

Alasoo, K., Rodrigues, J., Mukhopadhyay, S., Knights, A. J., Mann, A. L., Kundu, K., Hale, C., Dougan, G., & Gaffney, D. J. (2018). Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nature Genetics*, 50(3), 424–431. <https://doi.org/10.1038/s41588-018-0046-7>

Alcamo, E. A., Chirivella, L., Dautzenberg, M., Dobreva, G., Fariñas, I., Grosschedl, R., & McConnell, S. K. (2008). Satb2 Regulates Callosal Projection Neuron Identity in the Developing Cerebral Cortex. *Neuron*, 57(3), 364–377. <https://doi.org/10.1016/j.neuron.2007.12.012>

Alerasool, N., Segal, D., Lee, H., & Taipale, M. (2020). An efficient KRAB domain for CRISPRi applications in human cells. *Nature Methods* 2020 17:11, 17(11), 1093–1096. <https://doi.org/10.1038/s41592-020-0966-x>

Andersson, E. K. I., Irvin, D. K., Ahlsjö, J., & Parmar, M. (2007). Ngn2 and Nurr1 act in synergy to induce midbrain dopaminergic neurons from expanded neural stem and progenitor cells. *Experimental Cell Research*, 313(6), 1172–1180. <https://doi.org/10.1016/j.yexcr.2006.12.014>

Ashuach, T., Fischer, D. S., Kreimer, A., Ahituv, N., Theis, F. J., & Yosef, N. (2019). MPRAAnalyze: Statistical framework for massively parallel reporter assays. *Genome Biology*, 20(1), 1–17. <https://doi.org/10.1186/s13059-019-1787-z>

Avramopoulos, D. (2018). Recent Advances in the Genetics of Schizophrenia. *Molecular Neuropsychiatry*, 4(1), 35–51. <https://doi.org/10.1159/000488679>

Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J. R., Grabska-Barwinska, A., Taylor, K. R., Assael, Y., Jumper, J., Kohli, P., & Kelley, D. R. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. *Nature Methods*, 18(10), 1196–1203. <https://doi.org/10.1038/s41592-021-01252-x>

Badner, J. A., & Gershon, E. S. (2002). Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Molecular Psychiatry*, 7(4), 405–411. <https://doi.org/10.1038/sj.mp.4001012>

Baghbaderani, B. A., Syama, A., Sivapatham, R., Pei, Y., Mukherjee, O., Fellner, T., Zeng, X., & Rao, M. S. (2016). Detailed Characterization of Human Induced Pluripotent Stem Cells Manufactured for Therapeutic Applications. *Stem Cell Reviews and Reports*, 12(4), 394–420. <https://doi.org/10.1007/s12015-016-9662-8>

Banito, A., Rashid, S. T., Acosta, J. C., Li, S. De, Pereira, C. F., Geti, I., Pinho, S., Silva, J. C., Azuara, V., Walsh, M., Vallier, L., & Gil, J. (2009). Senescence impairs successful reprogramming to pluripotent stem cells. *Genes and Development*, 23(18), 2134–2139. <https://doi.org/10.1101/gad.1811609>

Bardy, C., van den Hurk, M., Eames, T., Marchand, C., Hernandez, R. V., Kellogg, M., Gorris, M., Galet, B., Palomares, V., Brown, J., Bang, A. G., Mertens, J., Böhnke, L., Boyer, L., Simon, S., & Gage, F. H. (2015). Neuronal medium that supports basic synaptic functions and activity of human neurons in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 112(20), E2725–34. <https://doi.org/10.1073/pnas.1504393112>

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>

Beagan, J. A., Pastuzyn, E. D., Fernandez, L. R., Guo, M. H., Feng, K., Titus, K. R., Chandrashekhar, H., Shepherd, J. D., & Phillips-Cremins, J. E. (2020). Three-dimensional genome restructuring across timescales of activity-induced neuronal gene expression. *Nature Neuroscience*, 23(6), 707–717. <https://doi.org/10.1038/s41593-020-0634-6>

Beers, J., Linask, K. L., Chen, J. A., Siniscalchi, L. I., Lin, Y., Zheng, W., Rao, M., & Chen, G. (2015). A cost-effective and efficient reprogramming platform for large-scale production of integration-free human induced pluripotent stem cells in chemically defined culture. *Scientific Reports*, 5(1). <https://doi.org/10.1038/srep11319>

Ben-Yosef, D., Telias, M., & Segal, M. (2014). Electrical maturation of neurons derived from human embryonic stem cells. *F1000Research*, 3. <https://doi.org/10.12688/f1000research.4943.2>

Bennett, K. P., Brown, E. M., Santos, H. D. los, Poegel, M., Kiehl, T. R., Patton, E. W., Norris, S., Temple, S., Erickson, J., McGuinness, D. L., & Boles, N. C. (2019). Identifying Windows of Susceptibility by Temporal Gene Analysis. *Scientific Reports*, 9(1), 1–14. <https://doi.org/10.1038/s41598-019-39318-8>

Bergsland, M., Werme, M., Malewicz, M., Perlmann, T., & Muhr, J. (2006). The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes and Development*, 20(24), 3475–3486. <https://doi.org/10.1101/gad.403406>

Beyfuss, K., & Hood, D. A. (2018). A systematic review of p53 regulation of oxidative stress in skeletal muscle. In *Redox Report* (Vol. 23, Issue 1, pp. 100–117). Taylor and Francis Ltd. <https://doi.org/10.1080/13510002.2017.1416773>

Bharathan, S. P., Manian, K. V., Aalam, S. M. M., Palani, D., Deshpande, P. A., Pratheesh, M. D., Srivastava, A., & Velayudhan, S. R. (2017). Systematic evaluation of markers used for the identification of human induced pluripotent stem cells. *Biology Open*, 6(1), 100–108. <https://doi.org/10.1242/bio.022111>

Bleuler, E. (2011). Eugen Bleuler: Dementia Praecox oder Gruppe der Schizophrenien. In *Psychiatrische Praxis* (Vol. 38, Issue 08). Deuticke. <https://doi.org/10.1055/s-0031-1295586>

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z. D., Ziller, M., Croft, G. F., Amoroso, M. W., Oakley, D. H., Gnarke, A., Eggan, K., & Meissner, A. (2011). Reference maps of human es and ips cell variation enable high-throughput characterization of pluripotent cell lines. *Cell*, 144(3), 439–452. <https://doi.org/10.1016/j.cell.2010.12.032>

Boer, G. J. (1994). Ethical guidelines for the use of human embryonic or fetal tissue for experimental and clinical neurotransplantation and research. *Journal of Neurology*, 242(1), 1–13. <https://doi.org/10.1007/BF00920568>

Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezhuk, Y., Raytselis, Y., Sayers, E. W., Tao, T., Ye, J., & Zaretskaya, I. (2013). BLAST: a more efficient report with usability improvements. *Nucleic Acids Research*, 41(Web Server issue). <https://doi.org/10.1093/nar/gkt282>

Børglum, A. D., Demontis, D., Grove, J., Pallesen, J., Hollegaard, M. V., Pedersen, C. B., Hedemand, A., Mattheisen, M., Uitterlinden, A., Nyegaard, M., Ørntoft, T., Wiuf, C., Didriksen, M., Nordentoft, M., Nöthen, M. M., Rietschel, M., Ophoff, R. A., Cichon, S., Yolken, R. H., ... Mors, O. (2014). Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. *Molecular Psychiatry*, 19(3), 325–333. <https://doi.org/10.1038/mp.2013.2>

Brennand, K., Savas, J. N., Kim, Y., Tran, N., Simone, A., Hashimoto-Torii, K., Beaumont, K. G., Kim, H. J., Topol, A., Ladran, I., Abdelrahim, M., Matikainen-Ankney, B., Chao, S., Mrksich, M., Rakic, P., Fang, G., Zhang, B., Yates, J. R., & Gage, F. H. (2015). Phenotypic differences in hiPSC NPCs derived from patients with schizophrenia. *Molecular Psychiatry*, 20(3), 361–368. <https://doi.org/10.1038/mp.2014.22>

Brennand, K., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., Li, Y., Mu, Y., Chen, G., Yu, D., McCarthy, S., Sebat, J., & Gage, F. H. (2011). Modelling schizophrenia using human induced pluripotent stem cells. *Nature*, 473(7346), 221–225. <https://doi.org/10.1038/nature09915>

Bronner, A. (1961). Psychiatrie: Ein Lehrbuch für Studierende und Ärzte. *American Journal of Psychotherapy*, 15(4), 688–689. <https://doi.org/10.1176/appi.psychotherapy.1961.15.4.688>

Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, 2015, 21.29.1–21.29.9. <https://doi.org/10.1002/0471142727.mb2129s109>

Burke, E. E., Chenoweth, J. G., Shin, J. H., Collado-Torres, L., Kim, S. K., Micali, N., Wang, Y., Colantuoni, C., Straub, R. E., Hoeppner, D. J., Chen, H. Y., Sellers, A., Shabbani, K., Hamersky, G. R., Diaz Bustamante, M., Phan, B. D. N., Ulrich, W. S., Valencia, C., Jaishankar, A., ... Jaffe, A. E. (2020). Dissecting transcriptomic signatures of neuronal differentiation and maturation using iPSCs. *Nature Communications*, 11(1), 462. <https://doi.org/10.1038/s41467-019-14266-z>

Cánoyas, J., Berndt, F. A., Sepúlveda, H., Aguilar, R., Veloso, F. A., Montecino, M., Oliva, C., Maass, J. C., Sierralta, J., & Kukuljan, M. (2015). The specification of cortical subcerebral projection neurons depends on the direct repression of TBR1 by CTIP1/BCL11a. *Journal of Neuroscience*, 35(19), 7552–7564. <https://doi.org/10.1523/JNEUROSCI.0169-15.2015>

Cárdenas, A., Villalba, A., de Juan Romero, C., Picó, E., Kyrousi, C., Tzika, A. C., Tessier-Lavigne, M., Ma, L., Drukker, M., Cappello, S., & Borrell, V. (2018). Evolution of Cortical Neurogenesis in Amniotes Controlled by Robo Signaling Levels. *Cell*, 174(3), 590–606.e21. <https://doi.org/10.1016/j.cell.2018.06.007>

Cardno, A. G., Marshall, E. J., Coid, B., Macdonald, A. M., Ribchester, T. R., Davies, N. J., Venturi, P., Jones, L. A., Lewis, S. W., Sham, P. C., Gottesman, I. I., Farmer, A. E., McGuffin, P., Reveley, A. M., & Murray, R. M. (1999). Heritability estimates for psychotic disorders: The Maudsley Twin psychosis series. *Archives of General Psychiatry*, 56(2), 162–168. <https://doi.org/10.1001/archpsyc.56.2.162>

Castle, J. C. (2011). SNPs occur in regions with less genomic sequence conservation. *PLoS ONE*, 6(6), e20660. <https://doi.org/10.1371/journal.pone.0020660>

Chambers, S. M., Fasano, C. A., Papapetrou, E. P., Tomishima, M., Sadelain, M., & Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnology*, 27(3), 275–280. <https://doi.org/10.1038/nbt.1529>

Chan, M. K., Tsang, T. M., Harris, L. W., Guest, P. C., Holmes, E., & Bahn, S. (2011). Evidence for disease and antipsychotic medication effects in post-mortem brain from schizophrenia patients. *Molecular Psychiatry*, 16(12), 1189–1202. <https://doi.org/10.1038/mp.2010.100>

Chang, M.-Y., Sun, W., Ochiai, W., Nakashima, K., Kim, S.-Y., Park, C.-H., Kang, J. S., Shim, J.-W., Jo, A.-Y., Kang, C.-S., Lee, Y.-S., Kim, J.-S., & Lee, S.-H. (2007). Bcl-X L /Bax Proteins Direct the Fate of Embryonic Cortical Precursor Cells. *Molecular and Cellular Biology*, 27(12), 4293–4305. <https://doi.org/10.1128/mcb.00031-07>

Chen, C., Lee, G. A., Pourmorady, A., Sock, E., & Donoghue, M. J. (2015). Orchestration of neuronal differentiation and progenitor pool expansion in the developing cortex by SoxC genes. *Journal of Neuroscience*, 35(29), 10629–10642. <https://doi.org/10.1523/JNEUROSCI.1663-15.2015>

Chen, C., Zhu, H., Stauffer, F., Caravatti, G., Vollmer, S., Machauer, R., Holzer, P., Möbitz, H., Scheufler, C., Klumpp, M., Tiedt, R., Beyer, K. S., Calkins, K., Guthy, D., Kiffe, M., Zhang, J., & Gaul, C. (2016). Discovery of Novel Dot1L Inhibitors through a Structure-Based Fragmentation Approach. *ACS Medicinal Chemistry Letters*, 7(8), 735–740. <https://doi.org/10.1021/acsmmedchemlett.6b00167>

Chen, L. F., Lin, Y. T., Gallegos, D. A., Hazlett, M. F., Gómez-Schiavon, M., Yang, M. G., Kalmeta, B., Zhou, A. S., Holtzman, L., Gersbach, C. A., Grandl, J., Buchler, N. E., & West, A. E. (2019). Enhancer Histone Acetylation Modulates Transcriptional Bursting Dynamics of Neuronal Activity-Inducible Genes. *Cell Reports*, 26(5), 1174–1188.e5. <https://doi.org/10.1016/j.celrep.2019.01.032>

Chen, M., Maimaitili, M., Habekost, M., Gill, K. P., Mermet-Joret, N., Nabavi, S., Febbraro, F., & Denham, M. (2020). Rapid generation of regionally specified CNS neurons by sequential patterning and conversion of human induced pluripotent stem cells. *Stem Cell Research*, 48, 101945. <https://doi.org/10.1016/j.scr.2020.101945>

Cheng, Y., Ma, Z., Kim, B. H., Wu, W., Cayting, P., Boyle, A. P., Sundaram, V., Xing, X., Dogan, N., Li, J., Euskirchen, G., Lin, S.,

Lin, Y., Visel, A., Kawli, T., Yang, X., Patacsil, D., Keller, C. A., Giardine, B., ... Snyder, M. P. (2014). Principles of regulatory information conservation between mouse and human. *Nature*, 515(7527), 371–375. <https://doi.org/10.1038/nature13985>

Chong, H. Y., Teoh, S. L., Wu, D. B. C., Kotirum, S., Chiou, C. F., & Chaiyakunapruk, N. (2016). Global economic burden of schizophrenia: A systematic review. In *Neuropsychiatric Disease and Treatment* (Vol. 12, pp. 357–373). Dove Medical Press Ltd. <https://doi.org/10.2147/NDT.S96649>

Chou, B.-K., Gu, H., Gao, Y., Dowey, S. N., Wang, Y., Shi, J., Li, Y., Ye, Z., Cheng, T., & Cheng, L. (2015). A Facile Method to Establish Human Induced Pluripotent Stem Cells From Adult Blood Cells Under Feeder-Free and Xeno-Free Culture Conditions: A Clinically Compliant Approach. *Stem Cells Translational Medicine*, 4(4), 320–332. <https://doi.org/10.5966/sctm.2014-0214>

Coleman, J. R. I., Euesden, J., Patel, H., Folarin, A. A., Newhouse, S., & Breen, G. (2016). Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray. *Briefings in Functional Genomics*, 15(4), 298–304. <https://doi.org/10.1093/bfgp/elv037>

Corces, M. R., Corces, M. R., Trevino, A. E., Hamilton, E. G., Greenside, P. G., Sinnott-Armstrong, N. A., Vesuna, S., Satpathy, A. T., Rubin, A. J., Montine, K. S., Wu, B., Kathiria, A., Cho, S. W., Mumbach, M. R., Carter, A. C., Kasowski, M., Orloff, L. A., Risca, V. I., Kundaje, A., ... Chang, H. Y. (2017). Omni-ATAC-seq: Improved ATAC-seq protocol. *Protocol Exchange*. <https://doi.org/10.1038/protex.2017.096>

Danecek, P., McCarthy, S. A., Consortium, H. S., & Durbin, R. (2016). A method for checking genomic integrity in cultured cell lines from SNP genotyping data. *PLoS ONE*, 11(5), e0155014. <https://doi.org/10.1371/journal.pone.0155014>

Davis, J. O., Phelps, J. A., & Bracha, H. S. (1995). Prenatal development of monozygotic twins and concordance for schizophrenia. *Schizophrenia Bulletin*, 21(3), 357–366. <https://doi.org/10.1093/schbul/21.3.357>

De Lera Ruiz, M., & Kraus, R. L. (2015). Voltage-Gated Sodium Channels: Structure, Function, Pharmacology, and Clinical Indications. *Journal of Medicinal Chemistry*, 58(18), 7093–7118. <https://doi.org/10.1021/jm501981g>

De Los Angeles, A., Ferrari, F., Xi, R., Fujiwara, Y., Benvenisty, N., Deng, H., Hochedlinger, K., Jaenisch, R., Lee, S., Leitch, H. G., Lensch, M. W., Lujan, E., Pei, D., Rossant, J., Wernig, M., Park, P. J., & Daley, G. Q. (2015). Hallmarks of pluripotency. *Nature*, 525(7570), 469–478. <https://doi.org/10.1038/nature15515>

de Wert, G., Berghmans, R. L. P., Boer, G. J., Andersen, S., Brambati, B., Carvalho, A. S., Dierickx, K., Elliston, S., Nunez, P., Osswald, W., & Vicari, M. (2002). Ethical guidance on human embryonic and fetal tissue transplantation: A European overview. *Medicine, Health Care, and Philosophy*, 5(1), 79–90. <https://doi.org/10.1023/A:1014213125573>

DeLisi, L. E., Shaw, S. H., Crow, T. J., Shields, G., Smith, A. B., Larach, V. W., Wellman, N., Loftus, J., Nanthakumar, B., Razi, K., Stewart, J., Comazzi, M., Vita, A., Heffner, T., & Sherrington, R. (2002). A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. *American Journal of Psychiatry*, 159(5), 803–812. <https://doi.org/10.1176/appi.ajp.159.5.803>

Díaz-García, C. M., Mongeon, R., Lahmann, C., Koveal, D., Zucker, H., & Yellen, G. (2017). Neuronal Stimulation Triggers Neuronal Glycolysis and Not Lactate Uptake. *Cell Metabolism*, 26(2), 361–374.e4. <https://doi.org/10.1016/j.cmet.2017.06.021>

Dick, E., Matsa, E., Young, L. E., Darling, D., & Denning, C. (2011). Faster generation of hiPSCs by coupling high-titer lentivirus and column-based positive selection. *Nature Protocols*, 6(6), 701–714. <https://doi.org/10.1038/nprot.2011.320>

Diecke, S., Lisowski, L., Kooreman, N. G., & Wu, J. C. (2014). Second generation codon optimized minicircle (CoMiC) for nonviral reprogramming of human adult fibroblasts. *Methods in Molecular Biology*, 1181, 1–13. [https://doi.org/10.1007/978-1-4939-1047-2\\_1](https://doi.org/10.1007/978-1-4939-1047-2_1)

Diecke, S., Lu, J., Lee, J., Termglinchan, V., Kooreman, N. G., Burridge, P. W., Ebert, A. D., Churko, J. M., Sharma, A., Kay, M. A., & Wu, J. C. (2015). Novel codon-optimized mini-intronic plasmid for efficient, inexpensive, and xeno-free induction of pluripotency. *Scientific Reports*, 5. <https://doi.org/10.1038/srep08081>

Dong, Y., Xiong, M., Chen, Y., Tao, Y., Li, X., Bhattacharyya, A., & Zhang, S. C. (2020). Plasticity of Synaptic Transmission in Human Stem Cell-Derived Neural Networks. *IScience*, 23(2), 100829. <https://doi.org/10.1016/j.isci.2020.100829>

Doni Jayavelu, N., Jajodia, A., Mishra, A., & Hawkins, R. D. (2020). Candidate silencer elements for the human and mouse genomes. 11(1), 1–15. <https://doi.org/10.1038/s41467-020-14853-5>

Drozd, A. M., Walczak, M. P., Piaskowski, S., Stoczyńska-Fidelus, E., Rieske, P., & Grzela, D. P. (2015). Generation of human iPSCs from cells of fibroblastic and epithelial origin by means of the oriP/EBNA-1 episomal reprogramming system. *Stem Cell Research and Therapy*, 6(1), 122. <https://doi.org/10.1186/s13287-015-0112-3>

Duke, C. G., Bach, S. V., Revanna, J. S., Sultan, F. A., Southern, N. T., Davis, M. N., Carullo, N. V. N., Bauman, A. J., Phillips, R. A., & Day, J. J. (2020). An Improved CRISPR/dCas9 Interference Tool for Neuronal Gene Suppression. In *Frontiers in Genome Editing* (Vol. 2, p. 2020.05.26.116822). bioRxiv. <https://doi.org/10.3389/fgeed.2020.00009>

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B. K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., ... Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74. <https://doi.org/10.1038/nature11247>

E McKinney, C. (2014). Human iPSC Models: A Platform for Investigating Neurodevelopmental Diseases. *Journal of Molecular and Genetic Medicine*, 08(03), 1–10. <https://doi.org/10.4172/1747-0862.1000122>

Ebrahimi, B. (2015). Reprogramming barriers and enhancers: Strategies to enhance the efficiency and kinetics of induced pluripotency. *Cell Regeneration*, 4(1), 4:10. <https://doi.org/10.1186/s13619-015-0024-9>

Esteban, M. A., Wang, T., Qin, B., Yang, J., Qin, D., Cai, J., Li, W., Weng, Z., Chen, J., Ni, S., Chen, K., Li, Y., Liu, X., Xu, J., Zhang, J. (2014). Human iPSC Models: A Platform for Investigating Neurodevelopmental Diseases. *Journal of Molecular and Genetic Medicine*, 08(03), 1–10. <https://doi.org/10.4172/1747-0862.1000122>

S., Li, F., He, W., Labuda, K., Song, Y., ... Pei, D. (2010). Vitamin C Enhances the Generation of Mouse and Human Induced Pluripotent Stem Cells. *Cell Stem Cell*, 6(1), 71–79. <https://doi.org/10.1016/j.stem.2009.12.001>

Farh, K. K. H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., Shores, N., Whitton, H., Ryan, R. J. H., Shishkin, A. A., Hatan, M., Carrasco-Alfonso, M. J., Mayer, D., Luckey, C. J., Patsopoulos, N. A., De Jager, P. L., Kuchroo, V. K., Epstein, C. B., Daly, M. J., ... Bernstein, B. E. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518(7539), 337–343. <https://doi.org/10.1038/nature13835>

Farmer, A. E., McGuffin, P., & Gottesman, I. I. (1987). Twin Concordance for DSM-III Schizophrenia: Scrutinizing the Validity of the Definition. *Archives of General Psychiatry*, 44(7), 634–641. <https://doi.org/10.1001/archpsyc.1987.01800190054009>

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. B., Okada, Y., Raychaudhuri, S., Daly, M. J., ... Price, A. L. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, 47(11), 1228–1235. <https://doi.org/10.1038/ng.3404>

Finucane, H. K., Reshef, Y. A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P. R., Lareau, C., Shores, N., Genovese, G., Saunders, A., Macosko, E., Pollack, S., Perry, J. R. B., Buenrostro, J. D., Bernstein, B. E., Raychaudhuri, S., McCarroll, S., ... Price, A. L. (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature Genetics*, 50(4), 621–629. <https://doi.org/10.1038/s41588-018-0081-4>

Forrest, M. P., Zhang, H., Moy, W., McGowan, H., Leites, C., Dionisio, L. E., Xu, Z., Shi, J., Sanders, A. R., Greenleaf, W. J., Cowan, C. A., Pang, Z. P., Gejman, P. V., Penzes, P., & Duan, J. (2017). Open Chromatin Profiling in hiPSC-Derived Neurons Prioritizes Functional Noncoding Psychiatric Risk Variants and Highlights Neurodevelopmental Loci. *Cell Stem Cell*, 21(3), 305–318.e8. <https://doi.org/10.1016/j.stem.2017.07.008>

Fromer, M., Roussos, P., Sieberts, S. K., Johnson, J. S., Kavanagh, D. H., Perumal, T. M., Ruderfer, D. M., Oh, E. C., Topol, A., Shah, H. R., Klei, L. L., Kramer, R., Pinto, D., Gümüş, Z. H., Cicek, A. E., Dang, K. K., Browne, A., Lu, C., Xie, L., ... Sklar, P. (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature Neuroscience*, 19(11), 1442–1453. <https://doi.org/10.1038/nn.4399>

Fulco, C. P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S. R., Perez, E. M., Kane, M., Cleary, B., Lander, E. S., & Engreitz, J. M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science*, 354(6313), 769–773. <https://doi.org/10.1126/science.aag2445>

Fullerton, J. M., & Nurnberger, J. I. (2019). Polygenic risk scores in psychiatry: Will they be useful for clinicians? [version 1; peer review: 4 approved]. *F1000Research*, 8, 1293. <https://doi.org/10.12688/f1000research.18491.1>

Garitaonandia, I., Amir, H., Boscolo, F. S., Wambua, G. K., Schultheisz, H. L., Sabatini, K., Morey, R., Waltz, S., Wang, Y. C., Tran, H., Leonardo, T. R., Nazon, K., Slavin, I., Lynch, C., Li, Y., Coleman, R., Romero, I. G., Altun, G., Reynolds, D., ... Laurent, L. C. (2015). Increased risk of genetic and epigenetic instability in human embryonic stem cells associated with specific culture conditions. *PLoS ONE*, 10(2), e0118307. <https://doi.org/10.1371/journal.pone.0118307>

Gasperini, M., Hill, A. J., McFaline-Figueroa, J. L., Martin, B., Kim, S., Zhang, M. D., Jackson, D., Leith, A., Schreiber, J., Noble, W. S., Trapnell, C., Ahituv, N., & Shendure, J. (2019). A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell*, 176(1–2), 377–390.e19. <https://doi.org/10.1016/j.cell.2018.11.029>

Genga, R. M. J., Kernfeld, E. M., Parsi, K. M., Parsons, T. J., Ziller, M. J., & Maehr, R. (2019). Single-Cell RNA-Sequencing-Based CRISPRi Screening Resolves Molecular Drivers of Early Human Endoderm Development. *Cell Reports*, 27(3), 708–718.e10. <https://doi.org/10.1016/j.celrep.2019.03.076>

Girdhar, K., Hoffman, G. E., Jiang, Y., Brown, L., Kundakovic, M., Hauberg, M. E., Francoeur, N. J., Wang, Y. chih, Shah, H., Kavanagh, D. H., Zharovsky, E., Jacobov, R., Wiseman, J. R., Park, R., Johnson, J. S., Kassim, B. S., Sloofman, L., Mattei, E., Weng, Z., ... Akbarian, S. (2018). Cell-specific histone modification maps in the human frontal lobe link schizophrenia risk to the neuronal epigenome. *Nature Neuroscience*, 21(8), 1126–1136. <https://doi.org/10.1038/s41593-018-0187-0>

Goes, F. S., McGrath, J., Avramopoulos, D., Wolyniec, P., Pirooznia, M., Ruczinski, I., Nestadt, G., Kenny, E. E., Vacic, V., Peters, I., Lencz, T., Darvasi, A., Mulle, J. G., Warren, S. T., & Pulver, A. E. (2015). Genome-wide association study of schizophrenia in Ashkenazi Jews. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, 168(8), 649–659. <https://doi.org/10.1002/ajmg.b.32349>

Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. *Science*, 249(4973), 1129–1133. <https://doi.org/10.1126/science.2204114>

Gore, A., Li, Z., Fung, H. L., Young, J. E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M. A., Kiskinis, E., Lee, J. H., Loh, Y. H., Manos, P. D., Montserrat, N., Panopoulos, A. D., Ruiz, S., Wilbert, M. L., Yu, J., Kirkness, E. F., ... Zhang, K. (2011). Somatic coding mutations in human induced pluripotent stem cells. *Nature*, 471(7336), 63–67. <https://doi.org/10.1038/nature09805>

Gottesman, I., & Bertelsen, A. (1989). Confirming Unexpressed Genotypes for Schizophrenia. *Archives of General Psychiatry*, 46(10), 867. <https://doi.org/10.1001/archpsyc.1989.01810100009002>

Grabundzija, I., Wang, J., Sebe, A., Erdei, Z., Kajdi, R., Devaraj, A., Steinemann, D., Szuhai, K., Stein, U., Cantz, T., Schambach, A., Baum, C., Izsák, Z., Sarkadi, B., & Ivics, Z. (2013). Sleeping Beauty transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells. *Nucleic Acids Research*, 41(3), 1829–1847. <https://doi.org/10.1093/nar/gks1305>

Grunwald, L. M., Stock, R., Haag, K., Buckenmaier, S., Eberle, M. C., Wildgruber, D., Storchak, H., Kriebel, M., Weißgraeber, S., Mathew, L., Singh, Y., Loos, M., Li, K. W., Kraushaar, U., Fallgatter, A. J., & Volkmer, H. (2019). Comparative

characterization of human induced pluripotent stem cells (hiPSC) derived from patients with schizophrenia and autism. *Translational Psychiatry*, 9(1), 179. <https://doi.org/10.1038/s41398-019-0517-3>

Gusev, A., Mancuso, N., Won, H., Kousi, M., Finucane, H. K., Reshef, Y., Song, L., Safi, A., McCarroll, S., Neale, B. M., Ophoff, R. A., O'Donovan, M. C., Crawford, G. E., Geschwind, D. H., Katsanis, N., Sullivan, P. F., Pasaniuc, B., & Price, A. L. (2018). Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nature Genetics*, 50(4), 538–548. <https://doi.org/10.1038/s41588-018-0092-1>

Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology*, 20(1), 296. <https://doi.org/10.1186/s13059-019-1874-1>

Hall, L. S., Medway, C. W., Pain, O., Pardiñas, A. F., Rees, E. G., Escott-Price, V., Pocklington, A., Bray, N. J., Holmans, P. A., Walters, J. T. R., Owen, M. J., & O'Donovan, M. C. (2020). A transcriptome-wide association study implicates specific pre-and post-synaptic abnormalities in schizophrenia. *Human Molecular Genetics*, 29(1), 159–167. <https://doi.org/10.1093/hmg/ddz253>

Haraldsson, H. M., Ettlinger, U., & Sigurdsson, E. (2011). Developments in schizophrenia genetics: From linkage to microchips, deletions and duplications. In *Nordic Journal of Psychiatry* (Vol. 65, Issue 2, pp. 82–88). Taylor & Francis. <https://doi.org/10.3109/08039488.2011.552734>

Harrison, P. J., & Weinberger, D. R. (2005). Schizophrenia genes, gene expression, and neuropathology: On the matter of their convergence. In *Molecular Psychiatry* (Vol. 10, Issue 1, pp. 40–68). Nature Publishing Group. <https://doi.org/10.1038/sj.mp.4001558>

Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., Van De Lagemaat, L. N., Smith, K. A., Ebbert, A., Riley, Z. L., Abajian, C., Beckmann, C. F., Bernard, A., Bertagnolli, D., Boe, A. F., Cartagena, P. M., Mallar Chakravarty, M., Chapin, M., Chong, J., ... Jones, A. R. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, 489(7416), 391–399. <https://doi.org/10.1038/nature11405>

Hilker, R., Helenius, D., Fagerlund, B., Skytthe, A., Christensen, K., Werge, T. M., Nordentoft, M., & Glenthøj, B. (2018). Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biological Psychiatry*, 83(6), 492–498. <https://doi.org/10.1016/j.biopsych.2017.08.017>

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America*, 106(23), 9362–9367. <https://doi.org/10.1073/pnas.0903106>

Ho, S. M., Hartley, B. J., TCW, J., Beaumont, M., Stafford, K., Slesinger, P. A., & Brennand, K. J. (2016). Rapid Ngn2-induction of excitatory neurons from hiPSC-derived neural progenitor cells. *Methods*, 101, 113–124. <https://doi.org/10.1016/j.ymeth.2015.11.019>

Hockemeyer, D., Soldner, F., Cook, E. G., Gao, Q., Mitalipova, M., & Jaenisch, R. (2008). A Drug-Inducible System for Direct Reprogramming of Human Somatic Cells to Pluripotency. *Cell Stem Cell*, 3(3), 346–353. <https://doi.org/10.1016/j.stem.2008.08.014>

Hoffman, G. E., Hartley, B. J., Flaherty, E., Ladran, I., Gochman, P., Ruderfer, D. M., Stahl, E. A., Rapoport, J., Sklar, P., & Brennand, K. J. (2017). Transcriptional signatures of schizophrenia in hiPSC-derived NPCs and neurons are concordant with post-mortem adult brains. *Nature Communications*, 8(1), 2225. <https://doi.org/10.1038/s41467-017-02330-5>

Hulme, A. J., McArthur, J. R., Maksour, S., Miellet, S., Ooi, L., Adams, D. J., Finol-Urdaneta, R. K., & Dottori, M. (2020). Molecular and Functional Characterization of Neurogenin-2 Induced Human Sensory Neurons. *Frontiers in Cellular Neuroscience*, 14, 425. <https://doi.org/10.3389/fncel.2020.600895>

Huo, Y., Li, S., Liu, J., Li, X., & Luo, X.-J. (2019). Functional genomics reveal gene regulatory mechanisms underlying schizophrenia risk. *Nature Communications*, 10(1), 670. <https://doi.org/10.1038/s41467-019-08666-4>

Huschtscha, L. I., Napier, C. E., Noble, J. R., Bower, K., Au, A. Y. M., Campbell, H. G., Braithwaite, A. W., & Reddel, R. R. (2012). Enhanced isolation of fibroblasts from human skin explants. *BioTechniques*, 53(4), 239–244. <https://doi.org/10.2144/0000113939>

Hwang, Y., Kim, J., Shin, J. Y., Kim, J. I. I., Seo, J. S., Webster, M. J., Lee, D., & 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), 1–9. <https://doi.org/10.1038/tp.2013.94>

Inoue, F., & Ahituv, N. (2015). Decoding enhancers using massively parallel reporter assays. In *Genomics* (Vol. 106, Issue 3, pp. 159–164). Academic Press Inc. <https://doi.org/10.1016/j.ygeno.2015.06.005>

Inoue, F., Kircher, M., Martin, B., Cooper, G. M., Witten, D. M., McManus, M. T., Ahituv, N., & Shendure, J. (2017). A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Research*, 27(1), 38–52. <https://doi.org/10.1101/gr.212092.116>

Johnson, T. S., Xiang, S., Helm, B. R., Abrams, Z. B., Neidecker, P., Machiraju, R., Zhang, Y., Huang, K., & Zhang, J. (2020). Spatial cell type composition in normal and Alzheimer's human brains is revealed using integrated mouse and human single cell RNA sequencing. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-74917-w>

Jones, C., Watson, D., & Fone, K. (2011). Animal models of schizophrenia. In *British Journal of Pharmacology* (Vol. 164, Issue 4, pp. 1162–1194). Br J Pharmacol. <https://doi.org/10.1111/j.1476-5381.2011.01386.x>

Kaech, S., & Bunker, G. (2006). Culturing hippocampal neurons. *Nature Protocols*, 1(5), 2406–2415. <https://doi.org/10.1038/nprot.2006.356>

Kang, H. M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C. M., Gate, R. E., Mostafavi, S., Marson, A., Zaitlen, N., Criswell, L. A., & Ye, C. J. (2018). Multiplexed droplet single-cell RNA-

sequencing using natural genetic variation. *Nature Biotechnology*, 36(1), 89–94. <https://doi.org/10.1038/nbt.4042>

Kang, X., Yu, Q., Huang, Y., Song, B., Chen, Y., Gao, X., He, W., Sun, X., & Fan, Y. (2015). Effects of integrating and non-integrating reprogramming methods on copy number variation and genomic stability of human induced pluripotent stem cells. *PLoS ONE*, 10(7), 131128. <https://doi.org/10.1371/journal.pone.0131128>

Kathuria, A., Lopez-Lengowski, K., Watmuff, B., McPhie, D., Cohen, B. M., & Karmacharya, R. (2019). Synaptic deficits in iPSC-derived cortical interneurons in schizophrenia are mediated by NLGN2 and rescued by N-acetylcysteine. *Translational Psychiatry*, 9(1), 321. <https://doi.org/10.1038/s41398-019-0660-x>

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K., & Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications*, 10(1), 1–10. <https://doi.org/10.1038/s41467-019-09982-5>

Kety, S. S. (1987). The significance of genetic factors in the etiology of schizophrenia: Results from the national study of adoptees in Denmark. *Journal of Psychiatric Research*, 21(4), 423–429. [https://doi.org/10.1016/0022-3956\(87\)90089-6](https://doi.org/10.1016/0022-3956(87)90089-6)

Kety, S. S., Wender, P. H., Jacobsen, B., Ingraham, L. J., Janson, L., Faber, B., & Kinney, D. K. (1994). Mental Illness in the Biological and Adoptive Relatives of Schizophrenic Adoptees: Replication of the Copenhagen Study in the Rest of Denmark. *Archives of General Psychiatry*, 51(6), 442–455. <https://doi.org/10.1001/archpsyc.1994.03950060006001>

Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., Alston, J., Mikkelsen, T. S., & Kellis, M. (2013). Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Research*, 23(5), 800–811. <https://doi.org/10.1101/gr.144899.112>

Kiecker, C., & Lumsden, A. (2012). The role of organizers in patterning the nervous system. *Annual Review of Neuroscience*, 35(1), 347–367. <https://doi.org/10.1146/annurev-neuro-062111-150543>

Kim-Hellmuth, S., Bechheim, M., Pütz, B., Mohammadi, P., Nédélec, Y., Giangreco, N., Becker, J., Kaiser, V., Fricker, N., Beier, E., Boor, P., Castel, S. E., Nöthen, M. M., Barreiro, L. B., Pickrell, J. K., Müller-Myhsok, B., Lappalainen, T., Schumacher, J., & Hornung, V. (2017). Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nature Communications*, 8(1). <https://doi.org/10.1038/s41467-017-00366-1>

Kirkeby, A., Grealish, S., Wolf, D. A., Nelander, J., Wood, J., Lundblad, M., Lindvall, O., & Parmar, M. (2012). Generation of Regionally Specified Neural Progenitors and Functional Neurons from Human Embryonic Stem Cells under Defined Conditions. *Cell Reports*, 1(6), 703–714. <https://doi.org/10.1016/j.celrep.2012.04.009>

Klein, J. C., Agarwal, V., Inoue, F., Keith, A., Martin, B., Kircher, M., Ahituv, N., & Shendure, J. (2020). A systematic evaluation of the design and context dependencies of massively parallel reporter assays. *Nature Methods*, 17(11), 1083–1091. <https://doi.org/10.1038/s41592-020-0965-y>

Klemm, S. L., Shipony, Z., & Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics*, 20(4), 207–220. <https://doi.org/10.1038/s41576-018-0089-8>

Komarova, E. A., Neznanov, N., Komarov, P. G., Chernov, M. V., Wang, K., & Gudkov, A. V. (2003). P53 Inhibitor Pifithrin A Can Suppress Heat Shock and Glucocorticoid Signaling Pathways. *Journal of Biological Chemistry*, 278(18), 15465–15468. <https://doi.org/10.1074/jbc.C300011200>

Kreimer, A., Zeng, H., Edwards, M. D., Guo, Y., Tian, K., Shin, S., Welch, R., Wainberg, M., Mohan, R., Sinnott-Armstrong, N. A., Li, Y., Eraslan, G., Amin, T. Bin, Tewhey, R., Sabeti, P. C., Goke, J., Mueller, N. S., Kellis, M., Kundaje, A., ... Yosef, N. (2017). Predicting gene expression in massively parallel reporter assays: A comparative study. *Human Mutation*, 38(9). <https://doi.org/10.1002/humu.23197>

Kumasaka, N., Knights, A. J., & Gaffney, D. J. (2016). Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nature Genetics*, 48(2), 206–213. <https://doi.org/10.1038/ng.3467>

Lai, C.-Y., Scarr, E., Udawela, M., Everall, I., Chen, W. J., & Dean, B. (2016). Biomarkers in schizophrenia: A focus on blood based diagnostics and theranostics. *World Journal of Psychiatry*, 6(1), 102. <https://doi.org/10.5498/wjp.v6.i1.102>

Lam, M., Chen, C. Y., Li, Z., Martin, A. R., Bryois, J., Ma, X., Gaspar, H., Ikeda, M., Benyamin, B., Brown, B. C., Liu, R., Zhou, W., Guan, L., Kamatani, Y., Kim, S. W., Kubo, M., Kusumawardhani, A. A. A. A., Liu, C. M., Ma, H., ... Huang, H. (2019). Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nature Genetics*, 51(12), 1670–1678. <https://doi.org/10.1038/s41588-019-0512-x>

Lancaster, M. A., Renner, M., Martin, C. A., Wenzel, D., Bicknell, L. S., Hurles, M. E., Homfray, T., Penninger, J. M., Jackson, A. P., & Knoblich, J. A. (2013). Cerebral organoids model human brain development and microcephaly. *Nature*, 501(7467), 373–379. <https://doi.org/10.1038/nature12517>

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>

Leask, S. J. (2004). Environmental influences in schizophrenia: The known and the unknown. *Advances in Psychiatric Treatment*, 10(5), 323–330. <https://doi.org/10.1192/apt.10.5.323>

Lee, E. J., Lee, H. G., Park, S. H., Choi, E. Y., & Park, S. H. (2003). CD99 type II is a determining factor for the differentiation of primitive neuroectodermal cells. *Experimental and Molecular Medicine*, 35(5), 438–447. <https://doi.org/10.1038/emm.2003.57>

Li, C. L., Li, K. C., Wu, D., Chen, Y., Luo, H., Zhao, J. R., Wang, S. S., Sun, M. M., Lu, Y. J., Zhong, Y. Q., Hu, X. Y., Hou, R., Zhou, B. B., Bao, L., Xiao, H. S., & Zhang, X. (2016). Somatosensory neuron types identified by high-coverage single-cell RNA-sequencing and functional heterogeneity. *Cell Research*, 26(1), 83–102. <https://doi.org/10.1038/cr.2015.149>

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14),

1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>

Li, Y., Wang, R., Qiao, N., Peng, G., Zhang, K., Tang, K., Han, J. D. J., & Jing, N. (2017). Transcriptome analysis reveals determinant stages controlling human embryonic stem cell commitment to neuronal cells. *Journal of Biological Chemistry*, 292(48), 19590–19604. <https://doi.org/10.1074/jbc.M117.796383>

Li, Z., Chen, J., Yu, H., He, L., Xu, Y., Zhang, D., Yi, Q., Li, C., Li, X., Shen, J., Song, Z., Ji, W., Wang, M., Zhou, J., Chen, B., Liu, Y., Wang, J., Wang, P., Yang, P., ... Shi, Y. (2017). Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nature Genetics*, 49(11), 1576–1583. <https://doi.org/10.1038/ng.3973>

Lichtenstein, P., Yip, B. H., Björk, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., & Hultman, C. M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *The Lancet*, 373(9659), 234–239. [https://doi.org/10.1016/S0140-6736\(09\)60072-6](https://doi.org/10.1016/S0140-6736(09)60072-6)

Lin, H.-C., He, Z., Ebert, S., Schörnig, M., Santel, M., Weigert, A., Hevers, W., Kasri, N. N., Taverna, E., Camp, J. G., & Treutlein, B. (2020). D R A F T Ngn2 induces diverse neuronal lineages from human pluripotency. *BioRxiv*, 2020.11.19.389445. <https://doi.org/10.1101/2020.11.19.389445>

Liu, D., & Xu, Y. (2011). P53, oxidative stress, and aging. In *Antioxidants and Redox Signaling* (Vol. 15, Issue 6, pp. 1669–1678). Mary Ann Liebert, Inc. <https://doi.org/10.1089/ars.2010.3644>

Liu, Y.-N., Lu, S.-Y., & Yao, J. (2017). Application of induced pluripotent stem cells to understand neurobiological basis of bipolar disorder and schizophrenia. *Psychiatry and Clinical Neurosciences*, 71(9), 579–599. <https://doi.org/10.1111/pcn.12528>

Liu, Y., Yu, S., Dhiman, V. K., Brunetti, T., Eckart, H., & White, K. P. (2017). Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biology*, 18(1), 219. <https://doi.org/10.1186/s13059-017-1345-5>

Loo, L., Simon, J. M., Xing, L., McCoy, E. S., Niehaus, J. K., Guo, J., Anton, E. S., & Zylka, M. J. (2019). Single-cell transcriptomic analysis of mouse neocortical development. *Nature Communications*, 10(1). <https://doi.org/10.1038/s41467-018-08079-9>

Lu, X., Chen, X., Forney, C., Donmez, O., Miller, D., Parameswaran, S., Hong, T., Huang, Y., Pujato, M., Cazares, T., Miraldi, E. R., Ray, J. P., de Boer, C. G., Harley, J. B., Weirauch, M. T., & Kottyan, L. C. (2021). Global discovery of lupus genetic risk variant allelic enhancer activity. *Nature Communications*, 12(1), 1–13. <https://doi.org/10.1038/s41467-021-21854-5>

Luo, C., Lancaster, M. A., Castanon, R., Nery, J. R., Knoblich, J. A., & Ecker, J. R. (2016). Cerebral Organoids Recapitulate Epigenomic Signatures of the Human Fetal Brain. *Cell Reports*, 17(12), 3369–3384. <https://doi.org/dx.doi.org/10.1016/j.celrep.2016.12.001>

Ma, L., Semick, S. A., Chen, Q., Li, C., Tao, R., Price, A. J., Shin, J. H., Jia, Y., Brandon, N. J., Cross, A. J., Hyde, T. M., Kleinman, J. E., Jaffe, A. E., Weinberger, D. R., & Straub, R. E. (2020). Schizophrenia risk variants influence multiple classes of transcripts of sorting nexin 19 (SNX19). *Molecular Psychiatry*, 25(4), 831–843. <https://doi.org/10.1038/s41380-018-0293-0>

Mali, P., Chou, B. K., Yen, J., Ye, Z., Zou, J., Dowey, S., Brodsky, R. A., Ohm, J. E., Yu, W., Baylin, S. B., Yusa, K., Bradley, A., Meyers, D. J., Mukherjee, C., Cole, P. A., & Cheng, L. (2010). Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. *Stem Cells*, 28(4), 713–720. <https://doi.org/10.1002/stem.402>

Malik, A. N., Vierbuchen, T., Hemberg, M., Rubin, A. A., Ling, E., Couch, C. H., Stroud, H., Spiegel, I., Farh, K. K. H., Harmin, D. A., & Greenberg, M. E. (2014). Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nature Neuroscience*, 17(10), 1330–1339. <https://doi.org/10.1038/nn.3808>

Mariani, J., Simonini, M. V., Palejev, D., Tomasini, L., Coppola, G., Szekely, A. M., Horvath, T. L., & Vaccarino, F. M. (2012). Modeling human cortical development in vitro using induced pluripotent stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 109(31), 12770–12775. <https://doi.org/10.1073/pnas.1202944109>

Marión, R. M., Strati, K., Li, H., Murga, M., Blanco, R., Ortega, S., Fernandez-Capetillo, O., Serrano, M., & Blasco, M. A. (2009). A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature*, 460(7259), 1149–1153. <https://doi.org/10.1038/nature08287>

Marmigère, F., & Ernfors, P. (2007). Specification and connectivity of neuronal subtypes in the sensory lineage. In *Nature Reviews Neuroscience* (Vol. 8, Issue 2, pp. 114–127). Nature Publishing Group. <https://doi.org/10.1038/nrn2057>

Maroof, A. M., Keros, S., Tyson, J. A., Ying, S.-W., Ganat, Y. M., Merkle, F. T., Liu, B., Goulburn, A., Stanley, E. G., Elefanty, A. G., Widmer, H. R., Eggan, K., Goldstein, P. A., Anderson, S. A., & Studer, L. (2013). Directed differentiation and functional maturation of cortical interneurons from human embryonic stem cells. *Cell Stem Cell*, 12(5), 559–572. <https://doi.org/10.1016/j.stem.2013.04.008>

Martinez-Ara, M., Comoglio, F., van Arensbergen, J., & van Steensel, B. (2022). Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome. *Molecular Cell*, 2021.10.21.465269. <https://doi.org/10.1016/j.molcel.2022.04.009>

Marwaha, S., Johnson, S., Bebbington, P., Stafford, M., Angermeyer, M. C., Brugha, T., Azorin, J. M., Kilian, R., Hansen, K., & Toumi, M. (2007). Rates and correlates of employment in people with schizophrenia in the UK, France and Germany. *British Journal of Psychiatry*, 191(JULY), 30–37. <https://doi.org/10.1192/bjp.bp.105.020982>

McGrath, J., Saha, S., Chant, D., & Welham, J. (2008). Schizophrenia: A Concise Overview of Incidence, Prevalence, and Mortality. *Epidemiologic Reviews*, 30(1), 67–76. <https://doi.org/10.1093/epirev/mxn001>

McGuffin, P., Asherson, P., Owen, M., & Farmer, A. (1994). The strength of the genetic effect. Is there room for an environmental influence in the aetiology of schizophrenia? *British Journal of Psychiatry*, 164(MAY), 593–599.

<https://doi.org/10.1192/bjp.164.5.593>

McPhie, D. L., Nehme, R., Ravichandran, C., Babb, S. M., Ghosh, S. D., Staskus, A., Kalinowski, A., Kaur, R., Douvaras, P., Du, F., Ongur, D., Fossati, V., Eggan, K., & Cohen, B. M. (2018). Oligodendrocyte differentiation of induced pluripotent stem cells derived from subjects with schizophrenia implicate abnormalities in development. *Translational Psychiatry*, 8(1), 230. <https://doi.org/10.1038/s41398-018-0284-6>

Meijer, M., Rehbach, K., Brunner, J. W., Classen, J. A., Lammertse, H. C. A., van Linge, L. A., Schut, D., Krutenko, T., Hebisch, M., Cornelisse, L. N., Sullivan, P. F., Peitz, M., Toonen, R. F., Brüstle, O., & Verhage, M. (2019). A Single-Cell Model for Synaptic Transmission and Plasticity in Human iPSC-Derived Neurons. *Cell Reports*, 27(7), 2199-2211.e6. <https://doi.org/10.1016/j.celrep.2019.04.058>

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnrke, A., Callan, C. G., Kinney, J. B., Kellis, M., Lander, E. S., & Mikkelsen, T. S. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nature Biotechnology*, 30(3), 271-277. <https://doi.org/10.1038/nbt.2137>

Melnikov, A., Zhang, X., Rogov, P., Wang, L., & Mikkelsen, T. S. (2014). Massively parallel reporter assays in cultured mammalian cells. *Journal of Visualized Experiments*, 90. <https://doi.org/10.3791/51719>

Messias, E. L., Chen, C. Y., & Eaton, W. W. (2007). Epidemiology of Schizophrenia: Review of Findings and Myths. *Psychiatric Clinics of North America*, 30(3), 323-338. <https://doi.org/10.1016/j.psc.2007.04.007>

Meyer-Lindenberg, A. (2010). Imagenología de la expresión genética en la esquizofrenia. *Dialogues in Clinical Neuroscience*, 12(4), 449-456. [www.dialogues-cns.org](http://www.dialogues-cns.org)

Mill, J., Tang, T., Kaminsky, Z., Khare, T., Yazdanpanah, S., Bouchard, L., Jia, P., Assadzadeh, A., Flanagan, J., Schumacher, A., Wang, S. C., & Petronis, A. (2008). Epigenomic Profiling Reveals DNA-Methylation Changes Associated with Major Psychosis. *American Journal of Human Genetics*, 82(3), 696-711. <https://doi.org/10.1016/j.ajhg.2008.01.008>

Miller, J. D., Ganat, Y. M., Kishinevsky, S., Bowman, R. L., Liu, B., Tu, E. Y., Mandal, P. K., Vera, E., Shim, J. W., Kriks, S., Taldone, T., Fusaki, N., Tomishima, M. J., Krainc, D., Milner, T. A., Rossi, D. J., & Studer, L. (2013). Human iPSC-based modeling of late-onset disease via progerin-induced aging. *Cell Stem Cell*, 13(6), 691-705. <https://doi.org/10.1016/j.stem.2013.11.006>

Mohamed, Z. I., Tee, S. F., Chow, T. J., Loh, S. Y., Yong, H. S., Sen, Bakar, A. K. A., & Tang, P. Y. (2019). Functional characterization of two variants in the 3'-untranslated region (UTR) of transcription factor 4 gene and their association with schizophrenia in sib-pairs from multiplex families. *Asian Journal of Psychiatry*, 40, 76-81. <https://doi.org/10.1016/j.ajp.2019.02.001>

Mongiat, L. A., Espósito, M. S., Lombardi, G., & Schinder, A. F. (2009). Reliable activation of immature neurons in the adult hippocampus. *PLoS ONE*, 4(4), 5320. <https://doi.org/10.1371/journal.pone.0005320>

Mora, A., Sandve, G. K., Gabrielsen, O. S., & Eskeland, R. (2016). In the loop: promoter-enhancer interactions and bioinformatics. In *Briefings in bioinformatics* (Vol. 17, Issue 6, pp. 980-995). Oxford University Press. <https://doi.org/10.1093/bib/bbv097>

Mortensen, P. B., Pedersen, C. B., Westergaard, T., Wohlfahrt, J., Ewald, H., Mors, O., Andersen, P. K., & Melbye, M. (1999). Effects of Family History and Place and Season of Birth on the Risk of Schizophrenia. *New England Journal of Medicine*, 340(8), 603-608. <https://doi.org/10.1056/nejm199902253400803>

Movva, R., Greenside, P., Marinov, G. K., Nair, S., Shrikumar, A., & Kundaje, A. (2019). Deciphering regulatory DNA sequences and noncoding genetic variants using neural network models of massively parallel reporter assays. *PLoS ONE*, 14(6). <https://doi.org/10.1371/journal.pone.0218073>

Muerdter, F., Boryn, L. M., Woodfin, A. R., Neumayr, C., Rath, M., Zabidi, M. A., Pagani, M., Haberle, V., Kazmar, T., Catarino, R. R., Schernhuber, K., Arnold, C. D., & Stark, A. (2018). Resolving systematic errors in widely used enhancer activity assays in human cells. *Nature Methods*, 15(2), 141-149. <https://doi.org/10.1038/nmeth.4534>

Murata, T., Ohnishi, H., Okazawa, H., Murata, Y., Kusakari, S., Hayashi, Y., Miyashita, M., Itoh, H., Oldenborg, P. A., Furuya, N., & Matozaki, T. (2006). CD47 promotes neuronal development through Src- and FRG/Vav2-mediated activation of Rac and Cdc42. *Journal of Neuroscience*, 26(48), 12397-12407. <https://doi.org/10.1523/JNEUROSCI.3981-06.2006>

Muratore, C. R., Srikanth, P., Callahan, D. G., & Young-Pearse, T. L. (2014). Comparison and optimization of hiPSC forebrain cortical differentiation protocols. *PLoS ONE*, 9(8), e105807. <https://doi.org/10.1371/journal.pone.0105807>

Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N. E., Ahfeldt, T., Sachs, K. V., Li, X., Li, H., Kuperwasser, N., Ruda, V. M., Pirruccello, J. P., Muchmore, B., Prokunina-Olsson, L., Hall, J. L., Schadt, E. E., Morales, C. R., Lund-Katz, S., Phillips, M. C., Wong, J., ... Rader, D. J. (2010). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*, 466(7307), 714-719. <https://doi.org/10.1038/nature09266>

Myint, L., Wang, R., Boukas, L., Hansen, K. D., Goff, L. A., & Avramopoulos, D. (2020). A screen of 1,049 schizophrenia and 30 Alzheimer's-associated variants for regulatory potential. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics : The Official Publication of the International Society of Psychiatric Genetics*, 183(1), 61-73. <https://doi.org/10.1002/ajmg.b.32761>

Nasser, J., Bergman, D. T., Fulco, C. P., Guckelberger, P., Doughty, B. R., Patwardhan, T. A., Jones, T. R., Nguyen, T. H., Ulirsch, J. C., Lekschas, F., Mualim, K., Natri, H. M., Weeks, E. M., Munson, G., Kane, M., Kang, H. Y., Cui, A., Ray, J. P., Eisenhaure, T. M., ... Engreitz, J. M. (2021). Genome-wide enhancer maps link risk variants to disease genes. *Nature*, 593(7858), 238-243. <https://doi.org/10.1038/s41586-021-03446-x>

Nehme, R., Zuccaro, E., Ghosh, S. D., Li, C., Sherwood, J. L., Pietilainen, O., Barrett, L. E., Limone, F., Worringer, K. A.,

Kommineni, S., Zang, Y., Cacchiarelli, D., Meissner, A., Adolfsson, R., Haggarty, S., Madison, J., Muller, M., Arlotta, P., Fu, Z., ... Eggan, K. (2018). Combining NGN2 Programming with Developmental Patterning Generates Human Excitatory Neurons with NMDAR-Mediated Synaptic Transmission. *Cell Reports*, 23(8), 2509–2523. <https://doi.org/10.1016/j.celrep.2018.04.066>

Ng, M. Y. M., Levinson, D. F., Faraone, S. V., Suarez, B. K., Delisi, L. E., Arinami, T., Riley, B., Paunio, T., Pulver, A. E., Irmansyah, Holmans, P. A., Escamilla, M., Wildenauer, D. B., Williams, N. M., Laurent, C., Mowry, B. J., Brzustowicz, L. M., Maziade, M., Sklar, P., ... Lewis, C. M. (2009). Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Molecular Psychiatry*, 14(8), 774–785. <https://doi.org/10.1038/mp.2008.135>

Nica, A. C., Parts, L., Glass, D., Nisbet, J., Barrett, A., Sekowska, M., Travers, M., Potter, S., Grundberg, E., Small, K., Hedman, Å. K., Bataille, V., Bell, J., Surdulescu, G., Dimas, A. S., Ingle, C., Nestle, F. O., Meglio, P., Min, J. L., ... Spector, T. D. (2011). The architecture of gene regulatory variation across multiple human tissues: The muTHER study. *PLoS Genetics*, 7(2), 1002003. <https://doi.org/10.1371/journal.pgen.1002003>

Nickolls, A. R., Lee, M. M., Espinoza, D. F., Szczot, M., Lam, R. M., Wang, Q., Beers, J., Zou, J., Nguyen, M. Q., Solinski, H. J., AlJanahi, A. A., Johnson, K. R., Ward, M. E., Chesler, A. T., & Bönnemann, C. G. (2020). Transcriptional Programming of Human Mechanosensory Neuron Subtypes from Pluripotent Stem Cells. *Cell Reports*, 30(3), 932–946.e7. <https://doi.org/10.1016/j.celrep.2019.12.062>

Ostermann, L., Ladewig, J., Müller, F. J., Kesavan, J., Tailor, J., Smith, A., Brüstle, O., & Koch, P. (2019). In Vitro Recapitulation of Developmental Transitions in Human Neural Stem Cells. *Stem Cells*, 37(11), 1429–1440. <https://doi.org/10.1002/stem.3065>

Page, G. P., George, V., Go, R. C., Page, P. Z., & Allison, D. B. (2003). "Are We There Yet?": Deciding When One Has Demonstrated Specific Genetic Causation in Complex Diseases and Quantitative Traits. In *American Journal of Human Genetics* (Vol. 73, Issue 4, pp. 711–719). University of Chicago Press. <https://doi.org/10.1086/378900>

Page, S. C., Hamersky, G. R., Gallo, R. A., Rannals, M. D., Calcaterra, N. E., Campbell, M. N., Mayfield, B., Briley, A., Phan, B. N., Jaffe, A. E., & Maher, B. J. (2018). The schizophrenia-and autism-associated gene, transcription factor 4 regulates the columnar distribution of layer 2/3 prefrontal pyramidal neurons in an activity-dependent manner. *Molecular Psychiatry*, 23(2), 304–315. <https://doi.org/10.1038/mp.2017.37>

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., ... Walters, J. T. R. R. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature Genetics*, 50(3), 381–389. <https://doi.org/10.1038/s41588-018-0059-2>

Patwardhan, R. P., Hiatt, J. B., Witten, D. M., Kim, M. J., Smith, R. P., May, D., Lee, C., Andrie, J. M., Lee, S. I., Cooper, G. M., Ahituv, N., Pennacchio, L. A., & Shendure, J. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. *Nature Biotechnology*, 30(3), 265–270. <https://doi.org/10.1038/nbt.2136>

Perkovic, M. N., Erjavec, G. N., Strac, D. S., Uzun, S., Kozumplik, O., & Pivac, N. (2017). Theranostic biomarkers for schizophrenia. In *International Journal of Molecular Sciences* (Vol. 18, Issue 4). MDPI AG. <https://doi.org/10.3390/ijms18040733>

Pollen, A. A., Nowakowski, T. J., Chen, J., Retallack, H., Sandoval-Espinosa, C., Nicholas, C. R., Shuga, J., Liu, S. J., Oldham, M. C., Diaz, A., Lim, D. A., Leyrat, A. A., West, J. A., & Kriegstein, A. R. (2015). Molecular Identity of Human Outer Radial Glia during Cortical Development. *Cell*, 163(1), 55–67. <https://doi.org/10.1016/j.cell.2015.09.004>

Powchik, P., Davidson, M., Haroutunian, V., Gabriel, S. M., Purohit, D. P., Perl, D. P., Harvey, P. D., & Davis, K. L. (1998). Postmortem studies in schizophrenia. *Schizophrenia Bulletin*, 24(3), 325–341. <https://doi.org/10.1093/oxfordjournals.schbul.a033330>

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5), 1173–1183. <https://doi.org/10.1016/j.cell.2013.02.022>

Qi, Y., Zhang, X.-J., Renier, N., Wu, Z., Atkin, T., Sun, Z., Ozair, M. Z., Tchieu, J., Zimmer, B., Fattah, F., Ganat, Y., Azevedo, R., Zeltner, N., Brivanlou, A. H., Karayiorgou, M., Gogos, J., Tomishima, M., Tessier-Lavigne, M., Shi, S.-H., & Studer, L. (2017). Combined small-molecule inhibition accelerates the derivation of functional cortical neurons from human pluripotent stem cells. *Nature Biotechnology*, 35(2), 154–163. <https://doi.org/10.1038/nbt.3777>

Qiu, J., McQueen, J., Bilican, B., Dando, O., Magnani, D., Punovuori, K., Selvaraj, B. T., Livesey, M., Hagh, G., Heron, S., Burr, K., Patani, R., Rajan, R., Sheppard, O., Kind, P. C., Simpson, T. I., Tybulewicz, V. L. J., Wyllie, D. J. A., Fisher, E. M. C., ... Hardingham, G. E. (2016). Evidence for evolutionary divergence of activity-dependent gene expression in developing neurons. *ELife*, 5(OCTOBER2016). <https://doi.org/10.7554/eLife.20337>

Raab, S., Klingenstein, M., Liebau, S., & Linta, L. (2014). A Comparative View on Human Somatic Cell Sources for iPSC Generation. *Stem Cells International*, 2014, 1–12. <https://doi.org/10.1155/2014/768391>

Rajarajan, P., Borrmann, T., Liao, W., Schröde, N., Flaherty, E., Casiño, C., Powell, S., Yashaswini, C., LaMarca, E. A., Kassim, B., Javidfar, B., Espeso-Gil, S., Li, A., Won, H., Geschwind, D. H., Ho, S. M., MacDonald, M., Hoffman, G. E., Roussos, P., ... Akbarian, S. (2018). Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. *Science*, 362(6420), 4311. <https://doi.org/10.1126/science.aat4311>

Ramirez, J. M., Tryba, A. K., & Peña, F. (2004). Pacemaker neurons and neuronal networks: An integrative view. *Current Opinion in Neurobiology*, 14(6), 665–674. <https://doi.org/10.1016/j.conb.2004.10.011>

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9

system. *Nature Protocols*, 8(11), 2281–2308. <https://doi.org/10.1038/nprot.2013.143>

Rasmussen, M. A., Holst, B., Tümer, Z., Johnsen, M. G., Zhou, S., Stummann, T. C., Hyttel, P., & Clausen, C. (2014). Transient p53 suppression increases reprogramming of human fibroblasts without affecting apoptosis and DNA damage. *Stem Cell Reports*, 3(3), 404–413. <https://doi.org/10.1016/j.stemcr.2014.07.006>

Rees, E., Walters, J. T. R., 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., & Kirov, G. (2014). Analysis of copy number variations at 15 schizophrenia-associated loci. *British Journal of Psychiatry*, 204(2), 108–114. <https://doi.org/10.1192/bj.p.113.131052>

Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., Lin, D. Y., Duan, J., Ophoff, R. A., Andreassen, O. A., Scolnick, E., Cichon, S., St. Clair, D., Corvin, A., Gurling, H., Werge, T., Rujescu, D., Blackwood, D. H. R., Pato, C. N., ... Gejman, P. V. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics*, 43(10), 969–978. <https://doi.org/10.1038/ng.940>

Rosa, F., Dhingra, A., Uysal, B., Mendis, G. D. C., Loeffler, H., Elsen, G., Mueller, S., Schwarz, N., Castillo-Lizardo, M., Cuddy, C., Becker, F., Heutink, P., Reid, C. A., Petrou, S., Lerche, H., & Maljevic, S. (2020). In Vitro Differentiated Human Stem Cell-Derived Neurons Reproduce Synaptic Synchronicity Arising during Neurodevelopment. *Stem Cell Reports*, 15(1), 22–37. <https://doi.org/10.1016/j.stemcr.2020.05.015>

Roussos, P., Guennewig, B., Kaczorowski, D. C., Barry, G., & Brennand, K. J. (2016). Activity-Dependent Changes in Gene Expression in Schizophrenia Human-Induced Pluripotent Stem Cell Neurons. *JAMA Psychiatry*, 73(11), 1180. <https://doi.org/10.1001/jamapsychiatry.2016.2575>

Sabaie, H., Moghaddam, M. M., Moghaddam, M. M., Ahangar, N. K., Asadi, M. R., Hussen, B. M., Taheri, M., & Rezazadeh, M. (2021). Bioinformatics analysis of long non-coding RNA-associated competing endogenous RNA network in schizophrenia. *Scientific Reports*, 11(1), 1–13. <https://doi.org/10.1038/s41598-021-03993-3>

Sack, L. M., Davoli, T., Xu, Q., Li, M. Z., & Edledge, S. J. (2016). Sources of error in mammalian genetic screens. *G3: Genes, Genomes, Genetics*, 6(9), 2781–2790. <https://doi.org/10.1534/g3.116.030973>

Sankaran, V. G., Ludwig, L. S., Sicinska, E., Xu, J., Bauer, D. E., Eng, J. C., Patterson, H. C., Metcalf, R. A., Natkunam, Y., Orkin, S. H., Sicinski, P., Lander, E. S., & Lodish, H. F. (2012). Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number. *Genes and Development*, 26(18), 2075–2087. <https://doi.org/10.1101/gad.197020.112>

Santos, R., Vadodaria, K. C., Jaeger, B. N., Mei, A., Lefcochilos-Fogelquist, S., Mendes, A. P. D., Erikson, G., Shokhirev, M., Randolph-Moore, L., Fredlender, C., Dave, S., Oefner, R., Fitzpatrick, C., Pena, M., Barron, J. J., Ku, M., Denli, A. M., Kerman, B. E., Charnay, P., ... Gage, F. H. (2017). Differentiation of Inflammation-Responsive Astrocytes from Glial Progenitors Generated from Human Induced Pluripotent Stem Cells. *Stem Cell Reports*, 8(6), 1757–1769. <https://doi.org/10.1016/j.stemcr.2017.05.011>

Sarnat, H. B. (2013). Clinical neuropathology practice guide 5-2013: Markers of neuronal maturation. *Clinical Neuropathology*, 32(5), 340–369. <https://doi.org/10.5414/NP300638>

Satir, T. M., Nazir, F. H., Vizlin-Hodzic, D., Hardselius, E., Blennow, K., Wray, S., Zetterberg, H., Agholme, L., & Bergström, P. (2020). Accelerated neuronal and synaptic maturation by BrainPhys medium increases Aβ secretion and alters Aβ peptide ratios from iPSC-derived cortical neurons. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-57516-7>

Schaid, D. J., Chen, W., & Larson, N. B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature Reviews Genetics*, 19(8), 491–504. <https://doi.org/10.1038/s41576-018-0016-z>

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. In *Nature Methods* (Vol. 9, Issue 7, pp. 676–682). NIH Public Access. <https://doi.org/10.1038/nmeth.2019>

Schizophrenia Working Group of the Psychiatric Genomics Consortium, Ripke, S., Neale, B. M., Corvin, A., Walters, J. T. R., 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., ... O'Donovan, M. C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511(7510), 421–427. <https://doi.org/10.1038/nature13595>

Schlaeger, T. M., Daheron, L., Brickler, T. R., Entwistle, S., Chan, K., Cianci, A., DeVine, A., Ettenger, A., Fitzgerald, K., Godfrey, M., Gupta, D., McPherson, J., Malwadkar, P., Gupta, M., Bell, B., Doi, A., Jung, N., Li, X., Lynes, M. S., ... Daley, G. Q. (2015). A comparison of non-integrating reprogramming methods. *Nature Biotechnology*, 33(1), 58–63. <https://doi.org/10.1038/nbt.3070>

Schneider, K. (1959). *Clinical psychopathology*. Grune & Stratton. <https://www.worldcat.org/title/clinical-psychopathology/oclc/576893>

Schröde, N., Ho, S.-M., Yamamoto, K., Dobbyn, A., Huckins, L., Matos, M. R., Cheng, E., Deans, P. J. M., Flaherty, E., Barreto, N., Topol, A., Alganem, K., Abadali, S., Gregory, J., Hoelzli, E., Phatnani, H., Singh, V., Girish, D., Aronow, B., ... Brennand, K. J. (2019). Synergistic effects of common schizophrenia risk variants. *Nature Genetics*, 51(10), 1475–1485. <https://doi.org/10.1038/s41588-019-0497-5>

Schuurmans, C., Armant, O., Nieto, M., Stenman, J. M., Britz, O., Klenin, N., Brown, C., Langevin, L. M., Seibt, J., Tang, H., Cunningham, J. M., Dyck, R., Walsh, C., Campbell, K., Polleux, F., & Guillemot, F. (2004). Sequential phases of cortical specification involve neurogenin-dependent and -independent pathways. *EMBO Journal*, 23(14), 2892–2902.

https://doi.org/10.1038/sj.emboj.7600278

Schwartzentruber, J., Foskolou, S., Kilpinen, H., Rodrigues, J., Alasoo, K., Knights, A. J., Patel, M., Goncalves, A., Ferreira, R., Benn, C. L., Wilbrey, A., Bictash, M., Impey, E., Cao, L., Lainez, S., Loucif, A. J., Whiting, P. J., Gutteridge, A., & Gaffney, D. J. (2018). Molecular and functional variation in iPSC-derived sensory neurons. *Nature Genetics*, 50(1), 54–61. <https://doi.org/10.1038/s41588-017-0005-8>

Sepp, M., Kannike, K., Eesmaa, A., Urb, M., & Timmus, T. (2011). Functional diversity of human basic helix-loop-helix transcription factor TCF4 isoforms generated by alternative 5' exon usage and splicing. *PLoS ONE*, 6(7). <https://doi.org/10.1371/journal.pone.0022138>

Seto, E., & Yoshida, M. (2014). Erasers of histone acetylation: The histone deacetylase enzymes. *Cold Spring Harbor Perspectives in Biology*, 6(4). <https://doi.org/10.1101/cshperspect.a018713>

Shao, Z., Noh, H., Bin Kim, W., Ni, P., Nguyen, C., Cote, S. E., Noyes, E., Zhao, J., Parsons, T., Park, J. M., Zheng, K., Park, J. J., Coyle, J. T., Weinberger, D. R., Straub, R. E., Berman, K. F., Apud, J., Ongur, D., Cohen, B. M., ... Chung, S. (2019). Dysregulated protocadherin-pathway activity as an intrinsic defect in induced pluripotent stem cell-derived cortical interneurons from subjects with schizophrenia. *Nature Neuroscience*, 22(2), 229–242. <https://doi.org/10.1038/s41593-018-0313-z>

Sheng, M., & Greenberg, M. E. (1990). The regulation and function of c-fos and other immediate early genes in the nervous system. *Neuron*, 4(4), 477–485. [https://doi.org/10.1016/0896-6273\(90\)90106-P](https://doi.org/10.1016/0896-6273(90)90106-P)

Sher, L., & Kahn, R. S. (2019). Suicide in Schizophrenia: An Educational Overview. *Medicina (Kaunas, Lithuania)*, 55(7). <https://doi.org/10.3390/medicina55070361>

Sheridan, S. D., Surampudi, V., & Rao, R. R. (2012). Analysis of embryoid bodies derived from human induced pluripotent stem cells as a means to assess pluripotency. *Stem Cells International*, 2012. <https://doi.org/10.1155/2012/738910>

Shershov, J. C., Savodnik, I., Kety, S., Lidz, T., Klerman, G. L., Creese, I., Snyder, S. H., Hollister, L., Borus, J. F., Hatow, E., Schwartz, D. P., Mosher, L. R., & Menn, A. Z. (1978). *Schizophrenia: Science and Practice*. Harvard University Press.

Shi, Y., Kirwan, P., Smith, J., Robinson, H. P. C., & Livesey, F. J. (2012). Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. *Nature Neuroscience*, 15(3), 477–486. <https://doi.org/10.1038/nn.3041>

Shin, J., Berg, D. A., Zhu, Y., Shin, J. Y., Song, J., Bonaguidi, M. A., Enikolopov, G., Nauen, D. W., Christian, K. M., Ming, G. L., & Song, H. (2015). Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. *Cell Stem Cell*, 17(3), 360–372. <https://doi.org/10.1016/j.stem.2015.07.013>

Siggers, T., & Gordân, R. (2014). Protein-DNA binding: Complexities and multi-protein codes. *Nucleic Acids Research*, 42(4), 2099–2111. <https://doi.org/10.1093/nar/gkt1112>

Spicuglia, S., & Vanhille, L. (2012). Chromatin signatures of active enhancers. *Nucleus*, 3(2), 126–131. <https://doi.org/10.4161/nucl.19232>

Spike, B. T., & Wahl, G. M. (2011). P53, stem cells, and reprogramming: Tumor suppression beyond guarding the genome. *Genes and Cancer*, 2(4), 404–419. <https://doi.org/10.1177/1947601911410224>

Srinivasan, K., Leone, D. P., Bateson, R. K., Dobreva, G., Kohwi, Y., Kohwi-Shigematsu, T., Grosschedl, R., & McConnell, S. K. (2012). A network of genetic repression and derepression specifies projection fates in the developing neocortex. *Proceedings of the National Academy of Sciences of the United States of America*, 109(47), 19071–19078. <https://doi.org/10.1073/pnas.1216793109>

Stadtfeld, M., Brennand, K., & Hochedlinger, K. (2008). Reprogramming of Pancreatic  $\beta$  Cells into Induced Pluripotent Stem Cells. *Current Biology*, 18(12), 890–894. <https://doi.org/10.1016/j.cub.2008.05.010>

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, 177(7), 1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>

Sullivan, P. F., Kendler, K. S., & Neale, M. C. (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Archives of General Psychiatry*, 60(12), 1187–1192. <https://doi.org/10.1001/archpsyc.60.12.1187>

Sun, Z., Williams, D. J., Xu, B., & Gogos, J. A. (2018). Altered function and maturation of primary cortical neurons from a 22q11.2 deletion mouse model of schizophrenia. *Translational Psychiatry*, 8(1), 85. <https://doi.org/10.1038/s41398-018-0132-8>

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, 131(5), 861–872. <https://doi.org/10.1016/j.cell.2007.11.019>

Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126(4), 663–676. <https://doi.org/10.1016/j.cell.2006.07.024>

Tang, X., Zhou, L., Wagner, A. M., Marchetto, M. C. N., Muotri, A. R., Gage, F. H., & Chen, G. (2013). Astroglial cells regulate the developmental timeline of human neurons differentiated from induced pluripotent stem cells. *Stem Cell Research*, 11(2), 743–757. <https://doi.org/10.1016/j.scr.2013.05.002>

Tansey, K. E., & Hill, M. J. (2018). Enrichment of schizophrenia heritability in both neuronal and glia cell regulatory elements. *Translational Psychiatry*, 8(1), 7. <https://doi.org/10.1038/s41398-017-0053-y>

Tanskanen, A., Tiihonen, J., & Taipale, H. (2018). Mortality in schizophrenia: 30-year nationwide follow-up study. *Acta Psychiatrica Scandinavica*, 138(6), 492–499. <https://doi.org/10.1111/acps.12913>

TCW, J., Wang, M., Pimenova, A. A., Bowles, K. R., Hartley, B. J., Lacin, E., Machlovi, S. I., Abdelaal, R., Karch, C. M., Phatnani, H., Slesinger, P. A., Zhang, B., Goate, A. M., & Brennand, K. J. (2017). An Efficient Platform for Astrocyte Differentiation from Human Induced Pluripotent Stem Cells. *Stem Cell Reports*, 9(2), 600–614.

<https://doi.org/10.1016/j.stemcr.2017.06.018>

Tewhey, R., Kotliar, D., Park, D. S., Liu, B., Winnicki, S., Reilly, S. K., Andersen, K. G., Mikkelsen, T. S., Lander, E. S., Schaffner, S. F., & Sabeti, P. C. (2016). Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell*, 165(6), 1519–1529. <https://doi.org/10.1016/j.cell.2016.04.027>

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178–192. <https://doi.org/10.1093/bib/bbs017>

Tian, R., Gachechiladze, M. A., Ludwig, C. H., Laurie, M. T., Hong, J. Y., Nathaniel, D., Prabhu, A. V., Fernandopulle, M. S., Patel, R., Abshari, M., Ward, M. E., & Kampmann, M. (2019). CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron*, 104(2), 239–255.e12. <https://doi.org/10.1016/j.neuron.2019.07.014>

Trubetskoy, V., Pardiñas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T. B., Bryois, J., Chen, C.-Y., Dennis, 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., ... Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*, 604(7906), 502–508. <https://doi.org/10.1038/s41586-022-04434-5>

Tyssowski, K. M., DeStefino, N. R., Cho, J.-H. H., Dunn, C. J., Poston, R. G., Carty, C. E., Jones, R. D., Chang, S. M., Romeo, P., Wurzelmann, M. K., Ward, J. M., Andermann, M. L., Saha, R. N., Dudek, S. M., & Gray, J. M. (2018). Different Neuronal Activity Patterns Induce Different Gene Expression Programs. *Neuron*, 98(3), 530–546.e11. <https://doi.org/10.1016/j.neuron.2018.04.001>

Ulirsch, J. C., Nandakumar, S. K., Wang, L., Giani, F. C., Zhang, X., Rogov, P., Melnikov, A., McDonel, P., Do, R., Mikkelsen, T. S., & Sankaran, V. G. (2016). Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell*, 165(6), 1530–1545. <https://doi.org/10.1016/j.cell.2016.04.048>

van Arensbergen, J., Pagie, L., FitzPatrick, V. D., de Haas, M., Baltissen, M. P., Comoglio, F., van der Weide, R. H., Teunissen, H., Vösa, U., Franke, L., de Wit, E., Vermeulen, M., Bussemaker, H. J., & van Steensel, B. (2019). High-throughput identification of human SNPs affecting regulatory element activity. *Nature Genetics*, 51(7), 1160–1169. <https://doi.org/10.1038/s41588-019-0455-2>

Vanhille, L., Griffon, A., Maqbool, M. A., Zacarias-Cabeza, J., Dao, L. T. M., Fernandez, N., Ballester, B., Andrau, J. C., & Spicuglia, S. (2015). High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nature Communications*, 6(1), 1–10. <https://doi.org/10.1038/ncomms7905>

Velasco, S., Ibrahim, M. M., Kakumanu, A., Garippler, G., Aydin, B., Al-Sayegh, M. A., Hirsekorn, A., Abdul-Rahman, F., Satija, R., Ohler, U., Mahony, S., & Mazzoni, E. O. (2017). A Multi-step Transcriptional and Chromatin State Cascade Underlies Motor Neuron Programming from Embryonic Stem Cells. *Cell Stem Cell*, 20(2), 205–217.e8. <https://doi.org/10.1016/j.stem.2016.11.006>

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F. C. P., Clarke, D., Gu, M., Emani, P., Yang, Y. T., Min, X., Gandal, M. J., Lou, S., Zhang, J., Park, J. J., Yan, C., KyongRhe, S., Manakongtreeep, K., Zhou, H., ... Gerstein, M. B. (2018). Comprehensive functional genomic resource and integrative model for the human brain. *Science*, 362(6420), 8464. <https://doi.org/10.1126/science.aat8464>

Wang, J., Zhang, H., Young, A. G., Qiu, R., Argalian, S., Li, X., Wu, X., Lemke, G., & Lu, Q. (2011). Transcriptome analysis of neural progenitor cells by a genetic dual reporter strategy. *Stem Cells*, 29(10), 1589–1600. <https://doi.org/10.1002/stem.699>

Wang, X., Tucker, N. R., Rizki, G., Mills, R., Krijger, P. H. L., de Wit, E., Subramanian, V., Bartell, E., Nguyen, X.-X., Ye, J., Leyton-Mange, J., Dolmatova, E. V., van der Harst, P., de Laat, W., Ellinor, P. T., Newton-Cheh, C., Milan, D. J., Kellis, M., & Boyer, L. A. (2016). Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures. *ELife*, 5(MAY2016). <https://doi.org/10.7554/eLife.10557>

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B. E., & Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature*, 448(7151), 318–324. <https://doi.org/10.1038/nature05944>

West, A. E., Chen, W. G., Dalva, M. B., Dolmetsch, R. E., Kornhauser, J. M., Shaywitz, A. J., Takasu, M. A., Tao, X., & Greenberg, M. E. (2001). Calcium regulation of neuronal gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 98(20), 11024–11031. <https://doi.org/10.1073/pnas.191352298>

White, J., Gray, R., & Jones, M. (2009). The development of the serious mental illness physical Health Improvement Profile: Practice Development. *Journal of Psychiatric and Mental Health Nursing*, 16(5), 493–498. <https://doi.org/10.1111/j.1365-2850.2009.01375.x>

Wickham, H. (2016). ggplot2. In R. Gentleman, K. Hornik, & G. Parmigiani (Eds.), *ggplot2* (Second Edi). Springer International Publishing. <http://link.springer.com/10.1007/978-3-319-24277-4>

Wilson, C., & Terry, A. V. (2010). Neurodevelopmental animal models of schizophrenia: Role in novel drug discovery and development. In *Clinical Schizophrenia and Related Psychoses* (Vol. 4, Issue 2, pp. 124–137). NIH Public Access. <https://doi.org/10.3371/CSRP.4.2.4>

Wirgenes, K. V., Sønderby, I. E., Haukvik, U. K., Mattingdal, M., Tesli, M., Athanasiu, L., Sundet, K., Røssberg, J. I., Dale, A. M., Brown, A. A., Agartz, I., Melle, I., Djurovic, S., & Andreassen, O. A. (2012). *TCF4 sequence variants and mRNA levels are associated with neurodevelopmental characteristics in psychotic disorders*. 2(5). <https://doi.org/10.1038/tp.2012.39>

Wu, A. R., Neff, N. F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M. E., Mburu, F. M., Mantalas, G. L., Sim, S., Clarke, M. F., & Quake, S. R. (2014). Quantitative assessment of single-cell RNA-sequencing methods. *Nature Methods*, 11(1),

41–46. <https://doi.org/10.1038/nmeth.2694>

Wu, S. J., Furlan, S. N., Mihalas, A. B., Kaya-Okur, H. S., Feroze, A. H., Emerson, S. N., Zheng, Y., Carson, K., Cimino, P. J., Keene, C. D., Sarthy, J. F., Gottardo, R., Ahmad, K., Henikoff, S., & Patel, A. P. (2021). Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nature Biotechnology*, 39(7), 819–824. <https://doi.org/10.1038/s41587-021-00865-z>

Xi, J., Liu, Y., Liu, H., Chen, H., Emborg, M. E., & Zhang, S. C. (2012). Specification of midbrain dopamine neurons from primate pluripotent stem cells. *Stem Cells*, 30(8), 1655–1663. <https://doi.org/10.1002/stem.1152>

Xu, X., & Qi, L. S. (2019). A CRISPR–dCas Toolbox for Genetic Engineering and Synthetic Biology. In *Journal of Molecular Biology* (Vol. 431, Issue 1, pp. 34–47). Academic Press. <https://doi.org/10.1016/j.jmb.2018.06.037>

Yang, J., Horton, J. R., Li, J., Huang, Y., Zhang, X., Blumenthal, R. M., & Cheng, X. (2019). Structural basis for preferential binding of human TCF4 to DNA containing 5-carboxylcytosine. *Nucleic Acids Research*, 47(16), 8375–8387. <https://doi.org/10.1093/nar/gkz381>

Yap, E. L., & Greenberg, M. E. (2018). Activity-Regulated Transcription: Bridging the Gap between Neural Activity and Behavior. *Neuron*, 100(2), 330–348. <https://doi.org/10.1016/j.neuron.2018.10.013>

Yu, C. H., Pal, L. R., & Moult, J. (2016). Consensus Genome-Wide Expression Quantitative Trait Loci and Their Relationship with Human Complex Trait Disease. *OMICS A Journal of Integrative Biology*, 20(7), 400–414. <https://doi.org/10.1089/omi.2016.0063>

Yuan, F., Fang, K.-H., Cao, S.-Y., Qu, Z.-Y., Li, Q., Krencik, R., Xu, M., Bhattacharyya, A., Su, Y.-W., Zhu, D.-Y., & Liu, Y. (2015). Efficient generation of region-specific forebrain neurons from human pluripotent stem cells under highly defined condition. *Scientific Reports*, 5(1), 18550. <https://doi.org/10.1038/srep18550>

Yuzwa, S. A., Borrett, M. J., Innes, B. T., Voronova, A., Ketela, T., Kaplan, D. R., Bader, G. D., & Miller, F. D. (2017). Developmental Emergence of Adult Neural Stem Cells as Revealed by Single-Cell Transcriptional Profiling. *Cell Reports*, 21(13), 3970–3986. <https://doi.org/10.1016/j.celrep.2017.12.017>

Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., & Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*, 518(7540), 556–559. <https://doi.org/10.1038/nature13994>

Zeisel, A., Hochgerner, H., Lönnberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L. E., La Manno, G., Codeluppi, S., Furlan, A., Lee, K., Skene, N., Harris, K. D., Hjerling-Leffler, J., Arenas, E., Ernfors, P., Marklund, U., & Linnarsson, S. (2018). Molecular Architecture of the Mouse Nervous System. *Cell*, 174(4), 999–1014.e22. <https://doi.org/10.1016/j.cell.2018.06.021>

Zhang, S., Zhang, H., Zhou, Y., Qiao, M., Zhao, S., Kozlova, A., Shi, J., Sanders, A. R., Wang, G., Luo, K., Sengupta, S., West, S., Qian, S., Streit, M., Avramopoulos, D., Cowan, C. A., Chen, M., Pang, Z. P., Gejman, P. V., ... Duan, J. (2020). Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants. *Science*, 369(6503), 561–565. <https://doi.org/10.1126/science.aay3983>

Zhang, Xinmin, Odom, D. T., Koo, S. H., Conkright, M. D., Canettieri, G., Best, J., Chen, H., Jenner, R., Herboldsheimer, E., Jacobsen, E., Kadam, S., Ecker, J. R., Emerson, B., Hogenesch, J. B., Unterman, T., Young, R. A., & Montminy, M. (2005). Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. *Proceedings of the National Academy of Sciences of the United States of America*, 102(12), 4459–4464. <https://doi.org/10.1073/pnas.0501076102>

Zhang, Xuejing, Hamblin, M. H., & Yin, K. J. (2017). The long noncoding RNA Malat1: Its physiological and pathophysiological functions. In *RNA Biology* (Vol. 14, Issue 12, pp. 1705–1714). Taylor and Francis Inc. <https://doi.org/10.1080/15476286.2017.1358347>

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., & Südhof, T. C. (2013). Rapid Single-Step Induction of Functional Neurons from Human Pluripotent Stem Cells. *Neuron*, 78(5), 785–798. <https://doi.org/10.1016/j.neuron.2013.05.029>

Zhang, Z., Xiang, D., & Wu, W. S. (2014). Sodium butyrate facilitates reprogramming by derepressing OCT4 transactivity at the promoter of embryonic stem cell-specific miR-302/367 cluster. *Cellular Reprogramming*, 16(2), 130–139. <https://doi.org/10.1089/cell.2013.0070>

Zhao, Y., Liu, H., Zhang, Q., & Zhang, Y. (2020). The functions of long non-coding RNAs in neural stem cell proliferation and differentiation. In *Cell and Bioscience* (Vol. 10, Issue 1, p. 74). BioMed Central Ltd. <https://doi.org/10.1186/s13578-020-00435-x>

Ziller, M. J., Ortega, J. A., Quinlan, K. A., Santos, D. P., Gu, H., Martin, E. J., Galonska, C., Pop, R., Maidl, S., Di Pardo, A., Huang, M., Meltzer, H. Y., Gnrke, A., Heckman, C. J., Meissner, A., & Kiskinis, E. (2018). Dissecting the Functional Consequences of De Novo DNA Methylation Dynamics in Human Motor Neuron Differentiation and Physiology. *Cell Stem Cell*, 22(4), 559–574.e9. <https://doi.org/10.1016/j.stem.2018.02.012>

Zirlinger, M., Lo, L., McMahon, J., McMahon, A. P., & Anderson, D. J. (2002). Transient expression of the bHLH factor neurogenin-2 marks a subpopulation of neural crest cells biased for a sensory but not a neuronal fate. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), 8084–8089. <https://doi.org/10.1073/pnas.122231199>

Zorin, V., Zorina, A., Smetanina, N., Kopnina, P., Ozerov, I. V., Leonov, S., Isaev, A., Klokov, D., & Osipov, A. N. (2017). Diffuse colonies of human skin fibroblasts in relation to cellular senescence and proliferation. *Aging*, 9(5), 1404–1413. <https://doi.org/10.18632/aging.101240>

---

## 7 Acknowledgement

This PhD thesis would not have been possible without the guidance and help of a lot of people which supported me in one way or another. My sincere thanks go to Prof. Dr. Michael Ziller who gave me the opportunity to perform my PhD thesis in his laboratory and trusted in my abilities to work on this exciting project. Additionally, I would like to thank Prof. Dr. Wolfgang Enard and Prof. Dr. Moritz Rossner for their guidance and advice during the thesis advisory committee meetings. Special thanks to Prof. Dr. Wolfgang Enard, for giving me the opportunity to defend my thesis at the biological faculty of the LMU. I also would like to thank Prof. Dr. Keays, Prof Dr. Bolle, PD Dr. Marín, and Prof. Dr. Schneeberger for their time and support.

Besides my direct advisors, I am grateful to the people who enabled this research by donating skin tissue or blood to generate patient and healthy control derived induced pluripotent stem cells. Likewise, I would like to thank my fellow lab mates for their encouragement, insightful comments, and moral support. Obviously, dear wet lab people in crime, thank you very much for sharing the duty of feeding needy iPSCs on the weekends, bank holidays, during vacations, over Christmas, and while I was home with a shattered heel. Special thanks to those people who performed the review experiments for my first-author paper, which is based on a more recent analysis of my PhD data sets, when I had already transitioned to a new working place.

This thesis would not have been possible without several fruitful collaborations. First of all, I would like to thank Alex who kindly provided me with several million primary mouse neurons whenever yet another MPRA library was ready for transduction. Second of all, I would like to thank Barbara for patching all those neurons (and heavily trying for those cultures with the changed media compositions) and Dr. Mathias Eder for analyzing the mini EPSC and MEA experiments.

The bioinformatic side of the lab, I would like to especially thank for making me less afraid of R, supporting my first ggplottings, and explanations how to best increase batch effects ;-)... I will never forget all the cakes we had in our kitchen (special thanks to Susi!) and all the “we must leave now and go for food” dinner plans. In the line of food – thank you Lucia for introducing me to the Italian way of cooking and drinking a cappuccino (never when I wanted),

---

and that Aperol Spritz is tasty after all. Also, I am very thankful for all the morning breakfasts at Cotidiano with Laura which kept on refueling my energy.

I am heartfully thankful to all my friends which kept me sane during the PhD time and the time when I could not walk (and my parents for additionally taking care of me while with crutches). I am grateful for all the beer walks and political discussions I had with Vanessa. I am thankful to Liesa to support and foster my love for plants, buying plants, and exchanging plant pictures. I am absolutely grateful for Jassi who made me repeatedly take holidays to visit far away and nearby, and additionally always listened in cases of emergency.

Furthermore, I offer my regards and blessings to all of those who supported me in any respect before and during the completion of this thesis. Special thanks to my Bachelor of Science crew which has been a stable friend group during my whole scientific career. My friends which brought me to the mountains for hikes (especially Lisa for the little holidays at her residency at the Chiemsee) allowed me to have a great excuse for the heavily needed nature time on the weekends. Last but not least, I would like to show my gratitude to my family for their loving support.

## 8 Supplementary tables

In this chapter, the fragments are described which were tested in the MPRA (Chr. = Chromosome, Var. Pos. = Variant position, Ref. = reference allele, Alt. = Alternative allele).

## 8.1 Control fragments for the MPRA experiments







Random.28	CTTAATCTTCTCGCTTAACCTATTGAGGCTTAAACGCAACGCTAGTAAACGAGGACTACTCTCATCCTGGCTTCATCAAGGATCTGTTAGGGCCAGTAGTGGC
Random.29	GACCAAAAGACTGGCTTCCGAAGGATTGGGAAATTCACCGCGAACATCTTAAAGTACGAGTACGGTCAAGGACTATGAACGCTGGAGCCACATTGAGATGA
Random.30	TCCTGAGCTACGACCAACTAACCATGGGGCTCTAGCGTAAAGC
Random.31	TAATTAGCAAGTAAATCAGTGGACGAGGACCCAAAGTAACTGGGATTTCAATTACTCGAGGTATCGAGGAAATGAGTTCCATAGCGAACCTACCG
Random.32	AGTTGGGTTATGTTACAGTCGAGGATTATACCTCGCTTCAAGACG
Random.33	GCTTACAGAAGAACCTACGGCTACGAGTCAGTCAGTGGAGAATCTTAAAGGATTAGCGATACTCTCTTGGGTTAGCGAGTCAAAGAGCTGAAGCTGCTCA
Random.34	ATTGCTTATGTTGAGAACGACTAACATCGAGAGGCCATTAGAAG
Random.35	CGGAGCTGGGACAGTGGGACCCCTGTGGGCTCTGACTTAAACGCTACTAAAACGAGCTAACGCTACGGTATCGGGTACAGTGGCTAGGCTAGGCTAGGACTCG
Random.36	CTGGTACGGCCGAACTCGCGCTTCAAGATGCTGCTCAAGTCGAGGAG
Random.37	CTGAGATAACAGAGTATGAGATGTATCTTCTGGTTAGTTGAGGCAAAT
Random.38	TAAGGAGCTGGGATTTAACAGGGCCCTGTGGGAGCTAACAGGCTACAGTAAAGGAGCTTAACTCGAGCTTGGCTTAACCCGGTAAGGGTGCAGTACGAATATCTGGCTA
Random.39	AGAATTAGGTTCTATTCTCTGTTAGAGTGTAGTCTGGCTGGACACAG
Random.40	CTTGGCAGAAGTATGGCAGACTCAGTTAGGCCGAAGTCACCCCTAGCTGGACTACGAGAATCCAAAATAGAACATACACCAAGCTAACCTACAGTAAAGG
Random.41	GGGGACGCTTACCAAGCATCACACTCCGCTTGTGGGATGCTGCAAAACTC
Random.42	GTGGAGCTGGGACTTACCTGGCTTAACTTACGGTACTCTGGCTCTTACAGCTTCTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.43	GTGGAGGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.44	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.45	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.46	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.47	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.48	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.49	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.50	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.51	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.52	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.53	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.54	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.55	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.56	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.57	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.58	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.59	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.60	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.61	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.62	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.63	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.64	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.65	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.66	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.67	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.68	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.69	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA
Random.70	CGAGAAATCTTACCTGGCTTAACTGGGCTTACGGTACTCTGGCTGGGAGCTTACGTTGAAATTGCAACGGACTGATACTATTCTGTAACAGTTAGTCCTAACAGCTGA

## 8.2 Fragments of the DevLib

All control fragments indicated in 8.1 were included. The SNP was located at position 81 of 160 nt.

Location start (hg19)	SNP								
chr6:31325058	C>T	chr6:31040766	A>G	chr7:104842645	G>A	chr3:52719090	C>T	chr10:104386154	A>G
chr6:31797589	C>G	chr6:31600822	A>G	chr6:56573658	T>C	chr6:25925989	G>A	chr10:104844874	T>C
chr10:104594509	A>G	chr6:28641959	A>G	chr6:27837185	T>C	chr6:26022246	A>G	chr10:104609376	T>C
chr17:17866899	A>G	chr6:31430723	A>G	chr6:28092309	C>G	chr6:26365348	G>C	chr10:104609468	T>C
chr6:28367357	G>A	chr6:32145207	C>T	chr2:198362526	C>T	chr10:104629467	T>G	chr10:104609678	T>C
chr6:28367625	C>G	chr6:32933495	G>C	chr10:3821563	G>A	chr12:132469649	G>T	chr11:130758155	A>T
chr6:28367660	G>A	chr6:31433695	T>C	chr2:145196198	T>G	chr6:29644042	C>T	chr11:130786310	T>C
chr5:139064269	C>G	chr6:30748166	C>G	chr6:27730336	T>C	chr17:17748647	T>C	chr11:1803017916	C>T
chr5:152897409	G>C	chr6:32338285	G>T	chr2:76309966	C>G	chr16:462621	A>G	chr11:130758115	G>A
chr5:139071355	T>C	chr6:30761574	C>T	chr6:27775030	C>T	chr8:10283428	G>A	chr11:130758154	G>T
chr14:104314185	C>G	chr6:31240182	A>C	chr16:45488334	C>G	chr8:10283604	C>A	chr11:130758558	G>A
chr14:104314184	G>A	chr6:30706363	T>A	chr6:27835220	G>A	chr11:130786347	C>A	chr11:47013443	G>A
chr6:31783210	A>C	chr11:57435298	G>A	chr11:46958301	G>T	chr6:26045907	G>C	chr11:46673346	C>T
chr8:38236892	G>A	chr11:64683534	C>T	chr14:104286212	C>T	chr1:36616339	T>C	chr11:113377490	C>A
chr14:104190529	A>G	chr6:288066305	G>C	chr1:98499845	T>C	chr2:201245274	G>A	chr14:46513251	C>T
chr2:198171100	T>C	chr6:31799078	T>G	chr1:98498900	C>A	chr5:109075511	C>T	chr11:57481157	G>A
chr11:130742790	T>C	chr6:31870328	G>A	chr1:98525223	A>G	chr5:109026665	G>C	chr11:46934371	C>T
chr12:2321870	A>T	chr6:30495862	A>G	chr11:65302850	T>C	chr5:137673169	A>C	chr11:65378030	C>T

chr2:220042677	G>A	chr6:31630970	G>A	chr11:65382566	T>C	chr6:26476784	T>A	chr11:130743390	T>C
chr2:198463092	G>T	chr6:29895101	C>A	chr10:104630414	A>T	chr6:28107359	T>C	chr11:46622394	C>T
chr2:220042629	A>C	chr6:33360576	T>G	chr12:123633426	G>A	chr10:104944746	C>G	chr11:46928457	G>A
chr2:27974973	A>G	chr6:29765293	A>G	chr12:123591598	G>T	chr12:123757146	T>G	chr12:123593384	C>T
chr2:172780134	T>C	chr11:57435213	C>T	chr12:123636843	A>G	chr12:123464281	T>C	chr22:41418231	T>G
chr2:76310298	G>A	chr6:29894009	A>T	chr12:123606741	G>A	chr12:110888859	C>T	chr12:2303488	G>A
chr7:2149969	A>G	chr12:123529058	A>G	chr14:104165929	T>C	chr6:296454504	C>T	chr15:85113492	G>C
chr7:2158392	A>G	chr11:64683701	C>T	chr6:31546852	G>A	chr16:4524062	C>A	chr14:104018653	A>G
chr12:2402248	A>G	chr6:32113573	A>T	chr22:42486725	G>A	chr6:29644635	C>T	chr14:104041376	C>T
chr7:86447947	C>A	chr6:28949962	T>C	chr3:180772749	G>C	chr12:205015701	C>T	chr14:104191919	G>C
chr7:104844521	G>A	chr6:28863371	A>G	chr3:180773882	T>G	chr1:205015801	A>T	chr14:104028272	T>C
chr7:104835890	T>C	chr6:32654151	A>C	chr3:180774849	C>A	chr2:225450163	G>C	chr10:18763196	C>T
chr19:2695663	C>T	chr6:31239804	C>G	chr3:180632128	T>G	chr2:201173622	A>G	chr10:18763345	C>G
chr19:50096423	C>G	chr6:31697560	G>A	chr3:180630567	T>C	chr2:201173102	C>G	chr10:18763388	C>T
chr19:2696957	A>G	chr6:32411378	T>G	chr3:180632032	T>C	chr2:201173050	A>G	chr1:243418184	C>T
chr10:3820789	T>C	chr6:30899054	G>A	chr3:180774989	A>T	chr2:201173001	C>T	chr6:27970717	T>C
chr6:31783509	G>C	chr12:123520073	C>T	chr3:180774430	G>A	chr2:37576138	A>G	chr6:26365915	G>T
chr17:19239434	C>T	chr6:32603658	C>G	chr3:180774675	G>A	chr7:110915341	C>T	chr6:26532100	A>C
chr17:19265799	C>T	chr12:2300609	G>C	chr2:22474140	C>G	chr7:110931477	C>T	chr6:27640991	A>G
chr6:26986794	C>T	chr6:32602954	T>A	chr2:21252964	C>G	chr7:2143210	C>A	chr6:27797418	C>A
chr10:104841481	C>T	chr6:32603009	A>G	chr6:32595423	T>A	chr7:2166321	T>G	chr6:27799472	A>G
chr6:27534530	A>G	chr6:28624896	T>A	chr17:19914401	T>A	chr10:104559721	C>T	chr6:56574221	T>C
chr6:26030494	C>A	chr6:28650525	T>C	chr6:32595385	T>C	chr10:104595422	T>G	chr5:109046212	A>T
chr6:26031813	T>G	chr6:31430067	C>T	chr6:32601836	A>G	chr14:103967047	G>A	chr14:103984618	C>T
chr6:26309910	G>A	chr6:31431815	A>G	chr10:88112941	G>A	chr11:130767488	G>T	chr14:104002236	G>A
chr2:185462043	T>C	chr6:28834144	G>C	chr5:32719400	C>A	chr11:130763997	C>G	chr15:91428638	G>A
chr11:130747223	C>G	chr6:32338869	A>G	chr17:13009042	T>C	chr12:2339255	C>G	chr15:91428591	T>G
chr11:130746148	T>C	chr6:32146646	C>T	chr6:32601887	C>T	chr1:18431609	C>T	chr16:9914403	C>G
chr12:99503624	G>C	chr6:31927344	A>G	chr6:26314235	G>C	chr2:172544485	A>G	chr16:4459836	A>G
chr1:150266561	A>C	chr6:30717260	C>T	chr17:17347472	G>A	chr3:52953232	C>T	chr16:4467535	C>T
chr5:152013791	G>C	chr6:25999151	C>T	chr10:104748461	A>C	chr4:1705040859	C>T	chr16:4460116	A>G
chr5:152013889	G>A	chr6:28696065	A>C	chr10:104748720	A>G	chr5:109077519	T>C	chr17:17907395	T>A
chr17:1281519	T>C	chr6:30796661	A>G	chr15:91428957	C>T	chr7:214380	C>T	chr17:19988545	C>A
chr11:130742632	G>A	chr6:32161432	C>A	chr15:91429044	G>A	chr7:104837803	C>A	chr17:2084307	A>G
chr10:104883339	A>G	chr6:31024798	G>A	chr6:32601870	T>C	chr11:1338227735	A>C	chr17:2085600	G>A
chr6:33282630	C>T	chr6:28920974	G>A	chr6:32595420	A>G	chr17:176965333	G>C	chr12:29934588	C>T
chr12:123743885	G>C	chr6:33359144	A>G	chr1:73313764	G>A	chr14:135514959	G>A	chr17:19250681	G>A
chr12:123743447	G>A	chr14:59951460	T>C	chr6:32119732	G>C	chr15:78832834	T>C	chr17:17818819	G>T
chr12:123743449	G>A	chr6:31240062	A>G	chr18:53067158	C>A	chr9:96213148	T>A	chr17:2085655	T>C
chr12:99532289	A>G	chr6:32439325	C>T	chr1:98473718	T>A	chr11:183184484	C>T	chr18:53074620	A>G
chr11:30429440	A>G	chr14:99712947	A>G	chr1:98472960	T>A	chr1:234498833	G>A	chr18:46714698	T>C
chr12:123721616	A>C	chr6:32147698	G>A	chr1:98472881	C>T	chr5:140187324	C>T	chr19:50093574	A>G
chr12:123723737	G>C	chr6:28586284	T>A	chr1:98491734	T>C	chr5:140187104	A>G	chr19:50096861	G>C
chr1:98470607	A>G	chr6:28864851	A>G	chr6:19893265	A>G	chr7:104624628	C>T	chr19:50097549	A>G
chr10:104880238	T>C	chr6:30652783	C>T	chr8:10268807	G>A	chr10:104614352	C>T	chr8:38240394	C>T
chr10:104943988	T>A	chr6:28602633	G>A	chr11:130717442	C>T	chr10:104595851	G>C	chr8:60556511	C>A
chr10:104953549	T>C	chr6:32096003	C>A	chr6:28105198	C>T	chr12:2339252	G>A	chr8:60556828	C>A
chr11:130719063	G>T	chr6:28806336	T>C	chr14:60144733	C>A	chr3:52349206	A>G	chr20:58253440	T>C
chr11:130764542	A>T	chr15:85113339	G>T	chr14:60146350	C>T	chr3:52349414	G>T	chr1:150058579	G>T
chr1:2383041	A>G	chr6:31688219	A>G	chr14:60146557	A>G	chr1:163716906	A>G	chr1:150123764	G>T
chr11:30428277	G>A	chr15:83318204	T>C	chr12:110907575	C>G	chr6:26157764	C>T	chr1:150133509	C>A
chr12:123719882	C>A	chr6:32122474	C>A	chr10:104744408	A>G	chr14:104322028	A>G	chr1:150134223	G>C
chr12:123746963	A>T	chr6:32223787	G>C	chr3:181716297	T>G	chr7:111078914	T>G	chr1:150135293	C>T
chr12:1229692	G>A	chr15:85113974	C>A	chr2:39869999	T>G	chr17:176965333	G>C	chr1:150185441	G>A
chr12:2314428	C>T	chr6:32603244	G>A	chr22:41685939	G>C	chr6:25930090	G>A	chr1:150274491	C>A
chr12:2330460	C>T	chr6:32411309	C>T	chr6:27742037	A>G	chr6:26364586	C>G	chr1:150274611	A>G
chr10:104387737	T>C	chr6:28727213	G>A	chr8:10268807	G>A	chr6:32634469	C>T	chr1:150337291	G>T
chr10:104412310	C>T	chr6:29766127	G>A	chr6:26139935	C>T	chr6:32634320	C>A	chr1:150426125	C>T
chr10:104596398	C>T	chr6:29759887	G>A	chr1:6713284	A>G	chr1:205041544	T>C	chr1:150540183	C>T
chr10:104698525	G>A	chr6:32145995	C>G	chr1:6721236	T>A	chr20:58253506	C>T	chr20:58253585	G>A
chr10:104794949	G>A	chr6:30723783	G>A	chr6:27773834	T>A	chr6:27782537	C>T	chr20:58253585	G>A
chr10:104886190	C>G	chr6:29892734	G>A	chr6:27773906	G>A	chr19:19431425	T>G	chr5:60696325	A>G
chr10:104935595	G>C	chr6:30798699	C>T	chr6:27774826	G>A	chr10:104429641	G>A	chr5:152055864	G>T
chr10:104943344	A>G	chr6:31184198	A>G	chr6:27775676	A>G	chr6:27782540	T>A	chr5:152191957	A>T
chr10:104956829	A>G	chr6:31434522	G>C	chr6:27775699	G>C	chr6:27782530	T>C	chr5:152191997	C>T
chr10:104698525	G>A	chr6:32115280	C>A	chr6:27779508	C>T	chr1:150058729	C>T	chr6:25962953	C>A
chr10:104794949	G>A	chr6:31319228	C>T	chr6:27798259	T>C	chr6:56572479	A>G	chr6:25999177	G>A
chr10:130744979	G>A	chr6:29623741	G>A	chr6:27796937	G>T	chr17:17833646	A>T	chr6:26563866	A>C
chr15:85177299	G>A	chr6:31633498	G>T	chr6:27792642	C>T	chr6:26021674	C>T	chr6:26550146	C>T
chr1:2383758	A>G	chr6:31634195	A>G	chr6:27791867	G>T	chr2:233641012	A>T	chr7:27730084	C>T
chr1:2384110	T>C	chr6:33359010	A>G	chr6:27788944	C>T	chr22:41830876	C>T	chr2:48684798	A>G
chr11:130718632	T>G	chr6:28831613	C>T	chr6:27785822	T>C	chr22:41831324	G>A	chr22:50321625	C>T
chr11:130718713	T>G	chr6:31820645	A>G	chr7:20483373	A>G	chr2:73461555	C>T	chr22:39897165	G>A
chr11:130718815	G>C	chr6:29892734	G>A	chr6:27835274	T>C	chr1:150974973	T>A	chr2:198492318	T>G
chr11:130744979	G>A	chr6:30798699	C>T	chr6:27835437	A>G	chr5:60614881	G>T	chr18:53050648	A>C
chr15:85177299	G>A	chr6:31184198	A>G	chr6:27835774	A>G	chr17:179241507	G>A	chr18:53070170	C>T
chr1:2383758	A>G	chr6:31434522	G>C	chr6:27836978	C>T	chr18:53251727	G>A	chr18:53076738	C>T
chr1:2384110	T>C	chr6:32115280	C>A	chr6:27837269	T>C	chr2:233562733	C>G	chr18:53076792	G>A
chr1:243555221	A>G	chr6:31864549	G>T	chr6:27838766	C>G	chr2:233562199	T>C	chr18:53085129	C>A
chr11:46686262	G>A	chr6:32134512	G>A	chr6:27839748	T>C	chr10:104041723	G>A	chr18:53087986	T>A
chr12:123753494	C>G	chr6:31697389	C>G	chr6:278666945	G>C	chr10:1040401634	T>C	chr22:41047558	G>T
chr12:2288407	T>C	chr6:309315037	G>A	chr6:278666386	G>A	chr10:104536362	G>A	chr18:53125476	G>C
chr12:2288438	T>C	chr6:29763259	C>G	chr6:27862154	T>C	chr22:42487902	G>A	chr18:53144839	C>T
chr12:2314827	G>A	chr6:29764550	A>G	chr6:27862129	A>G	chr12:123450767	G>C	chr22:41047753	G>T
chr12:2315959	C>T	chr6:31051677	A>G	chr6:27859570	T>C	chr12:123569377	A>G	chr18:53533191	C>T

chr2:208330432	A>G	chr6:31239229	C>T	chr1:98493787	G>A	chr17:17839026	C>A	chr22:41060855	A>T
chr3:52719818	C>T	chr6:31154356	T>A	chr1:98493791	A>G	chr5:60586627	T>C	chr7:86758466	G>T
chr5:60696427	C>T	chr6:29759447	C>T	chr2:41697340	A>G	chr6:28367685	T>G	chr7:86847968	G>A
chr5:60727992	T>C	chr6:28583394	A>G	chr2:41699208	G>C	chr11:130717258	T>C	chr5:153624048	T>A
chr5:60751469	T>C	chr6:31197265	G>A	chr1:150039680	G>A	chr2:22709904	C>A	chr11:134296643	A>G
chr6:26233389	A>G	chr6:31239520	C>T	chr7:105008460	C>G	chr1:98395883	G>T	chr6:32914728	G>C
chr6:26361948	C>G	chr6:31214559	C>T	chr1:243878505	A>G	chr5:60598545	A>G	chr6:43337804	A>G
chr6:26364058	C>T	chr6:32223777	A>G	chr2:200905138	C>A	chr16:29931901	G>A	chr18:53085394	C>T
chr8:60544008	A>C	chr6:30742715	G>C	chr6:28048626	T>C	chr10:104616665	T>C	chr18:52985403	C>A
chr11:46543526	C>T	chr6:29895501	A>G	chr6:28048998	T>G	chr16:29938822	A>G	chr18:53052925	C>T
chr2:233561745	C>G	chr6:29894602	G>T	chr7:2206112	A>G	chr1:98409305	C>A	chr18:53053400	G>A
chr12:123720777	G>T	chr6:31319217	G>C	chr10:104871363	A>G	chr11:46684679	T>G	chr18:53071768	A>G
chr12:2312182	T>A	chr6:31138912	T>C	chr15:91428199	T>C	chr8:143388292	T>C	chr17:17741877	A>G
chr12:2329972	T>C	chr6:31620022	A>G	chr6:33394255	C>G	chr1:73684275	G>A	chr2:233641085	G>C
chr12:2334262	G>A	chr6:31588386	G>T	chr1:243418065	C>A	chr6:29647547	T>C	chr1:36622188	G>C
chr12:2348846	G>C	chr6:31026591	T>C	chr20:37503319	T>A	chr6:29647644	T>C	chr5:153622823	G>A
chr12:23626984	C>T	chr6:30296255	G>A	chr20:37502639	T>G	chr1:98404236	T>C	chr17:17942615	C>T
chr12:23627018	T>C	chr6:28852158	C>T	chr6:33540211	A>G	chr4:170263445	G>T	chr22:46477392	C>G
chr2:201003839	A>C	chr6:31182973	C>T	chr6:33546839	T>C	chr5:60589741	C>T	chr6:27739256	T>C
chr2:201004091	T>C	chr6:28831023	T>G	chr6:33546932	T>C	chr6:32808301	G>A	chr10:104913942	T>C
chr7:2041640	G>A	chr6:30170289	A>G	chr7:104616746	G>A	chr6:27840928	T>C	chr2:201711131	G>A
chr7:2041434	A>G	chr6:32935181	A>C	chr3:17261457	G>A	chr6:265959511	A>G	chr10:104597154	A>G
chr16:4460051	A>G	chr6:29765272	C>T	chr18:53214867	T>C	chr4:170261501	C>G	chr8:10254324	C>T
chr16:4460135	G>A	chr6:33359718	G>C	chr18:53214925	C>T	chr28:367665	T>C	chr2:23726722	G>T
chr16:4549620	G>A	chr2:201051152	A>G	chr2:194547898	G>C	chr16:29988943	A>G	chr12:123633059	C>T
chr17:17728985	A>G	chr6:28866530	T>C	chr6:28234399	T>C	chr5:60601436	A>G	chr2:200941369	T>C
chr17:17824980	G>T	chr6:33085352	A>G	chr6:28233940	T>C	chr7:104611878	C>T	chr12:123849098	C>G
chr17:17840825	G>T	chr6:32602704	T>C	chr6:28321995	T>C	chr11:46691739	A>G	chr11:64684992	C>T
chr17:17841085	T>C	chr6:32450703	A>G	chr7:110848331	T>A	chr16:29932066	C>T	chr6:32605886	C>T
chr17:17876298	C>G	chr6:29764665	T>C	chr7:110851555	A>G	chr3:36862982	C>T	chr18:53023172	G>A
chr12:2333640	T>C	chr6:29894586	C>A	chr4:23425655	C>T	chr6:26598006	A>G	chr15:91429178	G>T
chr12:2333673	G>A	chr6:32156910	G>A	chr4:23425590	A>T	chr20:48120779	G>A	chr6:25248008	T>C
chr19:50096146	T>C	chr2:209916008	G>A	chr4:23425476	T>G	chr1:2374760	A>G	chr2:201068686	C>A
chr19:33834563	G>A	chr6:31434113	A>G	chr4:23425042	C>G	chr1:2387103	C>T	chr17:17779460	C>T
chr19:33834682	T>C	chr6:30708957	C>T	chr4:23424547	T>C	chr1:73532153	T>C	chr12:123716468	G>A
chr3:52720082	A>C	chr6:32813770	G>A	chr4:23424062	T>G	chr1:73542902	A>G	chr11:133822571	A>G
chr1:8362462	G>A	chr6:32410212	T>C	chr1:98465530	G>A	chr1:73543030	G>A	chr1:36622905	A>G
chr4:176865768	A>T	chr2:209916008	G>A	chr12:2346395	C>T	chr1:73544072	C>T	chr1:20977085	A>G
chr6:29913139	A>G	chr2:209916241	C>T	chr12:2288947	T>A	chr1:366275544	C>T	chr6:96462498	G>T
chr6:31241666	G>A	chr2:20143411	T>C	chr5:109141529	T>G	chr2:200819179	C>A	chr6:43395103	G>A
chr16:29985627	G>A	chr2:57988196	A>T	chr17:2126506	G>C	chr2:200915397	T>C	chr1:36616943	T>G
chr6:29911248	C>T	chr2:20998618	A>G	chr17:2196090	C>T	chr3:52630550	G>A	chr1:98385832	T>C
chr18:53066330	A>C	chr2:209916232	T>C	chr17:2195148	T>A	chr3:52880545	G>A	chr1:73530817	C>T
chr14:66314110	A>G	chr20:4812036	C>T	chr17:21951532	A>G	chr4:23334813	G>A	chr12:123752421	C>T
chr7:2145631	T>C	chr3:180857107	T>A	chr7:214622918	T>C	chr15:91426562	G>A	chr6:33755713	G>C
chr8:27437575	C>T	chr3:52559707	G>A	chr15:85200948	G>C	chr6:27446361	A>T	chr6:33755543	G>A
chr1:243669014	T>C	chr6:96464062	C>T	chr5:87246545	C>T	chr6:25417425	G>T	chr2:201057493	G>A
chr10:104430253	G>A	chr6:32552450	A>C	chr6:33748833	T>G	chr6:26361432	C>T	chr2:28022677	G>A
chr10:104707018	A>T	chr6:105388720	C>T	chr7:82447247	C>A	chr6:28092687	C>T	chr18:53060417	T>C
chr10:104869040	T>C	chr6:33715326	A>T	chr7:104613207	T>C	chr7:20481388	T>C	chr22:41802440	T>C
chr10:104886376	G>A	chr7:1904711	C>T	chr7:214622918	T>C	chr12:2361462	C>T	chr12:2294265	A>G
chr10:104906213	T>C	chr7:2027450	T>C	chr6:26431984	C>T	chr7:105028372	T>C	chr12:2314999	T>A
chr10:104913655	G>A	chr7:2027313	G>T	chr22:41661156	G>T	chr12:2349586	G>C	chr12:2315230	G>A
chr10:104943050	A>C	chr7:2027326	G>A	chr17:17747368	A>C	chr12:2365906	G>A	chr12:2315708	T>C
chr10:104963895	C>T	chr7:1986740	G>A	chr6:26104632	C>T	chr12:2419898	A>T	chr2:198380637	C>T
chr10:123912582	G>A	chr6:31239579	A>C	chr6:31079646	G>A	chr16:4500124	C>T	chr2:145214609	T>G
chr1:73531200	C>T	chr7:78345209	C>T	chr6:26431984	C>T	chr16:45002124	C>T	chr2:22524866	G>A
chr1:73531288	C>T	chr7:2155252	C>G	chr22:41661156	G>T	chr16:4573042	A>G	chr2:198718632	A>G
chr1:735322696	G>A	chr7:2155253	C>G	chr17:17747368	A>C	chr16:29972692	C>T	chr2:201087159	C>G
chr1:44029355	C>G	chr7:2050403	C>T	chr6:26104632	C>T	chr17:2074979	T>C	chr3:525666916	C>A
chr6:29910912	C>T	chr7:2072356	C>T	chr7:24797766	T>C	chr17:55728226	G>A	chr3:52279596	G>A
chr6:299109665	G>A	chr7:30745209	C>T	chr7:104845255	G>C	chr19:31051408	T>C	chr3:52333673	C>G
chr6:299109799	C>T	chr7:2155252	C>G	chr12:2299825	C>T	chr19:19572110	T>G	chr2:86947785	C>T
chr6:31606378	T>C	chr7:2155253	C>G	chr12:2300000	A>C	chr19:19640526	T>C	chr3:17784820	T>C
chr1:44114975	G>C	chr7:2050403	C>T	chr12:2300010	A>G	chr12:299547719	G>A	chr2:58135012	C>G
chr6:29910912	C>T	chr7:2072356	C>T	chr12:2300209	A>G	chr22:41637121	G>T	chr3:52566684	G>A
chr6:29726048	C>T	chr8:38298649	C>T	chr7:24797766	T>C	chr5:137784256	C>T	chr3:52279596	G>A
chr11:130744672	A>G	chr8:10255404	A>G	chr12:3070769	G>A	chr15:137945753	C>G	chr3:52333673	C>G
chr11:130744769	T>G	chr15:34659519	G>C	chr12:3071777	C>A	chr8:9606266	G>A	chr2:22524866	G>A
chr6:31606378	T>C	chr10:104391629	A>G	chr12:3072798	G>A	chr8:10253100	A>G	chr2:198718632	A>G
chr1:44114975	G>C	chr14:184668769	G>A	chr12:3084007	T>G	chr16:29972692	C>T	chr2:201087159	C>G
chr6:29910912	C>T	chr14:104156110	C>G	chr12:30844330	C>T	chr17:2028000627	C>T	chr3:525666916	C>A
chr6:29910733	G>A	chr17:17715103	G>A	chr12:3335943	G>A	chr3:180851259	A>T	chr2:200929533	C>T
chr6:29895628	T>A	chr17:17826530	T>A	chr12:3374132	G>A	chr5:1378724167	G>T	chr6:96464630	G>T
chr6:29909023	C>G	chr19:2686463	C>A	chr14:133292745	A>C	chr6:337310935	G>A	chr18:52987163	T>A
chr6:29911224	C>T	chr19:50095963	T>C	chr14:71373582	T>C	chr14:104145480	G>A	chr16:4562514	A>G
chr7:86782141	G>A	chr14:470580639	C>T	chr7:104844330	C>T	chr14:104177100	G>T	chr2:23726721	A>C
chr6:29876871	A>G	chr8:10256056	A>G	chr14:104024951	G>A	chr14:103980136	A>T	chr2:200734370	C>T
chr16:4476091	T>C	chr1:30448219	T>C	chr14:123433656	A>G	chr14:104021143	C>A	chr6:25219584	T>C
chr6:31177096	G>A	chr1:73977109	T>G	chr10:104814164	T>C	chr5:139045215	A>G	chr5:139071634	G>A
chr6:31197295	C>T	chr1:150972541	T>G	chr10:104942512	G>T	chr20:48121001	C>A	chr2:198669756	G>C
chr6:32161368	T>C	chr1:150974083	C>T	chr17:130498126	C>A	chr17:13627531	C>T	chr5:60773231	C>T
chr6:31434368	G>A	chr1:150339518	T>C	chr1:150974178	T>G	chr15:139070653	A>G	chr1:243555107	G>A
chr6:32146999	A>G	chr1:207992548	G>A	chr1:150974178	T>G	chr5:139070782	A>T	chr5:60773970	G>A
chr6:32601224	T>C	chr1:150154354	C>T	chr1:2299050	C>T	chr10:104388712	T>C	chr5:60736951	A>G
chr6:32657438	T>C	chr1:115626394	G>C	chr1:20:37377141	C>T	chr10:10494			

chr6:29622222	C>G	chr1:207999607	C>T	chr22:42485673	T>C	chr6:31838547	G>A	chr6:32607256	G>A
chr6:31036785	A>G	chr1:57508680	G>C	chr6:32607398	T>C	chr11:130741608	A>G	chr6:27034183	A>G
chr6:28832790	A>G	chr14:104060394	G>A	chr5:109076490	T>C	chr11:130741763	A>G	chr6:28301145	T>C
chr6:32134658	T>C	chr14:104006718	T>C	chr5:109079369	C>T	chr11:130741942	A>G	chr6:26474046	C>T
chr6:28522697	A>G	chr14:103850908	G>C	chr2:233704254	A>G	chr11:134297112	A>G	chr6:32558125	A>G
chr6:30899197	G>A	chr7:1982183	A>G	chr2:233642232	T>A	chr6:111863197	G>C	chr7:104886671	T>C
chr6:28726817	C>T	chr22:42665119	C>T	chr2:233641926	C>T	chr6:31838715	G>A	chr7:121950967	A>G
chr6:31800870	G>C	chr6:26514942	C>T	chr9:37079663	T>G	chr6:27850875	G>A	chr2:233743111	A>G
chr6:32595280	C>T	chr1:207978782	A>G	chr9:131591834	G>A	chr1:30447763	G>A	chr5:140027218	T>C
chr6:29895543	T>C	chr6:26367220	C>T	chr14:104202306	G>A	chr1:305457879	T>G	chr5:140027306	T>C
chr6:29895471	G>A	chr10:104804969	A>G	chr14:104193101	C>G	chr1:98491250	A>C	chr10:104748011	G>A
chr6:31794594	C>T	chr10:104677128	C>T	chr14:104180446	C>T	chr1:98491384	G>A	chr13:40396139	C>A
chr6:28853199	A>G	chr11:46540474	A>G	chr14:104029821	T>G	chr5:152014063	C>T	chr7:1982781	C>T
chr6:29760525	C>T	chr11:46361980	C>G	chr14:103996207	T>C	chr5:152015599	T>G	chr7:86940654	C>T
chr6:29764376	A>T	chr12:124482258	T>C	chr17:19213337	C>G	chr14:35811701	C>T	chr7:104838052	G>A
chr6:31619654	T>A	chr12:2394783	A>T	chr17:17153139	C>G	chr11:5085275	T>C	chr18:53085275	T>C
chr6:28533948	C>T	chr12:2346832	T>A	chr10:104679980	C>T	chr11:133817057	A>G	chr8:89339723	T>G
chr6:29961276	A>G	chr11:130744492	T>G	chr10:104680139	T>A	chr17:17832462	A>G	chr14:72448234	G>T
chr6:31587872	A>T	chr11:130744270	A>C	chr10:104390305	A>G	chr10:104041220	A>C	chr18:53063678	G>T
chr6:31433560	G>A	chr11:130744511	A>G	chr19:1950169022	T>C	chr6:25873027	C>A	chr8:10190042	C>T
chr6:30811267	G>C	chr10:104871206	A>G	chr19:50168929	A>C	chr11:133817135	G>A	chr8:89338157	C>T
chr6:31430696	G>C	chr10:104957631	T>A	chr19:50168873	G>A	chr15:78857898	T>A	chr6:105472322	G>A
chr6:31239451	C>G	chr10:104928916	T>A	chr16:4475870	A>T	chr11:133817794	T>C	chr6:32635199	T>C
chr6:30781303	T>C	chr10:104629013	T>C	chr16:4500546	C>T	chr9:77342817	C>G	chr9:96213531	A>G
chr6:30855213	T>G	chr12:2314321	A>C	chr2:233660706	T>G	chr9:77342865	A>T	chr9:7171608	T>G
chr6:32637304	G>A	chr12:2322515	A>G	chr13:7438217	G>A	chr6:33174783	G>A	chr10:18745107	T>G
chr6:32223832	A>T	chr14:35809416	G>A	chr4:176867626	A>G	chr6:32577635	G>C	chr10:104872549	A>T
chr6:30882417	C>T	chr16:4464619	T>C	chr5:139063730	G>A	chr11:83278972	G>A	chr2:201057445	C>T
chr6:31430801	G>T	chr18:53252287	G>A	chr2:22598928	C>T	chr6:28092605	T>G	chr2:48573897	G>C
chr6:32442438	T>C	chr18:53250052	A>C	chr12:2354827	T>C	chr7:86782807	T>C	chr11:46486061	A>G
chr6:31198215	T>G	chr19:19431965	G>T	chr1:98474769	C>T	chr10:104941109	T>A	chr10:104877304	T>C
chr6:32943409	A>T	chr2:76310519	C>T	chr6:26421347	T>C	chr11:83178810	C>G	chr10:104950199	C>T
chr6:29856196	G>C	chr20:58252539	C>T	chr7:2156340	A>T	chr12:23466113	C>T	chr11:47207364	G>A
chr6:31430012	G>A	chr20:58252438	G>T	chr14:104011431	G>A	chr13:114897280	C>T	chr10:104840972	T>C
chr6:32659611	T>C	chr3:52566822	C>A	chr14:104021808	G>A	chr14:104028808	G>A	chr10:104940948	T>C
chr6:30762240	A>G	chr3:52567016	G>T	chr22:39872129	A>G	chr18:53126332	T>A	chr11:130745047	T>C
chr6:28678359	T>C	chr4:176731403	G>A	chr7:104584412	C>T	chr15:78857988	C>G	chr11:123395866	A>G
chr6:32442743	G>C	chr7:104947197	T>C	chr5:153640048	A>G	chr8:82421310	A>G	chr11:123395989	C>T
chr6:29760821	A>C	chr8:9539101	A>G	chr6:313022043	G>A	chr15:78857941	T>G	chr2:200915718	T>C
chr6:32152444	A>G	chr1:2374122	C>T	chr5:153640310	G>A	chr7:2145722	G>C	chr1:150971197	G>T
chr1:44086833	C>T	chr10:104891398	C>G	chr7:104583774	G>A	chr3:56846398	C>A	chr6:32636758	T>C
chr6:28701494	G>C	chr11:46704441	A>G	chr7:104583845	G>A	chr2:201049388	A>G	chr11:130745593	A>G
chr6:31864306	G>A	chr2:198171060	G>T	chr7:104585518	G>C	chr20:48123089	T>G	chr1:36637710	A>G
chr6:32194632	G>C	chr5:109078042	G>A	chr7:104585762	C>A	chr7:878572450	C>G	chr11:46373313	T>C
chr6:31213813	G>A	chr7:2019877	G>A	chr1:2392650	G>C	chr2:2048122764	T>C	chr2:2324044	T>C
chr6:31040722	T>G	chr7:2046832	C>T	chr1:2372860	T>C	chr12:124480868	G>A	chr2:200716121	C>A
chr6:29894394	G>T	chr7:1907011	G>A	chr10:106539064	G>A	chr2:233640873	T>G	chr10:104910961	G>A
chr6:32635503	A>G	chr8:38137532	A>G	chr6:31324417	A>C	chr8:68384582	C>G	chr2:23741258	G>C
chr6:30899526	T>C	chr8:38240010	A>G	chr6:29893984	C>T	chr7:86782799	T>C	chr12:124481692	A>G
chr6:33290827	A>G	chr8:38236596	C>G	chr6:29893962	T>C	chr18:53125940	T>C	chr12:99463546	C>A
chr6:31878435	G>A	chr6:32080193	G>C	chr6:29893924	A>T	chr14:104188959	G>A	chr10:104952501	C>T
chr6:32171077	A>G	chr2:22735643	T>C	chr6:31326705	T>C	chr6:28271975	C>A	chr18:53009099	A>G
chr6:31762846	G>T	chr2:22738511	A>G	chr6:31322792	T>A	chr15:405774741	A>G	chr14:103981664	G>A
chr6:29855064	A>G	chr10:18804691	T>C	chr14:71374704	G>A	chr1:150058807	G>A	chr14:103853362	A>C
chr6:31240218	A>C	chr16:29937654	G>A	chr6:28411222	T>G	chr3:52878602	G>C	chr14:59951448	C>T
chr6:29631993	A>T	chr1:84284851	G>A	chr6:28411305	A>G	chr6:26089342	G>C	chr14:35513762	A>T
chr6:32224141	G>A	chr1:243645205	C>G	chr6:29645040	C>A	chr6:26474573	C>T	chr14:104157580	C>T
chr6:31346655	A>G	chr1:73542699	T>C	chr3:52845107	T>C	chr8:38243229	T>A	chr16:29931595	T>C
chr6:32635631	G>A	chr10:18804691	T>C	chr3:180891517	A>G	chr11:46695485	G>A	chr16:89541245	T>A
chr6:31589678	A>G	chr11:83077925	A>C	chr6:31323418	G>C	chr6:27301514	T>A	chr17:17728045	A>G
chr6:32408499	A>G	chr12:29935398	C>G	chr6:31321269	A>C	chr6:27640248	G>A	chr17:17698256	G>A
chr6:29759825	T>C	chr14:104095322	G>A	chr1:37001315	A>C	chr11:133819419	G>T	chr17:19912712	T>C
chr6:31805915	C>T	chr14:99727088	G>A	chr3:17302560	A>G	chr6:262864740	T>C	chr17:19211075	C>G
chr6:31609274	T>C	chr14:104008161	C>T	chr3:17235492	G>A	chr1:177275652	G>A	chr17:17831718	G>T
chr6:31698090	T>G	chr14:103883514	A>G	chr3:17238675	G>A	chr15:178833038	C>A	chr17:17744727	T>C
chr6:28806220	G>A	chr14:104017795	G>A	chr3:17245272	C>T	chr2:73460519	T>G	chr18:53075294	A>C
chr6:30857896	G>A	chr14:104035294	G>A	chr3:17259111	A>G	chr11:130764411	T>C	chr18:53449669	A>G
chr6:32807492	C>T	chr14:99712796	C>T	chr3:172207191	G>A	chr17:19141584	C>G	chr6:26226034	C>G
chr6:28908216	A>T	chr15:85201421	T>C	chr3:17221851	C>A	chr12:18755564	C>T	chr22:42475570	A>G
chr6:31164781	T>C	chr15:85114270	T>C	chr3:180700152	A>G	chr18:53014635	C>T	chr10:104436643	C>T
chr6:28584777	A>G	chr15:91427614	G>A	chr2:198362853	C>A	chr3:52322419	G>A	chr16:89541245	T>A
chr6:31434200	T>C	chr15:85114449	C>T	chr14:99726002	G>C	chr1:177273233	T>A	chr17:17728045	A>G
chr6:32410693	G>A	chr16:29934052	A>G	chr15:78730254	A>G	chr1:177273233	T>A	chr17:17698256	G>A
chr6:30112797	G>A	chr17:17726967	C>T	chr1:98520221	T>A	chr17:187875344	A>G	chr17:19912712	T>C
chr6:28891178	T>C	chr17:55743279	G>A	chr1:98513146	T>A	chr11:46697144	C>T	chr17:19211075	C>G
chr1:1164685316	C>T	chr17:19204434	A>C	chr1:98512129	G>T	chr5:153666792	C>T	chr17:17831718	G>T
chr6:30796547	T>C	chr17:19204533	T>A	chr2:233736246	C>T	chr14:104030903	T>C	chr17:17744727	T>C
chr6:31765866	T>A	chr17:17753830	C>A	chr3:528135874	A>G	chr18:53215741	A>T	chr18:53075294	A>C
chr6:30746333	A>G	chr17:17741671	T>C	chr2:58131997	T>C	chr11:129954502	A>C	chr18:53449669	A>G
chr6:32602700	C>T	chr17:17726650	A>G	chr2:200697541	A>T	chr7:2046330	A>G	chr2:58135025	T>C
chr6:31127564	C>T	chr18:53253412	G>C	chr2:58136222	C>G	chr1:150124617	T>C	chr1:2381570	A>G
chr6:30803528	A>G	chr18:53215144	T>A	chr1:36627193	T>C	chr12:123550815	A>G	chr6:28110150	C>T
chr6:31154495	G>A	chr2:57987203	C>G	chr1:36622872	C>G	chr6:26365681	T>G	chr7:2047847	A>G
chr6:31176950	C>T	chr2:86974320	A>G	chr3:172201010	C>T	chr1:16446017	T>C	chr7:2047877	G>A
chr6:31582027	G>C	chr2:233725366	G>A	chr3:17302357	T>C	chr1:1557480625	C>A	chr11:46370769	A>G
chr6:31184526</									

chr6:32936996	G>A	chr6:25993970	T>C	chr6:29648379	T>C	chr14:104061286	G>A	chr20:37455271	C>T
chr6:29895500	C>T	chr6:27143885	T>C	chr20:37376810	G>A	chr6:26054173	T>A	chr14:103984539	T>C
chr6:31348042	T>C	chr6:26044375	C>T	chr20:37376986	G>C	chr8:38299626	T>C	chr9:7172500	C>A
chr6:32114517	A>T	chr6:26122650	G>C	chr6:31786965	G>A	chr8:38240500	G>A	chr11:133817011	C>T
chr6:29764658	G>A	chr6:28174759	G>A	chr6:28234599	C>T	chr20:37434135	G>A	chrX:153626651	A>C
chr6:31319031	C>T	chr6:25874425	G>A	chr6:28235178	A>C	chr20:43688154	A>G	chrX:153626740	G>A
chr6:32224465	T>C	chr6:28323704	G>A	chr1:98549812	A>G	chr14:35514458	G>C	chr15:78731810	T>A
chr6:30687616	G>C	chr6:27834087	A>G	chr2:200819796	G>A	chr12:123849274	C>G	chr6:31514249	A>G
chr6:31025363	C>G	chr6:27834141	C>G	chr2:200820507	T>G	chr11:64692314	A>G	chr22:41418156	C>T
chr6:32765587	G>T	chr6:26122959	G>A	chr2:200847992	G>A	chr5:60699575	C>A	chr6:32552241	C>T
chr6:31348079	G>A	chr6:26057783	A>C	chr2:233644013	A>G	chr20:37455372	C>A	chr6:32554131	T>C
chr6:32224491	G>A	chr6:25993471	A>G	chr6:26030046	A>G	chr18:53014094	A>T	chr6:32561426	C>T
chr6:31176248	C>T	chr6:26189358	G>C	chr6:26365588	T>C	chr20:58252409	T>C	chr6:32577226	C>G
chr6:31318787	C>T	chr6:26119247	C>T	chr12:123717218	G>A	chr20:58252901	G>A	chr6:32577457	A>G
chr6:28911804	G>A	chr6:25901135	C>T	chr12:123717356	C>A	chr3:52273423	G>A	chr6:32577472	T>G
chr6:31177374	T>C	chr6:27520754	T>A	chr14:35591318	G>A	chr18:52968232	A>G	chr6:32577476	T>G
chr6:31040925	A>G	chr6:27265942	T>A	chr14:35625219	G>T	chr11:64692358	C>T	chr6:32603941	G>T
chr6:32224447	A>G	chr6:27688843	A>C	chr12:123850170	A>G	chr22:50312681	A>G	chr6:32604234	C>T
chr6:32904053	C>T	chr6:25983012	G>A	chr12:123849105	A>G	chr6:27124906	G>C	chr6:32604294	C>A
chr6:30850115	G>A	chr6:27806987	A>G	chr10:104574644	T>A	chr20:20828828	G>A	chr6:32608195	T>C
chr6:31430754	G>A	chr6:27647511	G>T	chr10:104574564	T>G	chr11:46923109	C>T	chr6:32623195	G>A
chr6:30782207	G>C	chr6:28306673	T>C	chr10:104574424	A>G	chr3:117640662	C>T	chr6:32623225	A>C
chr6:33048541	T>A	chr6:27136227	A>G	chr12:123850199	A>C	chr6:26090027	T>C	chr6:32623335	A>G
chr6:38126361	G>C	chr6:26055370	G>A	chr22:5032326	T>C	chr6:26305625	C>T	chr6:32623373	C>T
chr6:31040300	C>G	chr6:26441642	C>T	chr16:30134658	T>C	chr18:53014634	C>G	chr6:32623437	G>A
chr6:32148033	A>G	chr6:27356529	T>C	chr12:123873244	C>T	chr10:104596983	T>A	chr6:32631960	C>T
chr6:33048356	A>T	chr7:2048222	G>C	chr18:527617333	T>C	chr14:104096033	C>T	chr6:32632338	G>A
chr6:30738448	G>A	chr8:10255107	C>G	chr8:60557138	C>T	chr10:104596962	C>A	chr6:32632662	T>G
chr6:31125980	A>T	chr8:10257680	T>C	chr6:31326076	A>T	chr6:26272550	G>T	chr6:32632705	G>A
chr6:28891841	G>A	chr20:20830614	T>C	chr12:123882979	G>A	chr1:98513847	G>A	chr6:32632755	C>T
chr6:30899053	G>C	chr13:35194024	C>T	chr12:123883855	T>C	chr1:98513961	A>G	chr6:32632852	G>A
chr6:32223949	T>A	chr17:17740283	C>T	chr17:19175319	A>G	chr1:98385675	T>A	chr6:32632887	G>T
chr6:29757997	T>C	chr8:38269516	T>C	chr22:42466982	A>T	chr1:98386957	A>C	chr6:32632905	G>A
chr6:30789649	G>C	chr22:42465790	G>A	chr12:123849923	T>C	chr1:98404098	T>C	chr6:32633085	C>T
chr6:28757557	C>G	chr22:42465943	G>A	chr12:123849102	A>G	chr11:46368243	A>T	chr6:32633176	C>T
chr6:29725297	G>T	chr22:42466511	C>G	chr5:60712749	C>A	chr11:46368561	G>T	chr6:32633233	G>C
chr6:32161854	G>A	chr22:42466907	C>T	chr18:53446376	G>A	chr11:46479482	C>T	chr6:32633284	T>C
chr6:31589266	C>G	chr22:42466952	C>T	chr12:123849053	A>C	chr11:46639584	G>A	chr6:32633377	A>G
chr6:32443360	A>G	chr22:42467697	G>C	chr12:123875527	T>C	chr11:46684526	A>T	chr6:32633394	C>G
chr6:28924628	T>A	chr2:233734822	C>T	chr12:123849776	C>A	chr11:46705976	T>A	chr6:32633428	G>C
chr6:31240295	C>T	chr2:233736026	G>A	chr17:19172507	A>G	chr11:47198640	C>T	chr6:32633468	A>C
chr6:28913788	T>G	chr2:198493705	T>A	chr19:50087878	T>C	chr11:46922598	C>T	chr6:32633699	C>A
chr6:28250238	C>T	chr6:28103693	C>T	chr10:104261361	C>T	chr11:10744475	G>A	chr6:32633751	A>G
chr6:28243044	G>A	chr6:159076100	G>A	chr19:19495956	G>A	chr12:123451020	G>C	chr6:32633755	C>T
chr6:29892480	A>G	chr6:28305606	G>A	chr19:19516435	A>G	chr15:815196246	G>A	chr6:32633881	A>G
chr6:29609923	A>G	chr16:4549228	C>G	chr8:8168396	G>A	chr19:2686315	G>A	chr6:32633930	G>A
chr6:32059869	A>G	chr2:198172150	T>G	chr2:200715390	T>G	chr3:71547592	A>G	chr6:32633943	A>G
chr6:32050760	C>T	chr5:60774229	G>A	chr17:17941366	A>G	chr5:60139983	A>G	chr6:32634026	T>A
chr6:32038552	C>T	chr6:31198069	G>A	chr6:189524925	T>C	chr6:25371371	A>T	chr6:32634128	G>A
chr6:32146494	C>A	chr6:29765867	C>T	chr1:189541797	G>A	chr6:263055055	T>C	chr6:32634245	G>A
chr6:29759312	G>A	chr6:31786874	A>G	chr9:131580746	C>T	chr6:56558387	G>A	chr6:32634375	C>T
chr6:31154635	T>C	chr6:32601979	A>C	chr1:8469339	A>G	chr11:64685484	G>A	chr6:32634663	A>G
chr6:29764365	C>T	chr6:32443133	C>T	chr1:8469286	G>C	chr6:25885816	C>T	chr6:32634892	G>A
chr6:28892921	A>T	chr6:30722269	C>A	chr1:8486374	A>T	chr11:83279210	C>T	chr6:32635013	G>A
chr6:32361823	C>T	chr6:31624866	A>G	chr1:8468280	A>G	chr11:83279184	A>G	chr6:32635083	A>G
chr6:29764474	C>T	chr6:30703484	G>T	chr1:8467355	G>A	chr12:199462371	A>G	chr6:32635094	C>A
chr6:30851911	C>A	chr6:31232729	C>T	chr1:8498682	A>T	chr11:64684285	C>T	chr6:32635221	G>A
chr6:30838499	T>C	chr6:28805640	G>T	chr1:8481018	T>G	chr3:17418535	T>C	chr6:32635368	C>T
chr6:30907337	T>C	chr6:31214313	A>G	chr1:8482080	C>T	chr3:52880130	G>T	chr6:32635689	T>C
chr6:32605937	G>T	chr6:32441936	T>C	chr1:8484825	A>G	chr5:60712214	G>A	chr6:32635906	T>A
chr6:31434333	A>T	chr6:31671559	C>A	chr6:32623152	T>C	chr5:45534425	C>G	chr6:32635967	G>T
chr6:31239857	C>G	chr6:28885867	A>G	chr3:180707963	C>T	chr6:26477781	G>A	chr6:32635992	T>C
chr6:32987890	A>G	chr6:31801235	C>T	chr8:10226357	C>T	chr6:28305814	C>A	chr6:32636000	G>T
chr6:29856441	T>G	chr14:99733956	C>T	chr6:31465049	G>T	chr6:28306139	C>A	chr6:32636023	G>A
chr6:28716347	A>G	chr6:31323508	C>T	chr6:31082989	A>G	chr11:46657432	T>G	chr6:32636095	G>C
chr6:32450699	G>A	chr6:32153408	C>A	chr6:310813436	A>G	chr11:46927879	C>T	chr6:32636192	G>A
chr6:31348367	G>A	chr2:25599174	C>A	chr6:31082934	C>T	chr12:2338860	A>G	chr6:32636237	C>T
chr6:29732746	T>C	chr6:31832040	A>G	chr6:31082478	C>T	chr16:9915015	A>G	chr6:32636256	T>C
chr6:28983889	A>G	chr6:30851991	G>A	chr2:200792500	G>T	chr6:4502412	G>A	chr6:32636568	G>C
chr6:31745466	G>A	chr6:28302786	A>G	chr6:31705866	C>T	chr6:4502516	C>T	chr2:233639750	G>T
chr6:30644139	T>C	chr6:31239683	A>T	chr6:33282004	C>T	chr16:4573940	A>G	chr6:25450028	C>A
chr6:30894967	T>A	chr1:243669532	G>C	chr6:33280631	C>G	chr7:17746743	A>G	chr6:27689504	G>A
chr6:31143654	C>G	chr6:28242796	A>C	chr3:180893579	C>T	chr6:27730066	T>C	chr6:32636237	C>T
chr6:32155000	C>T	chr16:9914016	A>G	chr6:31718398	C>G	chr17:17847938	G>A	chr6:32636256	T>C
chr6:30747639	C>T	chr16:9914135	G>T	chr6:29648201	C>T	chr6:28305814	C>A	chr6:32636568	G>C
chr6:30780570	C>T	chr16:9914421	C>G	chr6:29645939	C>T	chr6:28306139	C>A	chr6:32636023	G>A
chr6:30063676	G>C	chr5:60791595	C>G	chr6:29645615	C>A	chr11:46657432	T>G	chr6:32636095	G>C
chr6:32224321	G>T	chr6:30761091	G>A	chr6:33282183	A>G	chr11:46927879	C>T	chr6:32636192	G>A
chr6:30787764	C>A	chr6:31781769	C>T	chr6:31715884	T>C	chr12:2338860	A>G	chr6:32636237	C>T
chr6:28852880	G>A	chr2:19815813	T>G	chr6:31094705	C>T	chr16:9915015	A>G	chr6:32636256	T>C
chr6:32408844	A>G	chr2:19815815	G>C	chr6:31462137	G>A	chr6:4502412	G>A	chr6:32636568	G>C
chr6:32339078	A>G	chr6:32601888	A>G	chr6:31081942	T>C	chr6:27975396	T>C	chr6:32636023	G>T
chr6:29760607	G>T	chr6:29628096	C>T	chr6:31085358	C>T	chr8:27438308	A>C	chr6:27688929	G>A
chr6:30731403	A>T	chr2:198178669	G>C	chr6:31087465	C>A	chr16:29974169	G>A	chr6:26286746	C>T
chr6:31443036	G>C	chr6:31198045	T>C	chr6:31704296	G>T	chr18:53445932	T>C	chr6:26363690	C>T
chr6:32224441	G>A	chr14:104060383	C>T	chr6:31712198	G>A	chr6:111872484	T>C	chr6:26363757	G>C
chr6:30893943	G>A	chr6:29610437							

chr6:29629346	T>C	chr14:72447997	C>T	chr1:214162575	A>T	chr6:26107090	C>T	chr6:26444734	C>T
chr6:29610283	C>A	chr6:29645772	A>G	chr1:214160550	A>T	chr6:26365768	T>C	chr6:26987902	A>T
chr6:31176228	T>C	chr14:72448084	A>G	chr1:214160531	G>A	chr6:27865904	A>T	chr6:27145313	C>T
chr6:31600853	G>T	chr2:198478322	G>A	chr1:124480803	A>G	chr1:2375846	T>G	chr6:28048537	A>G
chr6:32439755	T>A	chr6:30750927	T>A	chr6:26478254	C>T	chr17:19014833	G>C	chr6:28048540	A>G
chr6:31175948	A>C	chr3:17301089	G>C	chr6:26328355	T>C	chr1:2372399	A>G	chr6:28104478	C>T
chr15:91428292	C>T	chr3:17221019	T>C	chr6:28102380	C>T	chr6:26489868	A>T	chr6:28104636	G>T
chr6:32410217	T>C	chr3:17221254	C>A	chr6:26336574	G>A	chr1:204968789	T>C	chr6:28203302	A>G
chr14:103852166	C>G	chr6:30708697	C>T	chr14:104027597	T>G	chr2:201244263	A>C	chr6:114711097	T>C
chr6:32862960	C>G	chr17:19174876	T>C	chr1:2385238	T>C	chr2:18182272	G>C	chr6:28863266	C>G
chr6:28863266	C>G	chr6:31606400	C>G	chr3:52232799	A>G	chr6:27798889	C>G	chr6:32223772	T>C
chr6:32223772	T>C	chr6:28885865	A>G	chr3:53078592	C>A	chr2:220041930	C>T	chr6:26361013	G>A
chr6:29765049	A>T	chr17:19220668	G>A	chr14:104004970	C>T	chr2:22738727	A>T	chr6:26361502	G>A
chr6:33290549	T>A	chr6:30859346	G>A	chr12:57547751	T>C	chr2:172543915	G>T	chr6:26365149	C>T
chr6:29894213	C>T	chr20:37468447	T>C	chr6:28193104	C>A	chr2:201059373	C>A	chr6:26367835	A>G
chr6:29415466	C>A	chr6:105389955	G>A	chr6:264886909	C>T	chr6:271103656	G>A	chr6:27114054	G>T
chr6:31940899	A>G	chr6:32442298	C>T	chr6:27357416	C>G	chr6:28304386	A>C	chr6:28109264	T>C
chr6:32904041	T>A	chr6:31165220	C>T	chr6:25833012	T>C	chr5:152895821	A>T	chr6:28190832	T>A
chr6:31112739	C>A	chr6:31153022	T>C	chr2:200916405	C>T	chr6:25961354	C>T	chr6:33399780	C>T
chr6:32602762	A>C	chr6:32636615	T>C	chr1:207993639	A>G	chr2:76397510	A>T	chr6:25343446	T>A
chr6:32146001	C>G	chr3:181178011	G>A	chr6:26022650	A>C	chr6:72140868	T>A	chr14:104201871	C>G
chr6:31198368	G>A	chr6:28756904	C>T	chr1:180630193	A>G	chr6:25866234	T>A	chr14:104189863	G>A
chr6:30899652	G>A	chr6:30761134	C>T	chr6:25714961	A>G	chr6:26196595	A>G	chr14:35592540	C>G
chr6:29893141	G>A	chr6:31138109	G>A	chr5:139064837	C>T	chr2:76246647	T>G	chr11:57510296	A>G
chr6:28275806	G>C	chr6:28651093	T>G	chr6:25914855	G>A	chr11:130759412	G>A	chr1:2379707	A>G
chr6:31125779	C>A	chr10:104953046	A>G	chr6:26361987	A>G	chr20:43682553	T>G	chr10:104839154	T>C
chr6:31626015	C>T	chr14:72448164	G>A	chr5:87686128	A>C	chr3:36884372	C>A	chr6:33750088	A>G
chr6:31172153	T>C	chr1:30446086	A>G	chr5:60616028	T>G	chr6:25931022	G>A	chr1:50579239	A>C
chr6:31164903	T>G	chr1:30445993	C>T	chr18:53252891	C>T	chr3:44062009	G>A	chr1:98492464	C>T
chr6:30723171	A>G	chr6:274411725	T>A	chr14:99731733	A>C	chr14:104060068	A>T	chr6:25836929	C>G
chr6:30721145	C>T	chr6:31166354	A>G	chr14:104009632	A>G	chr3:136377379	T>C	chr6:27719377	G>T
chr6:29765718	T>A	chr6:32409062	G>A	chr6:26173480	T>G	chr1:150059071	A>G	chr6:33395201	G>A
chr6:32659647	A>G	chr6:31154436	C>T	chr6:27805257	A>C	chr1:245159107	T>G	chr6:25873748	C>A
chr6:31619578	A>G	chr14:104060386	G>A	chr6:27783943	C>T	chr6:26365761	T>C	chr6:26312172	G>C
chr6:31213904	T>C	chr6:31606394	G>A	chr5:60139550	C>T	chr6:25866245	A>T	chr6:26319488	C>T
chr6:30900152	G>T	chr2:225403617	G>A	chr10:104942246	T>G	chr15:78858402	C>G	chr6:26327907	G>C
chr6:29759925	T>C	chr3:36861198	T>G	chr1:150242342	G>A	chr12:111052467	T>A	chr6:26327958	T>C
chr6:28706914	G>A	chr17:19153177	G>C	chr6:26035808	G>A	chr5:137784517	A>G	chr6:26332607	A>C
chr6:30688577	G>C	chr14:99733386	G>C	chr6:27459925	G>A	chr5:153621017	G>A	chr6:27064277	C>T
chr6:29720999	G>A	chr7:110932741	G>A	chr6:28250181	T>C	chr5:60588574	C>G	chr6:27740668	C>T
chr6:31632136	C>A	chr1:23815451	C>T	chr14:104027997	C>T	chr5:158613791	G>A	chr1:27745721	T>G
chr6:31587563	G>C	chr12:123738680	A>G	chr5:87987691	A>G	chr5:152014668	T>C	chr6:28271200	C>T
chr6:31620522	T>G	chr2:201024514	T>C	chr6:26199905	C>T	chr5:152013318	G>C	chr6:28319109	A>G
chr6:32223927	T>C	chr6:28351730	C>T	chr6:275121858	A>G	chr11:133817227	C>T	chr6:28364951	G>A
chr6:33324529	G>T	chr3:17259759	T>C	chr6:281876734	T>G	chr11:133817225	A>G	chr12:123634045	T>C
chr6:32117973	T>G	chr3:17260190	C>T	chr14:104042755	A>G	chr5:60589023	T>G	chr12:123595165	G>A
chr6:29766160	A>G	chr1:8484230	C>T	chr6:28305510	G>A	chr5:60615888	A>G	chr15:78878543	G>A
chr6:31433833	A>G	chr14:104161283	G>A	chr6:27759117	G>C	chr11:133817039	C>T	chr17:19142228	C>T
chr6:32913248	C>T	chr12:123735939	G>A	chr6:26067362	A>G	chr5:152015210	G>A	chr22:41888114	G>A
chr6:31214356	G>A	chr12:123650337	C>T	chr5:159061822	A>G	chr6:26322155	C>G	chr22:41418299	T>C
chr6:31125707	C>G	chr12:123717364	A>G	chr6:26562271	G>T	chr6:26322117	C>T	chr22:41613190	C>T
chr6:28554545	G>T	chr6:29893405	C>G	chr6:27799516	C>T	chr6:26322528	A>G	chr22:41613305	C>T
chr6:31165102	A>C	chr6:28221266	G>A	chr6:56571462	T>G	chr6:28322122	C>G	chr22:41698544	C>G
chr6:30746521	G>T	chr6:32552094	A>T	chr2:145175860	A>G	chr6:28322298	A>G	chr22:50311975	A>G
chr6:30173540	G>A	chr1:2391254	A>C	chr22:46473847	C>A	chr6:32556009	A>T	chr8:10267542	C>A
chr6:31197076	C>T	chr1:177271315	G>A	chr6:26026165	C>G	chr6:26322863	C>T	chr10:104669623	T>C
chr6:28851574	T>C	chr1:177276008	T>C	chr6:26562271	G>T	chr6:32604223	A>G	chr11:46935187	G>C
chr1:150335536	A>G	chr6:56574182	T>G	chr4:170585508	A>G	chr6:40828287	C>T	chr5:140174624	G>A
chr1:150335538	A>G	chr6:26017544	T>C	chr1:831622	T>C	chr6:26327967	T>G	chr5:139065590	T>G
chr6:28852453	T>C	chr6:26124305	T>C	chr6:26534280	C>T	chr6:27115788	G>A	chr1:50180268	C>G
chr6:31239870	G>C	chr6:26477145	T>A	chr1:73542542	A>G	chr6:27158827	G>C	chr1:44087143	G>A
chr6:32352688	A>G	chr8:89340164	G>C	chr1:245139729	C>T	chr6:28325203	C>G	chr1:22852294	G>A
chr6:32224390	C>A	chr6:28414969	G>T	chr6:27799522	T>A	chr6:2823510	C>T	chr1:243672002	C>G
chr6:29707269	G>T	chr6:26555435	C>A	chr6:28091582	A>G	chr6:26313350	C>T	chr16:991206	C>G
chr6:32936122	C>T	chr17:19163054	A>C	chr15:91427874	G>C	chr6:26491236	A>G	chr2:233638353	G>A
chr6:29875994	G>A	chr17:19203034	A>G	chr6:28234830	C>T	chr6:27115788	G>A	chr19:50180268	C>G
chr6:30782004	C>T	chr22:41050988	C>G	chr1:73542542	A>G	chr6:26478827	G>C	chr1:44087143	G>A
chr6:31129312	C>T	chr2:76246434	A>G	chr17:1764063	C>G	chr6:11845572	T>A	chr12:122852294	G>A
chr6:31239754	C>T	chr1:98526169	T>C	chr2:76310282	G>A	chr6:27248933	T>C	chr2:172579956	A>G
chr1:1504264111	C>G	chr1:98524962	A>G	chr20:39764704	T>C	chr6:32604553	C>T	chr2:172330774	G>A
chr6:29766087	T>C	chr3:52593121	A>G	chr6:26105371	C>T	chr6:26534026	G>T	chr7:110904228	A>G
chr1:2379250	G>A	chr3:52593140	A>G	chr14:104025895	T>C	chr6:26323868	G>A	chr3:52931973	T>G
chr6:29759826	T>C	chr10:104386936	T>C	chr6:27739586	T>A	chr6:26328464	G>C	chr3:135914478	G>A
chr6:31765986	G>A	chr10:104391287	C>T	chr6:26044866	T>C	chr6:25931959	A>C	chr3:135914717	G>A
chr6:30735981	C>T	chr10:104391547	C>T	chr10:104941114	A>T	chr6:28106143	T>C	chr3:135524432	C>T
chr1:150039049	G>A	chr10:104591395	G>T	chr6:27145343	C>T	chr7:104837011	G>A	chr3:124079274	A>C
chr6:30709359	G>A	chr10:104681145	A>G	chr6:27219493	C>G	chr7:104652673	G>A	chr3:114171392	T>C
chr6:32223420	C>T	chr12:123716932	A>T	chr2:200775127	T>C	chr11:133817811	C>T	chr17:177444411	G>A
chr6:29892817	A>G	chr12:123716515	C>T	chr14:99712034	G>A	chr7:104618320	C>T	chr17:2084567	G>T
chr6:30873607	C>A	chr12:123635098	C>T	chr5:139071195	C>T	chr7:104842962	T>C	chr17:2126005	A>G
chr6:31164851	G>A	chr12:123632932	G>A	chr6:28191290	G>T	chr8:10193774	C>G	chr17:17753848	A>T
chr6:31762845	G>C	chr12:123633384	T>C	chr6:56571702	C>G	chr8:8168224	G>A	chr17:2115924	C>G
chr1:243910069	C>T	chr12:123718881	C>G	chr11:123396676	G>C	chr8:38243867	T>G	chr17:2076354	G>A
chr6:32131974	G>C	chr12:123718303	C>T	chr6:28318333	G>T	chr8:9607237	C>G	chr17:17727945	T>A
chr6:29856447	C>T	chr17:177271799	C>T	chr14:103981944	A>G	chr8:10283750	T>C	chr17:17733762	C>T
chr6:28916596	G>A	chr2:22735752	G>A	chr11:57547002	T>C	chr2:198668753	T>C	chr17:1783	

### 8.3 Fragments of the eQTLlib

All control fragments were included (8.1). SNP exchanged was at position 81 of 160 nt.

Start location (hg19)	SNP	Start location (hg19)	SNP	Start location (hg19)	SNP	Start location (hg19)	SNP	Start location (hg19)	SNP
chr1:111324727	T>C	chr1:243782745	D>I23	chr12:123750748	C>G	chr2:73713500	T>C	chr6:30650026	A>G
chr1:111356841	T>C	chr1:243789305	A>T	chr12:123750895	A>G	chr2:73713741	T>G	chr6:30687614	C>G
chr1:115618781	T>G	chr1:243791312	C>G	chr12:123752419	T>C	chr2:73788026	A>G	chr6:30688575	C>G
chr1:115619634	T>C	chr1:243792591	C>G	chr12:123752637	A>G	chr2:73788126	A>G	chr6:30689001	A>T
chr1:115626392	C>G	chr1:243792781	A>G	chr12:123753492	C>G	chr2:76384145	C>G	chr6:30699022	A>G
chr1:115629564	A>G	chr1:243795607	A>G	chr12:123757144	T>G	chr2:76396834	A>C	chr6:30717258	T>C
chr1:115629793	C>G	chr1:243802639	A>T	chr12:123757861	A>G	chr2:76397508	A>T	chr6:30721143	T>C
chr1:149998923	T>G	chr1:243803714	A>G	chr12:123758235	C>G	chr2:86947783	T>C	chr6:30723781	A>G
chr1:149999059	T>C	chr1:243803922	A>G	chr12:123771015	A>G	chr20:20817947	T>C	chr6:30731401	A>T
chr1:149999764	A>T	chr1:243803979	T>G	chr12:123771032	A>G	chr20:20830612	T>C	chr6:30732858	T>C
chr1:150001233	T>C	chr1:243806438	A>G	chr12:123771162	T>C	chr20:37278107	A>T	chr6:30748164	C>G
chr1:150001470	C>G	chr1:243811634	T>G	chr12:123771475	A>G	chr20:37376808	A>G	chr6:30758664	A>C
chr1:150001632	T>C	chr1:243812680	A>G	chr12:123840951	A>C	chr20:37376984	C>G	chr6:30758848	A>T
chr1:150001899	A>G	chr1:243814039	C>G	chr12:123840906	C>G	chr20:37377139	T>C	chr6:30758857	T>C
chr1:150002293	T>C	chr1:243815781	A>G	chr12:123849100	A>G	chr20:37434133	A>G	chr6:30761089	A>G
chr1:150002436	T>C	chr1:243816636	T>C	chr12:123849103	A>G	chr20:37455269	T>C	chr6:30761132	T>C
chr1:150003117	T>C	chr1:243818260	T>C	chr12:123849272	C>G	chr20:37455370	A>C	chr6:30761487	T>C
chr1:150003513	A>G	chr1:243819070	A>G	chr12:123849774	A>C	chr20:37455588	A>T	chr6:30761572	T>C
chr1:150004133	T>G	chr1:243820128	A>G	chr12:123849921	T>C	chr20:37455995	T>C	chr6:30796545	T>C
chr1:150004348	T>C	chr1:243822570	T>C	chr12:123850168	A>G	chr20:37456205	T>C	chr6:30796659	A>G
chr1:150005225	A>C	chr1:243823711	T>C	chr12:123850197	A>C	chr20:37457106	A>G	chr6:30802465	T>C
chr1:150007105	A>G	chr1:243826674	T>C	chr12:123873242	T>C	chr20:37457184	C>G	chr6:30803526	A>G
chr1:150008584	A>G	chr1:243829549	A>G	chr12:123875525	T>C	chr20:37458009	T>G	chr6:30850113	A>G
chr1:150008790	T>C	chr1:243833397	A>C	chr12:123882977	A>G	chr20:37461311	C>G	chr6:30851909	A>C
chr1:150009216	A>C	chr1:243838630	T>C	chr12:123883853	T>C	chr20:37461354	C>G	chr6:30851989	A>G
chr1:150009705	T>G	chr1:243846271	A>G	chr12:123886384	T>C	chr20:37466529	A>G	chr6:30855211	T>G
chr1:150009841	A>G	chr1:243856681	A>G	chr12:123886500	A>G	chr20:37485233	T>C	chr6:30857894	A>G
chr1:150010556	A>G	chr1:243868398	T>C	chr12:124428162	A>T	chr20:37485380	A>T	chr6:30873605	A>C
chr1:150011145	A>C	chr1:243875989	A>G	chr12:124480801	A>G	chr20:37485458	A>G	chr6:30882415	T>C
chr1:150011323	A>C	chr1:243878503	A>G	chr12:124480866	A>G	chr20:37487150	T>G	chr6:31022929	A>G
chr1:150011362	T>G	chr1:243882064	T>C	chr12:124481690	A>G	chr20:37491375	T>G	chr6:31023148	A>G
chr1:150012168	A>T	chr1:243883877	T>G	chr12:124482256	T>C	chr20:37492316	T>C	chr6:31023186	A>C
chr1:150012882	T>C	chr1:243891570	T>C	chr12:13688237	C>G	chr20:37492751	A>G	chr6:31023524	T>C
chr1:150012917	T>C	chr1:243897774	T>C	chr12:2288945	A>T	chr20:37493247	C>G	chr6:31023868	C>G
chr1:150013019	A>C	chr1:243908796	A>T	chr12:2289651	C>G	chr20:37493576	A>G	chr6:31024796	A>G
chr1:150013129	A>C	chr1:243909402	T>C	chr12:2306707	A>G	chr20:37493761	A>G	chr6:31025361	C>G
chr1:150013403	C>G	chr1:243910067	T>C	chr12:2307175	A>C	chr20:37498722	A>G	chr6:31026434	A>G
chr1:150013453	T>C	chr1:243911397	A>C	chr12:2307296	A>G	chr20:37502637	T>G	chr6:31026589	T>C
chr1:150013470	A>G	chr1:243919773	A>G	chr12:23121820	A>T	chr20:37503317	A>T	chr6:31036783	A>G
chr1:150013656	T>C	chr1:243922374	A>G	chr12:2321868	A>T	chr20:39764702	T>C	chr6:31038054	T>C
chr1:150014017	A>C	chr1:243924743	A>T	chr12:2322513	A>G	chr20:43682551	T>G	chr6:31038225	T>C
chr1:150014078	A>G	chr1:243932275	T>C	chr12:23240402	T>C	chr20:43689097	A>T	chr6:31038593	T>C
chr1:150014296	A>C	chr1:243939970	A>G	chr12:2348844	C>G	chr20:43690561	C>G	chr6:31038609	T>C
chr1:150015166	T>G	chr1:243943907	A>G	chr12:2349584	C>G	chr20:43696025	A>G	chr6:31038803	A>C
chr1:150015253	A>G	chr1:243945131	T>C	chr12:2350401	C>G	chr20:48042970	T>C	chr6:31038976	T>C
chr1:150015482	T>G	chr1:243946072	A>G	chr12:2350452	A>G	chr20:58252407	T>C	chr6:31039252	A>G
chr1:150015625	A>G	chr1:243946917	A>T	chr12:2350620	A>G	chr20:58252436	T>G	chr6:31040298	C>G
chr1:150015805	A>G	chr1:243953583	A>G	chr12:23639392	A>G	chr20:58252537	T>C	chr6:31040720	T>G
chr1:150016659	A>G	chr1:243956120	A>T	chr12:23639542	C>G	chr20:58252899	A>G	chr6:31040764	A>G
chr1:150017234	A>C	chr1:243959543	T>C	chr12:2374130	A>G	chr20:58253438	T>C	chr6:31047471	A>C
chr1:150017281	A>G	chr1:243963175	A>G	chr12:29935396	C>G	chr20:58253504	T>C	chr6:31047620	A>G
chr1:150017823	C>G	chr1:243963896	T>C	chr12:29936800	A>G	chr20:58253583	A>G	chr6:31081940	T>C
chr1:150018373	T>C	chr1:243967674	A>G	chr12:50594947	A>G	chr20:58293777	C>G	chr6:31082476	T>C
chr1:150018482	T>G	chr1:243975753	T>C	chr12:57497814	A>C	chr20:62133177	A>G	chr6:31082932	T>C
chr1:150018719	T>C	chr1:243976035	A>C	chr12:574990100	T>C	chr20:62153311	C>G	chr6:31082987	A>G
chr1:150018944	T>G	chr1:243983723	T>C	chr12:57569478	T>C	chr21:25101724	A>G	chr6:31085356	T>C
chr1:150019036	A>G	chr1:243983877	T>C	chr12:57824165	T>G	chr21:25105211	A>G	chr6:31085700	T>C
chr1:150019327	T>C	chr1:243985769	A>C	chr12:67595857	A>C	chr22:39869209	T>C	chr6:31094703	T>C
chr1:150019580	A>G	chr1:243990833	T>C	chr12:87669155	T>G	chr22:39897163	A>G	chr6:31105147	T>C
chr1:150020862	T>G	chr1:243994216	T>C	chr12:87699522	A>G	chr22:39930197	T>C	chr6:31105413	A>G
chr1:150021609	T>C	chr1:243995041	T>C	chr12:87703470	A>C	chr22:39942234	T>C	chr6:31125705	C>G
chr1:150022827	A>G	chr1:244002773	T>C	chr12:87705214	A>G	chr22:39988175	T>C	chr6:31125777	A>C
chr1:150023407	T>C	chr1:244015993	T>C	chr12:92254654	T>G	chr22:40087389	A>G	chr6:31125978	A>T
chr1:150024300	T>C	chr1:244017365	T>C	chr12:92256812	A>G	chr22:41047556	T>G	chr6:31164779	T>C
chr1:150024854	A>G	chr1:244017865	T>C	chr12:92256818	A>T	chr22:41047751	T>G	chr6:31164849	A>G
chr1:150025177	T>C	chr1:244023667	A>G	chr12:99444432	A>G	chr22:41050986	C>G	chr6:31164901	T>G
chr1:150025377	T>C	chr1:244023680	T>C	chr12:99450338	A>G	chr22:41060853	A>T	chr6:31165100	A>C
chr1:150025672	A>C	chr1:244023793	A>T	chr12:99475226	C>G	chr22:41215672	T>C	chr6:31165218	T>C
chr1:150025833	T>C	chr1:244023962	C>G	chr12:99481031	A>T	chr22:41252962	C>G	chr6:31175608	T>C
chr1:150026207	T>C	chr1:244024211	T>G	chr12:99481052	T>G	chr22:41418154	T>C	chr6:31175667	A>G
chr1:150026544	T>C	chr1:244024268	A>G	chr12:99481536	T>C	chr22:41418229	T>G	chr6:31175946	A>C
chr1:150026823	T>C	chr1:244024703	T>G	chr12:99481565	T>C	chr22:41418297	T>C	chr6:31176226	T>C
chr1:150027004	T>C	chr1:244025317	T>C	chr12:99545400	A>C	chr22:41487218	T>C	chr6:31176246	T>C

chr1:150027486	T>C	chr1:244025999	A>G	chr12:99547717	A>G	chr22:41592677	A>G	chr6:31176335	T>C
chr1:150029936	A>G	chr1:244100402	T>C	chr13:114897278	T>C	chr22:41617897	A>G	chr6:31176602	A>G
chr1:150031132	A>G	chr1:28885800	T>C	chr13:114900150	T>C	chr22:41618880	T>G	chr6:31176678	A>G
chr1:150031490	T>C	chr1:29088713	C>G	chr13:21951162	A>C	chr22:41619350	A>T	chr6:31176921	A>G
chr1:150032221	T>C	chr1:29110736	A>G	chr13:21985288	T>C	chr22:41637119	T>G	chr6:31182147	A>G
chr1:150037560	T>C	chr1:29136686	A>G	chr13:35194022	T>C	chr22:41642212	T>C	chr6:31182608	T>C
chr1:150039047	A>G	chr1:29138975	T>G	chr13:73419467	A>G	chr22:41644074	T>C	chr6:31196840	T>C
chr1:150039678	A>G	chr1:29141155	A>G	chr13:85272495	A>G	chr22:41644428	A>G	chr6:31196847	A>G
chr1:150043471	A>G	chr1:29170005	T>C	chr13:96149046	A>G	chr22:41645257	T>C	chr6:31196862	A>G
chr1:150043544	A>G	chr1:29170159	C>G	chr14:10385096	C>G	chr22:41652846	T>G	chr6:31196935	A>G
chr1:150046132	C>G	chr1:30412503	T>C	chr14:103852164	C>G	chr22:41653540	T>G	chr6:31197074	T>C
chr1:150046970	A>T	chr1:30413139	C>G	chr14:103853360	A>C	chr22:41656332	A>G	chr6:31197263	A>G
chr1:150047119	A>G	chr1:30413903	A>G	chr14:103883512	A>G	chr22:41656509	A>G	chr6:31197293	T>C
chr1:150047856	C>G	chr1:30414051	A>G	chr14:103898525	C>G	chr22:41657626	T>C	chr6:31197528	A>G
chr1:150048871	A>T	chr1:30414346	T>C	chr14:103899169	C>G	chr22:41658632	T>C	chr6:31235869	T>C
chr1:150049934	A>G	chr1:30417143	A>C	chr14:103949378	T>C	chr22:41661154	T>G	chr6:31236051	A>G
chr1:150049982	T>C	chr1:30419456	T>G	chr14:103955069	A>G	chr22:41684093	T>C	chr6:31236115	T>C
chr1:150050058	A>G	chr1:30420798	A>T	chr14:103959432	A>G	chr22:41685937	C>G	chr6:31236316	T>G
chr1:150050268	A>G	chr1:30422702	T>C	chr14:103967045	A>G	chr22:41697338	A>G	chr6:31236339	T>C
chr1:150050772	C>G	chr1:30427639	A>G	chr14:103980134	A>T	chr22:41698542	C>G	chr6:31236523	A>G
chr1:150056510	A>T	chr1:30428943	A>T	chr14:103984537	T>C	chr22:41703408	A>G	chr6:31236524	A>G
chr1:150058577	T>G	chr1:30429958	T>C	chr14:103984616	T>C	chr22:41703428	A>G	chr6:31238930	A>C
chr1:150058727	T>C	chr1:30430366	T>G	chr14:103991478	A>T	chr22:41703579	A>G	chr6:31239141	T>C
chr1:150058805	A>G	chr1:30431560	A>G	chr14:103992096	C>G	chr22:41704143	T>G	chr6:31239151	A>C
chr1:150059069	A>G	chr1:30432219	T>C	chr14:103996205	T>C	chr22:41802438	T>C	chr6:31239205	T>C
chr1:150062644	A>C	chr1:30433405	A>G	chr14:103999750	T>C	chr22:41830874	T>C	chr6:31239227	T>C
chr1:150064762	A>C	chr1:30433886	C>G	chr14:104000518	T>G	chr22:41831322	A>G	chr6:31239449	C>G
chr1:150065287	T>C	chr1:30433951	T>C	chr14:104004968	T>C	chr22:41888112	A>G	chr6:31239518	T>C
chr1:150065378	A>T	chr1:30434937	A>G	chr14:104006342	T>C	chr22:42315790	T>C	chr6:31239577	A>C
chr1:150066530	A>T	chr1:30436000	A>G	chr14:104006414	C>G	chr22:42321251	A>G	chr6:31239681	A>T
chr1:150068071	A>C	chr1:30437118	A>T	chr14:104006693	T>G	chr22:42324622	T>G	chr6:31239752	T>C
chr1:150072596	T>C	chr1:30437268	A>G	chr14:104006716	T>C	chr22:42330229	T>C	chr6:31239802	C>G
chr1:150073256	T>C	chr1:30447761	A>G	chr14:104008159	T>C	chr22:42331409	C>G	chr6:31239827	T>C
chr1:150073298	A>G	chr1:31829873	T>C	chr14:104009630	A>G	chr22:42333408	T>G	chr6:31239855	C>G
chr1:150074141	T>C	chr1:31831870	A>C	chr14:104011429	A>G	chr22:42335621	A>G	chr6:31239868	C>G
chr1:150074172	T>C	chr1:31833706	T>G	chr14:104017793	A>G	chr22:42337040	T>G	chr6:31240060	A>G
chr1:150074523	T>G	chr1:31854882	A>G	chr14:104018651	A>G	chr22:42339095	T>C	chr6:31240180	A>C
chr1:150074642	A>G	chr1:31857332	T>C	chr14:104021806	A>G	chr22:42339516	A>G	chr6:31240216	A>C
chr1:150074759	A>G	chr1:31858901	T>C	chr14:104027995	T>C	chr22:42430408	C>G	chr6:31240293	T>C
chr1:150074945	A>G	chr1:31859558	A>G	chr14:104028270	T>C	chr22:42340569	T>C	chr6:31324151	T>C
chr1:150074988	A>G	chr1:31859959	A>T	chr14:104028806	A>G	chr22:42430582	T>C	chr6:31324208	T>G
chr1:150076542	A>G	chr1:31861187	T>C	chr14:104029449	A>G	chr22:424343091	A>G	chr6:31324210	A>G
chr1:150076613	A>G	chr1:31864183	T>C	chr14:104029819	T>G	chr22:42363408	A>G	chr6:31324415	A>C
chr1:150076663	A>G	chr1:36616337	T>C	chr14:104035292	A>G	chr22:42364997	T>C	chr6:31324586	T>C
chr1:150076808	T>C	chr1:36621186	C>G	chr14:104042753	A>G	chr22:42375761	T>G	chr6:31324829	C>G
chr1:150077069	T>C	chr1:36622870	C>G	chr14:104060381	T>C	chr22:42395242	C>G	chr6:31324830	A>G
chr1:150077266	A>G	chr1:36622903	A>G	chr14:104060392	A>G	chr22:42466509	C>G	chr6:31324887	C>G
chr1:150079777	C>G	chr1:36662191	T>C	chr14:104061284	A>G	chr22:42466905	T>C	chr6:31324888	T>G
chr1:150080113	A>C	chr1:366627542	T>C	chr14:104095320	A>G	chr22:42466950	T>C	chr6:31324892	C>G
chr1:150080265	T>C	chr1:366632051	T>C	chr14:104096031	T>C	chr22:42470317	T>C	chr6:31324895	C>G
chr1:150080659	A>C	chr1:366644144	A>G	chr14:104154878	A>G	chr22:42470589	C>G	chr6:31324931	A>C
chr1:150081568	A>G	chr1:36667708	A>G	chr14:104156108	C>G	chr22:42470608	C>G	chr6:31324953	T>C
chr1:150085959	A>G	chr1:41832297	T>G	chr14:104157578	T>C	chr22:42471774	T>C	chr6:31325004	A>G
chr1:150087824	T>G	chr1:44029353	C>G	chr14:104160141	T>C	chr22:42475568	A>G	chr6:31325049	A>G
chr1:150088955	T>G	chr1:44037124	T>C	chr14:104161281	A>G	chr22:42475703	T>G	chr6:31325056	T>C
chr1:150089307	A>C	chr1:44037685	A>G	chr14:104165927	T>C	chr22:42486723	A>G	chr6:31325090	T>C
chr1:150089914	T>C	chr1:44040427	A>T	chr14:104170796	T>C	chr22:42525952	A>C	chr6:31325092	A>G
chr1:150090738	A>G	chr1:44045919	A>G	chr14:104178186	A>G	chr22:42526484	A>C	chr6:31325341	C>G
chr1:150093628	T>C	chr1:44047900	C>G	chr14:104188127	A>G	chr22:42527471	T>C	chr6:31325380	T>C
chr1:150094532	T>C	chr1:44049266	T>G	chr14:104189857	A>G	chr22:42538704	T>G	chr6:31325424	T>G
chr1:150095754	A>G	chr1:44049386	A>T	chr14:104189861	A>G	chr22:42538897	A>G	chr6:31325494	T>C
chr1:150097602	A>G	chr1:44055781	T>C	chr14:104190527	A>G	chr22:42540551	A>G	chr6:31325504	T>C
chr1:150097692	A>G	chr1:44062437	T>C	chr14:104193099	C>G	chr22:42610099	T>C	chr6:31325529	T>C
chr1:150098110	A>G	chr1:44062805	T>C	chr14:104193439	T>C	chr22:42611834	C>G	chr6:31325565	A>G
chr1:150101169	T>C	chr1:44063025	A>G	chr14:104195610	T>C	chr22:42665117	T>C	chr6:31325620	T>C
chr1:150103589	T>C	chr1:44065100	A>T	chr14:104195664	T>C	chr22:42680201	T>C	chr6:31325623	T>C
chr1:150104039	A>G	chr1:44066406	T>C	chr14:104196405	C>G	chr22:42705870	A>C	chr6:31325692	A>G
chr1:150104269	C>G	chr1:4407019	A>G	chr14:104196607	A>G	chr22:42711636	T>C	chr6:31325702	A>G
chr1:150106226	T>G	chr1:44076469	A>G	chr14:104201869	C>G	chr22:46473845	A>C	chr6:31325745	T>C
chr1:150106554	T>G	chr1:44076630	T>C	chr14:104202304	A>G	chr22:46474527	T>C	chr6:31325756	A>G
chr1:150106834	T>G	chr1:44078384	T>C	chr14:104306696	T>C	chr22:46477390	C>G	chr6:31325758	A>G
chr1:150107136	A>G	chr1:44079411	T>C	chr14:104314182	A>G	chr22:46478195	A>G	chr6:31334422	T>C
chr1:150107509	T>C	chr1:44080958	A>G	chr14:104314183	C>G	chr22:50311973	A>G	chr6:31334477	A>G
chr1:150108487	T>C	chr1:44081912	A>G	chr14:104322026	A>G	chr22:50312679	A>G	chr6:31335318	T>C
chr1:150109095	A>C	chr1:44083015	A>G	chr14:104322394	A>C	chr22:50316444	A>G	chr6:31362891	C>G
chr1:150110056	A>G	chr1:44084063	T>C	chr14:104338258	A>G	chr22:50317187	T>C	chr6:31363026	T>G
chr1:150110435	A>T	chr1:44084083	A>G	chr14:104345822	A>G	chr22:50321623	T>C	chr6:31430694	C>G
chr1:150110480	A>G	chr1:44084620	T>G	chr14:104346185	T>C	chr22:5032324	T>C	chr6:31430721	A>G
chr1:150111602	C>G	chr1:44086831	T>C	chr14:104354505	T>C	chr3:10398964	T>C	chr6:31430752	A>G
chr1:150111980	T>C	chr1:44087141	A>G	chr14:30000661	A>T	chr3:10400145	A>G	chr6:31430799	T>G
chr1:150112476	T>C	chr1:44091404	T>C	chr14:30001032	A>T	chr3:10800703	A>G	chr6:31461771	A>T
chr1:150113490	A>T	chr1:44091476	A>T	chr14:30154174	A>G	chr3:10800794	A>G	chr6:31461979	T>C

chr1:150115486	T>G	chr1:44093693	T>C	chr14:33292743	A>C	chr3:114171390	T>C	chr6:31462035	A>G
chr1:150115974	A>G	chr1:44094528	A>G	chr14:35514957	A>G	chr3:114171726	T>C	chr6:31462135	A>G
chr1:150119151	A>G	chr1:44097530	T>C	chr14:35591316	A>G	chr3:114171818	A>C	chr6:31462248	T>G
chr1:150119177	T>C	chr1:44099554	A>T	chr14:35592538	C>G	chr3:114983656	T>C	chr6:31465047	T>G
chr1:150119729	T>C	chr1:44100084	A>G	chr14:35625217	T>G	chr3:117640167	C>G	chr6:31514247	A>G
chr1:150123762	T>G	chr1:44101767	T>G	chr14:35806152	T>C	chr3:117640660	T>C	chr6:31543031	A>G
chr1:150124615	T>C	chr1:44102587	T>C	chr14:35809414	A>G	chr3:123967920	A>G	chr6:31583155	A>G
chr1:150128425	T>C	chr1:44102782	A>G	chr14:35811699	T>C	chr3:135473872	A>C	chr6:31583827	T>G
chr1:150129335	A>G	chr1:44103275	A>G	chr14:59951446	T>C	chr3:135474349	A>C	chr6:31587083	T>C
chr1:150129998	T>C	chr1:44105789	T>C	chr14:59951458	T>C	chr3:135914476	A>G	chr6:31587561	C>G
chr1:150130081	T>C	chr1:44107072	A>C	chr14:60099375	T>C	chr3:135914715	A>G	chr6:31587870	A>T
chr1:150130983	C>G	chr1:44107428	T>C	chr14:60144733	A>C	chr3:135953476	T>C	chr6:31588384	T>G
chr1:150132324	A>G	chr1:44109749	A>G	chr14:60146348	T>C	chr3:135953729	A>G	chr6:31589264	C>G
chr1:150132430	A>G	chr1:44114973	C>G	chr14:60146555	A>G	chr3:136031358	C>G	chr6:31600820	A>G
chr1:150132627	T>C	chr1:44118195	A>T	chr14:60155307	A>C	chr3:136110506	C>G	chr6:31600851	T>G
chr1:150133323	A>C	chr1:44118354	T>C	chr14:60169895	T>C	chr3:136377377	T>C	chr6:31601012	T>C
chr1:150133507	A>C	chr1:44122165	T>C	chr14:60170383	A>G	chr3:136470004	C>G	chr6:31601022	T>C
chr1:150133563	A>T	chr1:44126143	T>C	chr14:60170932	T>C	chr3:136471815	T>C	chr6:31606376	T>C
chr1:150134221	C>G	chr1:44126440	A>G	chr14:60171047	A>G	chr3:136577085	T>C	chr6:31606392	A>G
chr1:150134735	A>G	chr1:44130564	A>G	chr14:66315883	A>G	chr3:136739406	C>G	chr6:31606398	C>G
chr1:150135291	T>C	chr1:44140312	T>C	chr14:66317858	A>T	chr3:136740769	A>G	chr6:31619576	A>G
chr1:150137773	A>G	chr1:44185442	A>G	chr14:71361367	C>G	chr3:136741079	C>G	chr6:31619652	A>T
chr1:150138151	A>G	chr1:44186396	A>G	chr14:71361414	A>G	chr3:136741389	T>C	chr6:31620020	A>G
chr1:150138433	A>G	chr1:44198297	A>G	chr14:71362254	T>C	chr3:136741804	A>C	chr6:31620520	T>G
chr1:150138699	T>C	chr1:44371259	T>C	chr14:71365940	A>T	chr3:136742300	A>G	chr6:31633496	T>G
chr1:150138709	T>C	chr1:50572096	A>C	chr14:71366196	A>G	chr3:136751460	A>T	chr6:31634193	A>G
chr1:150138782	A>T	chr1:50579237	A>C	chr14:7171904	T>C	chr3:136751526	A>G	chr6:31671557	A>C
chr1:150139579	A>G	chr1:66310689	A>C	chr14:7173580	T>C	chr3:136751674	T>G	chr6:31697387	C>G
chr1:150140833	A>G	chr1:66311335	A>T	chr14:717374702	A>G	chr3:136751732	C>G	chr6:31697558	A>G
chr1:150140885	A>G	chr1:6701978	T>C	chr14:71584473	T>C	chr3:136751968	T>C	chr6:31698088	T>G
chr1:150143153	T>C	chr1:6711941	T>C	chr14:715851113	T>C	chr3:136760598	A>G	chr6:31704294	T>G
chr1:150143193	T>C	chr1:6713282	A>G	chr14:71586128	T>C	chr3:161408621	A>T	chr6:31705864	T>C
chr1:150143302	A>G	chr1:6721234	A>T	chr14:72447776	C>G	chr3:161463625	A>G	chr6:31708463	A>C
chr1:150146907	A>T	chr1:6723184	C>G	chr14:72447995	T>C	chr3:1614661182	C>G	chr6:31712196	A>G
chr1:150147150	A>G	chr1:6725230	C>G	chr14:72448082	A>G	chr3:161780887	A>G	chr6:31740454	A>G
chr1:150147475	T>C	chr1:6726401	A>G	chr14:72448162	A>G	chr3:161780948	C>G	chr6:31745464	A>G
chr1:150147962	T>C	chr1:6731501	T>C	chr14:72454935	A>T	chr3:161781149	T>C	chr6:31761213	A>G
chr1:150149299	T>C	chr1:6732424	A>G	chr14:84668767	A>G	chr3:161781288	T>G	chr6:31762843	C>G
chr1:150152860	A>G	chr1:6735264	A>C	chr14:99667179	A>G	chr3:161781337	T>C	chr6:31762844	T>G
chr1:150154352	T>C	chr1:6737212	T>C	chr14:99707933	T>C	chr3:161781423	A>G	chr6:31783208	A>C
chr1:150157712	A>G	chr1:73288520	A>G	chr14:99711331	A>G	chr3:161781431	A>G	chr6:31783507	C>G
chr1:150158723	A>T	chr1:73288601	A>T	chr14:99711394	T>C	chr3:161782037	T>C	chr6:31786872	A>G
chr1:150159616	T>C	chr1:73288933	A>G	chr14:99711953	A>G	chr3:161782551	C>G	chr6:31786963	A>G
chr1:150162158	T>C	chr1:73311542	A>G	chr14:99712032	A>G	chr3:161782714	T>C	chr6:31787167	T>C
chr1:150164069	T>C	chr1:73313762	A>G	chr14:99712794	T>C	chr3:161782729	A>G	chr6:31794592	T>C
chr1:150167235	T>C	chr1:73315916	A>G	chr14:99712945	A>G	chr3:161791586	T>C	chr6:31797587	C>G
chr1:150170853	T>C	chr1:73285979	T>C	chr14:997207086	A>G	chr3:17221252	A>C	chr6:31799076	T>G
chr1:150173801	T>G	chr1:73336107	A>G	chr14:99733384	C>G	chr3:17221849	A>C	chr6:31800868	C>G
chr1:150174366	A>G	chr1:73336263	A>C	chr15:34659517	C>G	chr3:17232635	A>G	chr6:31801233	T>C
chr1:150175613	A>G	chr1:73336918	C>G	chr15:40566759	T>C	chr3:17235874	T>C	chr6:31802101	T>C
chr1:150177282	A>G	chr1:73343966	A>G	chr15:40567237	A>G	chr3:17236255	A>G	chr6:31820400	T>C
chr1:150179127	A>G	chr1:73344155	C>G	chr15:405774469	A>G	chr3:17237409	A>T	chr6:31820643	A>G
chr1:150181911	T>G	chr1:73354728	A>T	chr15:40583560	T>G	chr3:17274801	T>C	chr6:31832038	A>G
chr1:150183322	T>G	chr1:73361777	C>G	chr15:44250313	T>C	chr3:17275173	C>G	chr6:31838441	T>C
chr1:150184102	A>G	chr1:73362466	T>G	chr15:59042012	T>C	chr3:17301087	C>G	chr6:31838490	T>C
chr1:150184794	T>C	chr1:73370644	A>G	chr15:61829676	A>C	chr3:17308296	A>C	chr6:31838545	A>G
chr1:150185439	A>G	chr1:73412289	T>G	chr15:78730252	A>G	chr3:17308770	A>G	chr6:31838713	A>G
chr1:150189284	A>G	chr1:73435171	A>G	chr15:78731808	A>T	chr3:17418533	T>C	chr6:31839309	T>C
chr1:150189738	A>G	chr1:73435629	A>G	chr15:78774676	A>G	chr3:17454766	A>T	chr6:31839331	T>C
chr1:150193165	A>G	chr1:73531160	C>G	chr15:78832832	T>C	chr3:17686288	A>G	chr6:31839494	T>C
chr1:150194912	A>T	chr1:73531198	T>C	chr15:78833036	A>C	chr3:17686445	T>C	chr6:31839756	C>G
chr1:150204973	A>T	chr1:73531286	T>C	chr15:78833450	A>C	chr3:17719764	A>G	chr6:31839782	A>C
chr1:150210249	T>C	chr1:73562776	A>G	chr15:78833453	T>C	chr3:17720104	A>C	chr6:31864304	A>G
chr1:150211896	T>C	chr1:73576140	T>G	chr15:78857896	A>T	chr3:17784818	T>C	chr6:31864547	T>G
chr1:150214166	A>G	chr1:73598422	A>C	chr15:78857939	T>G	chr3:17882668	A>G	chr6:31870326	A>G
chr1:150216357	A>G	chr1:73609533	A>G	chr15:78857986	C>G	chr3:180589804	A>G	chr6:31927342	A>G
chr1:150217786	C>G	chr1:73610030	T>C	chr15:78858400	C>G	chr3:180629399	T>G	chr6:31940897	A>G
chr1:150221068	T>C	chr1:73610214	A>G	chr15:78930510	A>G	chr3:180630191	A>G	chr6:32006886	A>G
chr1:150221653	T>C	chr1:73645866	T>C	chr15:83318202	T>C	chr3:180630565	T>C	chr6:32007063	T>C
chr1:150221726	T>C	chr1:73646085	T>C	chr15:83368514	T>G	chr3:180632030	T>C	chr6:32014828	A>G
chr1:150223780	A>G	chr1:73646317	T>C	chr15:83368738	T>G	chr3:180632126	T>G	chr6:32038550	T>C
chr1:150225097	A>G	chr1:73664320	A>G	chr15:83371566	A>G	chr3:180647263	T>C	chr6:32038700	T>C
chr1:150226321	T>G	chr1:73664395	C>G	chr15:83371598	T>C	chr3:180647410	T>C	chr6:32050653	T>C
chr1:150228699	T>C	chr1:73664563	T>C	chr15:84749477	T>C	chr3:180658176	T>C	chr6:32096001	A>C
chr1:150242340	A>G	chr1:73664609	T>C	chr15:84814202	T>C	chr3:180658544	A>G	chr6:32114515	A>T
chr1:150242604	A>C	chr1:73776673	A>G	chr15:85113337	T>G	chr3:180700150	A>G	chr6:32117971	T>G
chr1:150246070	T>C	chr1:73776696	C>G	chr15:85113490	C>G	chr3:180707961	T>C	chr6:32119730	C>G
chr1:150247451	A>T	chr1:73776850	A>C	chr15:85113972	A>C	chr3:180773880	T>G	chr6:32122472	A>C
chr1:150250991	A>G	chr1:73797927	A>T	chr15:85114268	T>C	chr3:180774428	A>G	chr6:321313972	C>G
chr1:150251915	A>G	chr1:73814159	A>G	chr15:85114447	T>C	chr3:180774673	A>G	chr6:321314656	T>C
chr1:150252762	A>G	chr1:73843213	T>G	chr15:85114547	T>C	chr3:180774847	A>C	chr6:32145205	T>C
chr1:150262094	T>C	chr1:73843468	T>C	chr15:85122620	A>C	chr3:180774987	A>T	chr6:32145993	C>G
chr1:150262455	A>G	chr1:73848944	T>C	chr15:85143277	A>G	chr3:180804471	A>G	chr6:32145999	C>G

chr1:150263832	A>G	chr1:73848985	A>G	chr15:85148231	T>G	chr3:180830337	T>C	chr6:32146492	A>C
chr1:150264311	T>G	chr1:73848987	A>C	chr15:85177297	A>G	chr3:180830910	A>T	chr6:32146644	T>C
chr1:150264930	T>C	chr1:73849016	A>G	chr15:85177723	T>G	chr3:180849838	A>G	chr6:32146997	A>G
chr1:150264936	T>G	chr1:73849297	A>G	chr15:85193869	T>C	chr3:180851257	A>T	chr6:32151934	A>G
chr1:150266559	A>C	chr1:73919230	A>G	chr15:85193937	A>G	chr3:180894780	T>C	chr6:32152442	A>G
chr1:150272626	A>G	chr1:73923349	T>C	chr15:85195430	T>C	chr3:180928466	T>C	chr6:32153406	A>C
chr1:150272811	T>C	chr1:73928755	A>T	chr15:85200946	C>G	chr3:181035370	A>T	chr6:32153409	A>C
chr1:150274489	A>C	chr1:73965869	C>G	chr15:85201419	T>C	chr3:181061017	T>C	chr6:32154285	A>G
chr1:150274609	A>G	chr1:73966712	A>G	chr15:85250253	T>C	chr3:181101544	T>C	chr6:32154998	T>C
chr1:150277163	A>G	chr1:73977107	T>G	chr15:91404705	A>G	chr3:181176295	T>G	chr6:32155581	A>G
chr1:150278823	T>C	chr1:73983073	T>G	chr15:91404788	T>C	chr3:181178009	A>G	chr6:32156908	A>G
chr1:150279087	T>C	chr1:74077588	T>C	chr15:91407197	T>G	chr3:18488631	T>C	chr6:32158319	A>G
chr1:150281795	C>G	chr1:83556669	A>C	chr15:91407275	A>C	chr3:2488667	A>T	chr6:32159956	T>C
chr1:150284311	T>C	chr1:8357221	T>C	chr15:91416550	A>C	chr3:2494042	A>G	chr6:32165444	A>G
chr1:150285715	T>C	chr1:8357338	T>C	chr15:91418297	A>T	chr3:2519380	C>G	chr6:32171075	A>G
chr1:150287518	T>C	chr1:8357940	T>C	chr15:91418394	T>C	chr3:2519703	T>G	chr6:32171683	T>C
chr1:150289614	A>G	chr1:8358767	T>C	chr15:91427612	A>G	chr3:2553398	A>T	chr6:32191041	A>G
chr1:150294925	A>G	chr1:8359799	A>G	chr15:91428782	C>G	chr3:2554097	T>C	chr6:32191620	T>C
chr1:150295673	A>G	chr1:8360487	A>G	chr15:91428197	T>C	chr3:2554297	T>G	chr6:32202817	T>C
chr1:150295738	T>G	chr1:8360494	A>G	chr15:91428290	T>C	chr3:2554612	T>G	chr6:32203591	A>G
chr1:150296058	A>C	chr1:8361312	A>T	chr15:91429042	A>G	chr3:30067575	A>C	chr6:32203608	T>C
chr1:150298192	A>G	chr1:8361620	T>C	chr15:91429176	T>G	chr3:30071445	T>C	chr6:32223109	A>G
chr1:150305992	T>C	chr1:8362460	A>G	chr15:91429287	A>C	chr3:30078414	C>G	chr6:32223418	T>C
chr1:150306462	T>C	chr1:8362754	T>C	chr16:29923510	A>G	chr3:36856030	T>C	chr6:32223632	A>G
chr1:150308032	A>G	chr1:8364393	A>G	chr16:29931593	T>C	chr3:36856328	A>G	chr6:32223770	T>C
chr1:150308309	T>C	chr1:8365102	A>G	chr16:29931899	A>G	chr3:36885392	T>C	chr6:32223775	A>G
chr1:150308919	A>G	chr1:8366090	A>T	chr16:29932064	T>C	chr3:36893965	T>C	chr6:32223785	C>G
chr1:150308958	T>G	chr1:8368463	T>C	chr16:29932691	T>C	chr3:36894048	T>C	chr6:32381280	C>G
chr1:150311662	A>G	chr1:8370068	A>C	chr16:29936654	A>G	chr3:36966060	A>C	chr6:32381461	T>C
chr1:150311890	T>G	chr1:8374297	C>G	chr16:29937652	A>G	chr3:44062007	A>G	chr6:32381472	T>C
chr1:150313918	A>G	chr1:8379967	A>G	chr16:29938820	A>G	chr3:52261031	A>G	chr6:32406100	T>C
chr1:150316291	A>G	chr1:8380123	A>G	chr16:29954654	T>G	chr3:52267732	C>G	chr6:32406887	T>C
chr1:150319081	A>T	chr1:8381159	A>G	chr16:29972690	T>C	chr3:52268866	T>C	chr6:32407153	A>T
chr1:150319272	T>G	chr1:8381412	C>G	chr16:29984839	C>G	chr3:52273421	A>G	chr6:32407302	A>G
chr1:150319628	A>G	chr1:8383778	C>G	chr16:29988349	A>G	chr3:52279594	A>G	chr6:32407310	T>G
chr1:150319892	T>C	chr1:8389154	A>G	chr16:29988941	A>G	chr3:52287293	T>C	chr6:32407404	C>G
chr1:150319901	T>C	chr1:8389894	T>C	chr16:30042677	T>G	chr3:52287468	T>G	chr6:32407433	C>G
chr1:150322333	C>G	chr1:8390008	A>G	chr16:30060655	A>G	chr3:52288945	T>C	chr6:32407440	T>C
chr1:150324284	T>C	chr1:8390051	A>G	chr16:30134656	T>C	chr3:52322417	A>G	chr6:32407468	T>C
chr1:150325231	A>G	chr1:8390054	A>C	chr16:4454170	A>G	chr3:52333671	C>G	chr6:32408012	A>G
chr1:150327455	A>G	chr1:8390246	T>C	chr16:4455675	A>G	chr3:52338852	T>C	chr6:32440879	T>C
chr1:150331232	A>G	chr1:8391251	T>C	chr16:452619	A>G	chr3:52341215	A>C	chr6:32440880	A>G
chr1:150331515	A>G	chr1:8392592	A>G	chr16:462832	T>C	chr3:52349204	A>G	chr6:32440910	T>G
chr1:150335534	A>G	chr1:8393871	T>C	chr16:462836	A>G	chr3:52349412	T>G	chr6:32440969	A>G
chr1:150335536	A>G	chr1:8412989	A>G	chr16:462897	A>C	chr3:52558008	T>C	chr6:32441178	A>G
chr1:150337289	T>G	chr1:8418644	A>C	chr16:467533	T>C	chr3:52558133	T>C	chr6:32441199	A>G
chr1:150339516	T>C	chr1:8418650	T>C	chr16:475868	A>T	chr3:52559705	A>G	chr6:32441268	A>G
chr1:150344961	A>G	chr1:8421203	T>C	chr16:476089	T>C	chr3:52566682	A>G	chr6:32441408	T>C
chr1:150345558	T>C	chr1:8422676	T>C	chr16:484328	A>G	chr3:52566880	A>C	chr6:32441422	T>C
chr1:150346087	A>C	chr1:8423510	A>G	chr16:484396	A>T	chr3:52566914	A>C	chr6:32441555	A>G
chr1:150349862	A>G	chr1:84266791	A>G	chr16:484613	A>G	chr3:52567014	T>G	chr6:32441641	A>G
chr1:150350927	T>C	chr1:84279139	A>G	chr16:4900544	T>C	chr3:52567617	A>G	chr6:32441679	A>G
chr1:150353050	T>C	chr1:84279279	T>C	chr16:4924060	A>C	chr3:52567779	A>G	chr6:32441934	T>C
chr1:150354578	A>G	chr1:84280055	A>C	chr16:4925785	A>G	chr3:52568805	T>G	chr6:32497975	A>G
chr1:150355625	A>G	chr1:84280341	T>C	chr16:4936621	T>C	chr3:52593119	A>G	chr6:32498174	T>C
chr1:150355690	T>C	chr1:84280513	T>G	chr16:4936934	A>T	chr3:52593138	A>G	chr6:32520273	A>G
chr1:150356842	A>G	chr1:84280704	T>C	chr16:4958832	C>G	chr3:52593230	T>C	chr6:32520313	T>G
chr1:150357010	C>G	chr1:84282586	A>T	chr16:4959226	C>G	chr3:52719088	T>C	chr6:32520465	A>G
chr1:150357777	A>C	chr1:84287969	A>G	chr16:4959618	A>G	chr3:52719398	A>C	chr6:32520595	A>G
chr1:150359424	T>C	chr1:842885573	T>C	chr16:4955759	T>C	chr3:52719816	T>C	chr6:32521083	T>C
chr1:150362014	A>G	chr1:84286576	A>G	chr16:4960929	C>G	chr3:52720080	A>C	chr6:32521155	T>G
chr1:150363356	A>G	chr1:84286897	A>G	chr16:4962351	T>C	chr3:52739520	T>C	chr6:32521195	T>G
chr1:150363472	A>T	chr1:84287573	A>T	chr16:4962512	A>G	chr3:52740182	C>G	chr6:32551291	T>C
chr1:150364081	A>T	chr1:84287969	A>G	chr16:4973040	A>G	chr3:52748857	T>C	chr6:32551329	A>G
chr1:150364447	A>G	chr1:84288165	A>G	chr16:4980034	A>G	chr3:52749334	T>C	chr6:32551334	T>C
chr1:150364873	A>G	chr1:84289032	C>G	chr16:4980214	T>C	chr3:52804487	T>C	chr6:32551468	A>T
chr1:150365275	T>C	chr1:84289093	A>G	chr16:53583771	A>G	chr3:52805093	A>G	chr6:32551469	C>G
chr1:150366212	A>T	chr1:84290288	T>C	chr16:58549932	T>C	chr3:52826846	A>G	chr6:32551568	C>G
chr1:150368080	T>C	chr1:84295070	A>G	chr16:58660691	T>C	chr3:52827566	A>G	chr6:32551571	A>C
chr1:150368195	C>G	chr1:84298042	T>C	chr16:58660692	A>G	chr3:52827915	T>C	chr6:32551653	A>C
chr1:150369716	T>C	chr1:84301640	C>G	chr16:58664602	A>G	chr3:52828628	T>C	chr6:32551703	T>G
chr1:150369895	A>G	chr1:84303894	A>T	chr16:58682616	T>C	chr3:52866157	C>G	chr6:32552075	A>G
chr1:150375173	T>C	chr1:84306747	T>C	chr16:58682833	T>C	chr3:52866535	A>C	chr6:32552092	A>T
chr1:150376141	T>G	chr1:84309863	T>G	chr16:63699425	T>G	chr3:52866845	A>C	chr6:32552219	A>G
chr1:150377101	A>T	chr1:84310896	A>G	chr16:63699511	T>C	chr3:52867595	A>C	chr6:32552239	T>C
chr1:150377172	A>G	chr1:84312316	T>C	chr16:67989523	T>C	chr3:52868145	A>G	chr6:32552291	A>T
chr1:150379542	C>G	chr1:84312696	T>G	chr16:68074857	A>G	chr3:52868445	A>G	chr6:32552316	T>C
chr1:150381835	A>G	chr1:84313288	T>C	chr16:68280893	T>C	chr3:52868610	T>C	chr6:32552404	T>C
chr1:150383021	A>G	chr1:84314270	A>G	chr16:68288331	T>C	chr3:52869263	A>C	chr6:32552448	A>C
chr1:150383281	T>C	chr1:8431607	T>C	chr16:68288393	C>G	chr3:52870132	A>G	chr6:32552686	A>G
chr1:150384458	T>G	chr1:8432136	T>C	chr16:68288983	T>G	chr3:52870618	A>T	chr6:32552779	T>C
chr1:150384514	A>G	chr1:8436802	T>C	chr16:68289059	T>C	chr3:52876389	T>G	chr6:32552795	T>G
chr1:150384895	A>G	chr1:8439625	A>C	chr16:68289313	A>C	chr3:52878600	C>G	chr6:32552826	A>C

chr1:150385282	C>G	chr1:8445360	T>C	chr16:68323654	A>G	chr3:52880128	T>G	chr6:32552870	T>C
chr1:150386431	T>C	chr1:8447404	T>C	chr16:68414975	C>G	chr3:52880543	A>G	chr6:32556475	T>C
chr1:150387715	A>G	chr1:8447722	T>C	chr16:89393562	T>C	chr3:52883070	T>C	chr6:32556493	T>C
chr1:150387733	A>G	chr1:8448404	A>G	chr16:89399455	A>G	chr3:52931971	T>G	chr6:32556542	T>C
chr1:150387786	T>C	chr1:8452725	T>C	chr16:89524923	T>C	chr3:52968726	T>C	chr6:32556689	T>G
chr1:150387870	C>G	chr1:8461061	C>G	chr16:89541795	A>G	chr3:52969128	T>C	chr6:32556858	C>G
chr1:150388647	A>T	chr1:8463176	T>C	chr16:89561475	A>C	chr3:53033057	A>C	chr6:32557243	A>T
chr1:150389029	T>G	chr1:8464509	T>C	chr16:89632725	T>G	chr3:53033081	C>G	chr6:32571647	A>C
chr1:150390972	A>G	chr1:8466730	T>C	chr16:89697625	T>C	chr3:53033295	A>G	chr6:32571690	T>C
chr1:150391034	C>G	chr1:8467353	A>G	chr16:89872445	A>C	chr3:53033796	T>C	chr6:32571845	T>C
chr1:150394234	C>G	chr1:8468278	A>G	chr16:89872827	A>G	chr3:53078590	A>C	chr6:32571962	A>G
chr1:150395254	A>G	chr1:8468372	A>T	chr16:89872970	T>C	chr3:63951765	T>C	chr6:32573039	T>C
chr1:150395830	T>G	chr1:8469284	C>G	chr16:9881953	T>C	chr3:63965093	A>C	chr6:32573040	A>G
chr1:150397056	C>G	chr1:8469337	A>G	chr16:9885807	T>G	chr3:71256309	T>C	chr6:32573058	T>C
chr1:150398992	A>C	chr1:8473331	A>G	chr16:9885832	T>C	chr3:71297450	A>T	chr6:32577455	A>G
chr1:150399671	A>G	chr1:8473336	A>G	chr16:9914014	A>G	chr3:71320898	A>G	chr6:32577470	T>G
chr1:150401971	T>C	chr1:8476428	A>G	chr16:9914095	T>C	chr3:71547590	A>G	chr6:32577474	T>G
chr1:150402419	T>C	chr1:8476441	C>G	chr16:9914133	T>G	chr3:71574051	T>C	chr6:32577633	C>G
chr1:150402613	C>G	chr1:8481016	T>G	chr16:9914401	C>G	chr4:102855110	C>G	chr6:32577784	T>C
chr1:150403238	A>T	chr1:8482078	T>C	chr16:9914419	C>G	chr4:105445205	A>C	chr6:32577785	A>G
chr1:150403379	C>G	chr1:8484228	T>C	chr16:9915013	A>G	chr4:118654892	A>G	chr6:32577889	T>C
chr1:150404288	A>G	chr1:8484823	A>G	chr16:9916204	C>G	chr4:17026605	T>C	chr6:32577907	T>C
chr1:150405474	A>G	chr1:8486131	A>C	chr16:9926348	T>G	chr4:170264344	T>G	chr6:32577975	T>C
chr1:150407299	T>C	chr1:8486341	C>G	chr17:11227300	T>G	chr4:1702636752	T>C	chr6:32577981	A>G
chr1:150407596	A>G	chr1:8487323	A>C	chr17:11227739	T>C	chr4:170236963	T>C	chr6:32578040	C>G
chr1:150407723	T>C	chr1:8488565	A>G	chr17:11227881	T>C	chr4:170266486	T>C	chr6:32578082	T>C
chr1:150409100	A>T	chr1:8489302	T>C	chr17:1282979	A>G	chr4:170319471	A>G	chr6:32578191	T>C
chr1:150409116	A>C	chr1:8490320	T>G	chr17:1283915	A>C	chr4:170328209	A>G	chr6:32578196	A>G
chr1:150409185	A>C	chr1:8490983	T>C	chr17:17649172	T>C	chr4:170328389	T>C	chr6:32578209	T>C
chr1:150411499	T>C	chr1:8495590	A>G	chr17:17661802	A>G	chr4:170331972	T>G	chr6:32578229	T>C
chr1:150412523	T>C	chr1:8495945	T>C	chr17:17696531	C>G	chr4:170540857	T>C	chr6:32578230	A>G
chr1:150412599	A>T	chr1:8497307	A>T	chr17:17708529	A>G	chr4:170579399	T>C	chr6:32594579	A>G
chr1:150412744	A>C	chr1:8498680	A>T	chr17:17708846	C>G	chr4:170580637	T>C	chr6:32594709	A>T
chr1:150415631	T>G	chr1:8501786	A>G	chr17:17709136	A>G	chr4:170585506	A>G	chr6:32594775	A>G
chr1:150415975	A>G	chr1:8503242	A>G	chr17:17715101	A>G	chr4:170586793	A>G	chr6:32594980	A>T
chr1:150416220	T>G	chr1:8503379	A>C	chr17:17715317	C>G	chr4:176731401	A>G	chr6:32595060	A>G
chr1:150416913	A>G	chr1:8504421	A>C	chr17:17724789	A>G	chr4:176859992	A>G	chr6:32595194	T>C
chr1:150418217	T>G	chr1:8505058	A>G	chr17:17726648	A>G	chr4:176865001	T>C	chr6:32595223	A>G
chr1:150418468	T>C	chr1:8535164	T>G	chr17:17726965	T>C	chr4:176865038	A>G	chr6:32595278	T>C
chr1:150418501	A>G	chr1:8536575	A>C	chr17:17727943	A>T	chr4:176865766	A>T	chr6:32603599	T>C
chr1:150419097	A>G	chr1:8548098	A>C	chr17:17728043	A>G	chr4:176866459	A>G	chr6:32603603	T>C
chr1:150419386	A>G	chr1:8549971	A>G	chr17:17728574	T>C	chr4:176867624	A>G	chr6:32603634	C>G
chr1:150419903	T>C	chr1:8554471	T>G	chr17:17730201	T>C	chr4:176868361	A>G	chr6:32603641	C>G
chr1:150420439	C>G	chr1:8559660	A>G	chr17:17733760	T>C	chr4:23334811	A>G	chr6:32603656	C>G
chr1:150421091	A>G	chr1:8560932	A>G	chr17:17733826	T>C	chr4:97762504	T>G	chr6:32603760	T>C
chr1:150421286	T>C	chr1:8567498	T>G	chr17:17733934	T>C	chr4:97763264	T>G	chr6:32603936	A>G
chr1:150421457	T>C	chr1:8601118	A>C	chr17:17734740	A>G	chr5:101677341	T>C	chr6:32603939	T>G
chr1:150422602	A>G	chr1:8605667	A>T	chr17:17735952	T>C	chr5:101707048	A>T	chr6:32603953	A>G
chr1:150423005	T>C	chr1:95850019	A>G	chr17:17740281	T>C	chr5:101756869	T>C	chr6:32603968	A>G
chr1:150423717	T>C	chr1:95890977	T>C	chr17:17741669	T>C	chr5:101757038	T>C	chr6:32604000	T>G
chr1:150423873	A>G	chr1:95902901	T>C	chr17:17741875	A>G	chr5:101757265	T>G	chr6:32604037	C>G
chr1:150423893	A>G	chr1:95905915	T>C	chr17:17742904	A>G	chr5:101757776	A>G	chr6:32604039	T>C
chr1:150423985	T>C	chr1:95906877	A>T	chr17:17744439	A>G	chr5:101777273	A>G	chr6:32604204	A>G
chr1:150424003	T>C	chr1:95906938	A>G	chr17:17744725	T>C	chr5:101777462	A>T	chr6:32604292	A>C
chr1:150424172	A>G	chr1:95909925	A>G	chr17:17746197	T>C	chr5:101777664	A>G	chr6:32604294	A>G
chr1:150426123	T>C	chr1:95911848	T>C	chr17:17747289	A>G	chr5:109005482	T>C	chr6:32604326	A>G
chr1:150426409	C>G	chr1:95912481	A>C	chr17:17747366	A>C	chr5:109006376	T>G	chr6:32604396	T>C
chr1:150427934	T>C	chr1:95912749	A>T	chr17:17747514	T>C	chr5:109007255	T>C	chr6:32604423	A>G
chr1:150427956	T>C	chr1:95913448	A>G	chr17:17748013	A>G	chr5:109014510	A>G	chr6:32604485	A>G
chr1:150428582	T>C	chr1:95914625	C>G	chr17:17753828	A>C	chr5:109026485	C>G	chr6:32604516	T>C
chr1:150428819	A>G	chr1:95914783	T>C	chr17:17754633	T>C	chr5:109026663	C>G	chr6:32604544	T>C
chr1:150428925	A>G	chr1:95915727	T>G	chr17:17755259	A>G	chr5:109037530	T>C	chr6:32604551	T>C
chr1:150429284	T>C	chr1:95915823	A>C	chr17:17764061	C>G	chr5:109039966	A>G	chr6:32604643	T>C
chr1:150429323	T>C	chr1:95920263	T>C	chr17:17804725	T>C	chr5:109040242	A>G	chr6:32604919	T>C
chr1:150429507	C>G	chr1:95921592	A>C	chr17:17811251	T>C	chr5:109075509	T>C	chr6:32604929	C>G
chr1:150429633	A>G	chr1:95921697	A>G	chr17:17822480	T>C	chr5:109076488	T>C	chr6:32604963	A>G
chr1:150429637	T>G	chr1:95923659	T>C	chr17:17823175	C>G	chr5:109077517	T>C	chr6:32604968	A>C
chr1:150429944	A>G	chr1:95926069	A>G	chr17:17824978	T>G	chr5:109078040	A>G	chr6:32604987	T>C
chr1:150430961	A>G	chr1:95926210	A>G	chr17:17826020	A>G	chr5:109079367	T>C	chr6:32604994	A>T
chr1:150431716	A>G	chr1:95927268	T>C	chr17:17829423	T>C	chr5:109084807	T>G	chr6:32605016	A>G
chr1:150433823	T>G	chr1:95929307	T>C	chr17:17829434	A>G	chr5:109141527	T>G	chr6:32605118	A>G
chr1:150437215	A>G	chr1:95929401	A>G	chr17:17830553	T>C	chr5:109189640	T>G	chr6:32605164	T>G
chr1:150437395	A>G	chr1:95929887	T>G	chr17:17831716	T>G	chr5:109201823	T>C	chr6:32605173	T>C
chr1:150437813	T>C	chr1:95930084	T>C	chr17:17832460	A>G	chr5:109202032	C>G	chr6:32605197	A>G
chr1:150438056	A>G	chr1:95933441	T>G	chr17:17833644	A>T	chr5:116537250	A>G	chr6:32605216	A>G
chr1:150438199	A>G	chr1:95935264	A>C	chr17:17834119	A>G	chr5:127212006	T>C	chr6:32605271	T>C
chr1:150438352	A>T	chr1:95936031	A>G	chr17:17839024	A>C	chr5:137673167	A>C	chr6:32605325	T>C
chr1:150438362	A>C	chr1:95937642	A>G	chr17:17840823	T>G	chr5:137775581	C>G	chr6:32605353	A>G
chr1:150439774	T>C	chr1:95941654	C>G	chr17:17841083	T>C	chr5:137780368	A>C	chr6:32605356	A>G
chr1:150440279	T>C	chr1:95944432	T>C	chr17:17841456	A>G	chr5:137784254	T>C	chr6:32605358	C>G
chr1:150440441	A>C	chr1:97120607	C>G	chr17:17847936	A>G	chr5:137784515	A>G	chr6:32605361	T>G
chr1:150440598	T>C	chr1:97120759	T>C	chr17:17848286	C>G	chr5:137840293	T>C	chr6:32605371	C>G
chr1:150442100	T>G	chr1:97803724	T>C	chr17:17851377	T>C	chr5:137840860	A>G	chr6:32605423	T>C

chr1:150448666	C>G	chr1:97842369	T>C	chr17:17864048	T>C	chr5:137863134	A>C	chr6:32605439	A>T
chr1:150448707	T>C	chr1:97842394	T>C	chr17:17866897	A>G	chr5:137889799	A>G	chr6:32605451	A>G
chr1:150448766	C>G	chr1:97845053	T>G	chr17:17875342	A>G	chr5:137903084	A>G	chr6:32605478	T>C
chr1:150450204	T>C	chr1:97846855	C>G	chr17:17875407	A>G	chr5:137939170	C>G	chr6:32605543	A>G
chr1:150450457	C>G	chr1:97846875	T>C	chr17:17875836	A>T	chr5:137940165	T>C	chr6:32605578	A>G
chr1:150451499	C>G	chr1:98105438	A>C	chr17:17876011	A>G	chr5:137942510	A>G	chr6:32605620	A>G
chr1:150453018	C>G	chr1:98298371	A>G	chr17:17876296	C>G	chr5:137945751	C>G	chr6:32605713	A>C
chr1:150453242	A>G	chr1:98299475	A>C	chr17:17878612	A>G	chr5:137946355	A>C	chr6:32605714	T>G
chr1:150454143	A>G	chr1:98304157	T>C	chr17:17883848	A>G	chr5:137946500	A>G	chr6:32626015	A>C
chr1:150455813	T>C	chr1:98305394	A>G	chr17:17894750	A>G	chr5:139064267	C>G	chr6:32626019	T>C
chr1:150456350	A>T	chr1:98309251	A>C	chr17:17897739	T>C	chr5:139064835	T>C	chr6:32626139	C>G
chr1:150456367	A>T	chr1:98310239	A>G	chr17:17898243	T>C	chr5:139065180	A>G	chr6:32632274	A>C
chr1:150456571	T>C	chr1:98315061	A>T	chr17:17902135	A>G	chr5:139065988	T>G	chr6:32632275	A>G
chr1:150456678	T>C	chr1:98315893	A>G	chr17:17907393	A>T	chr5:139067098	A>G	chr6:32632315	C>G
chr1:150459324	A>G	chr1:98316285	T>G	chr17:17909775	A>G	chr5:139068126	C>G	chr6:32632326	A>G
chr1:150460842	A>C	chr1:98316355	T>C	chr17:17930253	A>G	chr5:139070573	C>G	chr6:32632331	T>C
chr1:150460850	T>C	chr1:98318162	A>G	chr17:17939573	A>G	chr5:139070651	A>G	chr6:32632336	A>G
chr1:150461801	A>G	chr1:98318735	T>C	chr17:17942613	T>C	chr5:139070780	A>T	chr6:32632457	T>C
chr1:150461873	T>G	chr1:98320492	A>C	chr17:17944349	A>G	chr5:139070952	A>G	chr6:32632660	T>G
chr1:150463526	T>C	chr1:98322379	T>C	chr17:17990634	T>C	chr5:139071193	T>C	chr6:32632703	A>G
chr1:150463772	C>G	chr1:98325796	A>G	chr17:17990671	T>C	chr5:139071301	A>C	chr6:32632745	A>G
chr1:150464331	C>G	chr1:98325947	A>G	chr17:18009028	T>C	chr5:139071353	T>C	chr6:32632753	T>C
chr1:150465922	A>G	chr1:98326447	A>G	chr17:18009102	A>G	chr5:139073143	T>C	chr6:32632850	A>G
chr1:150466558	T>C	chr1:98327831	T>C	chr17:18011140	T>C	chr5:139534319	A>T	chr6:32632864	A>G
chr1:150466617	T>C	chr1:98328093	T>C	chr17:18011750	T>C	chr5:140027216	T>C	chr6:32632872	A>C
chr1:150466925	T>C	chr1:98334990	T>G	chr17:18012730	T>C	chr5:140027304	T>C	chr6:32632885	T>G
chr1:150467096	A>T	chr1:98335004	C>G	chr17:18021607	A>G	chr5:140105978	A>C	chr6:32632903	A>G
chr1:150467613	A>G	chr1:98336329	A>G	chr17:18021882	T>C	chr5:140174622	A>G	chr6:32632968	C>G
chr1:150467814	T>C	chr1:98336823	T>G	chr17:18022039	A>C	chr5:140174865	C>G	chr6:32632974	C>G
chr1:150467817	A>G	chr1:98337283	A>G	chr17:18917237	A>G	chr5:140186657	A>G	chr6:32633087	A>G
chr1:150468574	T>C	chr1:98337578	A>T	chr17:18917324	A>G	chr5:140187102	A>G	chr6:32633174	T>C
chr1:150468842	A>G	chr1:98341152	T>G	chr17:18917513	A>G	chr5:140187322	T>C	chr6:32633196	A>G
chr1:150470296	A>G	chr1:98342067	A>C	chr17:19014831	C>G	chr5:151936155	A>T	chr6:32633230	A>T
chr1:150476188	T>C	chr1:98342417	A>G	chr17:19141582	C>G	chr5:151936546	A>G	chr6:32633231	C>G
chr1:150481738	T>C	chr1:98342685	A>C	chr17:19143614	A>C	chr5:151936869	A>G	chr6:32633247	T>C
chr1:150482255	A>G	chr1:98344088	T>C	chr17:19146445	A>G	chr5:151940238	A>T	chr6:32633282	T>C
chr1:150486988	A>G	chr1:98344461	T>C	chr17:19153175	C>G	chr5:151945469	A>T	chr6:32633290	C>G
chr1:150489768	C>G	chr1:98345019	A>G	chr17:19153417	T>C	chr5:151947259	A>G	chr6:32633375	A>G
chr1:150492114	T>C	chr1:98345449	A>T	chr17:19163052	A>C	chr5:151964215	A>G	chr6:32633380	A>G
chr1:150495327	A>C	chr1:98345517	C>G	chr17:19172196	T>G	chr5:151975441	T>C	chr6:32633392	C>G
chr1:150496388	C>G	chr1:98345614	A>T	chr17:19172505	A>G	chr5:151975443	A>C	chr6:32633425	T>C
chr1:150498138	T>C	chr1:98346177	T>C	chr17:19174874	T>C	chr5:152009417	T>C	chr6:32633426	C>G
chr1:150498556	T>G	chr1:98348885	A>G	chr17:19175317	A>G	chr5:152010999	A>G	chr6:32633466	A>C
chr1:150498727	A>G	chr1:98352053	A>T	chr17:19203032	A>G	chr5:152011814	A>G	chr6:32633632	A>G
chr1:150498758	A>G	chr1:98353416	A>G	chr17:19204432	A>C	chr5:152012437	T>G	chr6:32633697	A>C
chr1:150499295	A>G	chr1:98357727	T>C	chr17:19204531	A>T	chr5:152012795	A>G	chr6:32633713	A>G
chr1:150500085	A>G	chr1:98359342	T>C	chr17:19204863	T>C	chr5:152013789	C>G	chr6:32633749	A>G
chr1:150500107	A>G	chr1:98360091	A>T	chr17:19211073	C>G	chr5:152013887	A>G	chr6:32633753	T>C
chr1:150504055	A>G	chr1:98362040	T>C	chr17:19213335	C>G	chr5:152014061	T>C	chr6:32633818	C>G
chr1:150506621	A>G	chr1:98365310	A>C	chr17:19220666	A>G	chr5:152014666	T>C	chr6:32633819	A>G
chr1:150507233	A>G	chr1:98369744	A>G	chr17:19223224	T>G	chr5:152014941	A>G	chr6:32633844	T>C
chr1:150509169	T>C	chr1:98370375	C>G	chr17:19224397	C>G	chr5:152015208	A>G	chr6:32633879	A>G
chr1:150509544	A>G	chr1:98371449	A>T	chr17:19239432	T>C	chr5:152055862	T>G	chr6:32633928	A>G
chr1:150509595	A>G	chr1:98375391	T>C	chr17:19241505	A>G	chr5:152056234	T>C	chr6:32633941	A>G
chr1:150510403	A>G	chr1:98376245	T>C	chr17:19242467	A>G	chr5:152149398	A>G	chr6:32634024	A>T
chr1:150512253	C>G	chr1:98376502	T>G	chr17:19247075	A>G	chr5:152150366	A>G	chr6:32634126	A>G
chr1:150514149	T>C	chr1:98378786	A>C	chr17:19250679	A>G	chr5:152191955	A>T	chr6:32634243	A>G
chr1:150540181	T>C	chr1:98379562	A>G	chr17:19265797	T>C	chr5:152191995	T>C	chr6:32634318	A>C
chr1:150616888	A>G	chr1:98381437	T>G	chr17:19281828	T>G	chr5:152193105	A>C	chr6:32634341	A>G
chr1:150636991	A>G	chr1:98381860	T>C	chr17:19291785	A>G	chr5:152196711	C>G	chr6:32634373	T>C
chr1:150672976	A>T	chr1:98382113	A>G	chr17:19912710	T>C	chr5:152286739	A>G	chr6:32634467	T>C
chr1:150685716	T>C	chr1:98382114	T>G	chr17:19914399	A>T	chr5:152517290	T>C	chr6:32634584	C>G
chr1:150742542	T>C	chr1:98383395	T>C	chr17:19914546	T>C	chr5:152518128	T>C	chr6:32634655	A>G
chr1:150776123	A>G	chr1:98385673	A>T	chr17:19988543	A>C	chr5:152528318	A>G	chr6:32634661	A>G
chr1:150807431	A>T	chr1:98385830	T>C	chr17:20205458	T>C	chr5:152584121	T>C	chr6:32813768	A>G
chr1:150822168	A>G	chr1:98386955	A>C	chr17:2025544	C>G	chr5:152584675	A>G	chr6:32862958	C>G
chr1:150857175	T>G	chr1:98390072	A>T	chr17:2027182	C>G	chr5:152785080	A>G	chr6:32904039	A>T
chr1:150959758	A>G	chr1:98391864	A>C	chr17:2040963	A>G	chr5:152871184	A>C	chr6:32904051	T>C
chr1:150961542	A>G	chr1:98394864	A>G	chr17:2074977	T>C	chr5:152897407	C>G	chr6:329044661	A>G
chr1:150961827	T>G	chr1:98395714	A>T	chr17:2076352	A>G	chr5:153497780	A>G	chr6:32905746	A>T
chr1:150962949	T>C	chr1:98395881	T>G	chr17:2083881	A>T	chr5:153498135	T>C	chr6:32935179	A>C
chr1:150963049	A>G	chr1:98396847	T>C	chr17:2084008	A>G	chr5:153529697	A>G	chr6:32936120	T>C
chr1:150964810	A>G	chr1:98397765	T>C	chr17:2084305	A>G	chr5:153530216	T>C	chr6:32936994	A>G
chr1:150968858	T>C	chr1:98398619	T>G	chr17:2084565	T>G	chr5:153621015	A>G	chr6:32952708	A>G
chr1:150969592	T>C	chr1:98399903	T>G	chr17:2085185	A>G	chr5:153622821	A>G	chr6:32996999	A>G
chr1:150969992	T>C	chr1:98400439	A>T	chr17:2085475	T>G	chr5:153624046	A>T	chr6:33041073	A>C
chr1:150971195	T>G	chr1:98404096	T>C	chr17:2085598	A>G	chr5:153644212	A>G	chr6:33047173	T>C
chr1:150972539	T>G	chr1:98404234	T>C	chr17:2085653	T>C	chr5:153644596	T>C	chr6:33047898	A>C
chr1:150974081	T>C	chr1:98405032	C>G	chr17:2117945	A>G	chr5:153644928	A>G	chr6:33048354	A>T
chr1:150974971	A>T	chr1:98405856	A>G	chr17:2119101	T>C	chr5:153666790	T>C	chr6:33048539	A>T
chr1:150986662	A>G	chr1:98409303	A>C	chr17:2126003	A>G	chr5:153681208	A>C	chr6:33048975	A>G
chr1:150988635	A>C	chr1:98409901	A>G	chr17:2126504	C>G	chr5:144880288	T>G	chr6:33049074	T>G
chr1:150989274	A>G	chr1:98410616	T>C	chr17:2195130	A>G	chr5:144908935	T>C	chr6:33049565	C>G

chr1:150990913	C>G	chr1:98411869	A>G	chr17:2195146	A>T	chr5:44914579	A>G	chr6:33049727	A>T
chr1:150992010	A>G	chr1:98416157	A>G	chr17:2196088	T>C	chr5:44926518	A>T	chr6:33049902	A>T
chr1:150992019	A>G	chr1:98416940	T>C	chr17:2196150	T>C	chr5:44927107	A>G	chr6:33049965	A>G
chr1:151002758	A>G	chr1:98417446	C>G	chr17:55728224	A>G	chr5:45435019	T>C	chr6:33160906	T>G
chr1:151004084	A>G	chr1:98417967	A>C	chr17:55743277	A>G	chr5:45440343	T>G	chr6:33282002	T>C
chr1:151005763	A>G	chr1:98418687	A>G	chr18:28513101	T>C	chr5:45441027	C>G	chr6:33282181	A>G
chr1:154251514	T>C	chr1:98419702	T>C	chr18:28513207	T>G	chr5:45534423	C>G	chr6:33283766	T>C
chr1:154252208	A>G	chr1:98420453	C>G	chr18:46711776	A>G	chr5:49573365	T>G	chr6:33290402	A>G
chr1:154252860	T>C	chr1:98421934	A>G	chr18:46712790	T>C	chr5:49621413	T>G	chr6:33290547	A>T
chr1:154258549	T>C	chr1:98421965	A>C	chr18:46714696	T>C	chr5:49698986	A>C	chr6:33290825	A>G
chr1:154789036	T>C	chr1:98422612	A>G	chr18:52722378	T>C	chr5:49760547	A>C	chr6:33324527	T>G
chr1:154791128	A>G	chr1:98423149	A>G	chr18:52746868	A>C	chr5:60139548	T>C	chr6:33332696	A>G
chr1:154990297	C>G	chr1:98424252	A>G	chr18:52747164	T>G	chr5:60139881	A>G	chr6:33341424	A>C
chr1:155876613	A>G	chr1:98425888	A>G	chr18:52747689	T>C	chr5:60508492	T>C	chr6:33359008	A>G
chr1:163644938	A>G	chr1:98426476	A>G	chr18:52747871	T>C	chr5:60528806	A>G	chr6:33359142	A>G
chr1:163650546	C>G	chr1:98428260	T>C	chr18:52754086	A>G	chr5:60581371	T>C	chr6:33359716	C>G
chr1:163653310	T>C	chr1:98430950	A>C	chr18:52931891	T>C	chr5:60585889	T>C	chr6:33360574	T>G
chr1:163662427	A>C	chr1:98431165	A>G	chr18:52942523	A>C	chr5:60588572	C>G	chr6:33394253	C>G
chr1:163662479	T>G	chr1:98431639	A>G	chr18:52967632	C>G	chr5:60589021	T>G	chr6:33395199	A>G
chr1:163666085	T>C	chr1:98432507	A>T	chr18:52968230	A>G	chr5:60597127	C>G	chr6:33399778	T>C
chr1:163666974	T>C	chr1:98432826	A>G	chr18:52985401	A>C	chr5:60615886	A>G	chr6:33401124	C>G
chr1:163667287	A>G	chr1:98432916	A>G	chr18:53014092	A>T	chr5:60616026	T>G	chr6:33401291	A>C
chr1:163671570	A>C	chr1:98433535	A>T	chr18:53014632	C>G	chr5:60624958	A>C	chr6:33421577	T>C
chr1:163678435	T>G	chr1:98435544	A>G	chr18:53014633	T>C	chr5:60625875	A>G	chr6:33556605	C>G
chr1:163679501	T>C	chr1:98436232	T>G	chr18:53023170	A>G	chr5:60699573	A>C	chr6:33558751	A>G
chr1:163680402	A>G	chr1:98437556	T>G	chr18:53052923	T>C	chr5:60704813	C>G	chr6:33661091	A>G
chr1:163682555	T>C	chr1:98438135	A>C	chr18:53053398	A>G	chr5:60711632	A>C	chr6:33662265	T>C
chr1:163688765	A>G	chr1:98440529	T>C	chr18:53063676	T>G	chr5:60712212	A>G	chr6:33700376	T>G
chr1:163699103	A>G	chr1:98441830	A>G	chr18:53063719	A>G	chr5:60712747	A>C	chr6:33702869	T>C
chr1:163701097	A>G	chr1:98441888	T>C	chr18:53065892	A>G	chr5:60727990	T>C	chr6:33703230	A>G
chr1:163701978	C>G	chr1:98442308	T>G	chr18:53068680	T>C	chr5:60735530	T>C	chr6:33710038	T>C
chr1:163703394	T>C	chr1:98443440	A>T	chr18:53070168	T>C	chr5:60735662	C>G	chr6:33710229	T>C
chr1:163705585	T>C	chr1:98443509	T>C	chr18:53071766	A>G	chr5:60736544	T>C	chr6:33710933	A>G
chr1:163708990	T>C	chr1:98443858	A>G	chr18:53074618	A>G	chr5:60736949	A>G	chr6:33711420	T>C
chr1:163709434	A>G	chr1:98445830	C>G	chr18:53075292	A>C	chr5:60751467	T>C	chr6:33713779	T>C
chr1:163709730	A>C	chr1:98446194	C>G	chr18:53082154	T>G	chr5:60751524	A>T	chr6:33714781	A>G
chr1:163710740	A>G	chr1:98447298	A>C	chr18:53082236	T>C	chr5:60760700	T>C	chr6:33719877	T>C
chr1:163714719	A>G	chr1:98448341	A>G	chr18:53085127	A>C	chr5:60777345	A>C	chr6:33729718	C>G
chr1:163714833	C>G	chr1:98487333	A>G	chr18:53085273	T>C	chr5:60831581	T>G	chr6:33738442	T>C
chr1:163716904	A>G	chr1:98487510	A>C	chr18:53085392	T>C	chr5:63713440	A>G	chr6:33739960	A>G
chr1:163718770	A>G	chr1:98487566	T>C	chr18:53087984	A>T	chr5:67076800	A>T	chr6:33748831	T>G
chr1:163721678	T>G	chr1:98491248	A>C	chr18:53106326	T>C	chr5:67089501	A>T	chr6:33749686	A>G
chr1:163722747	A>G	chr1:98497176	A>C	chr18:53125474	C>G	chr5:67123091	T>C	chr6:33750086	A>G
chr1:163722840	T>C	chr1:98499843	T>C	chr18:53125938	T>C	chr5:67246543	T>C	chr6:33755541	A>G
chr1:163725271	A>C	chr1:98501984	T>C	chr18:53126330	A>T	chr5:67686126	A>C	chr6:33755711	C>G
chr1:163726840	A>T	chr1:98502340	A>G	chr18:53144837	T>C	chr5:67987689	A>G	chr6:43337802	A>G
chr1:163727923	A>G	chr1:98502934	T>G	chr18:53147537	C>G	chr5:88224419	C>G	chr6:43395101	A>G
chr1:163731296	T>C	chr1:98507696	A>G	chr18:53150381	A>G	chr5:88716714	T>C	chr6:56558385	A>G
chr1:163733217	A>C	chr1:98508258	T>C	chr18:53158629	C>G	chr5:88743219	A>G	chr6:56559544	T>G
chr1:163733717	A>T	chr1:98512127	T>G	chr18:53158650	A>G	chr5:88743962	T>C	chr6:56571460	T>G
chr1:163734561	A>G	chr1:98513845	A>G	chr18:53168073	A>G	chr5:88744550	A>G	chr6:56571700	C>G
chr1:163734859	A>G	chr1:98513959	A>G	chr18:53173887	A>G	chr5:88745184	T>C	chr6:56572477	A>G
chr1:163736829	A>G	chr1:98520219	A>T	chr18:53214865	T>C	chr5:90228398	T>C	chr6:56573656	T>C
chr1:163736925	A>G	chr1:98524960	A>G	chr18:53214923	T>C	chr5:90228400	A>G	chr6:56574180	T>G
chr1:163747095	A>G	chr1:98525221	A>G	chr18:53215142	A>T	chr5:90251093	T>C	chr6:56574219	T>C
chr1:163748438	A>G	chr1:98526167	T>C	chr18:53215739	A>T	chr5:90278217	A>T	chr6:73121906	A>G
chr1:163750904	A>G	chr1:98527259	T>C	chr18:53215962	T>C	chr5:90940109	T>C	chr6:83890642	T>C
chr1:163752376	A>G	chr1:98528211	A>C	chr18:53222855	A>G	chr5:105388718	T>C	chr6:83901888	C>G
chr1:163752688	T>G	chr1:98528452	T>C	chr18:53252889	T>C	chr6:105389953	A>G	chr6:84173462	T>C
chr1:163753767	A>T	chr1:98529378	A>G	chr18:53253410	C>G	chr6:105400605	T>C	chr6:84225229	T>C
chr1:163758093	A>G	chr1:98530398	A>G	chr18:53253431	C>G	chr6:105400837	A>G	chr6:84264202	A>T
chr1:163762232	T>C	chr1:98530559	T>C	chr18:53346739	A>C	chr6:105453417	A>C	chr6:84264333	T>C
chr1:163763163	A>G	chr1:98531022	A>G	chr18:53412903	T>C	chr6:105457186	T>G	chr6:84306856	A>C
chr1:163763578	T>C	chr1:98532403	T>C	chr18:53445930	T>C	chr6:105462876	T>C	chr6:84307328	T>C
chr1:163764928	A>G	chr1:98533392	A>G	chr18:53446374	A>G	chr6:108944165	T>C	chr6:84307726	T>C
chr1:163766004	A>C	chr1:98535375	A>C	chr18:53449667	A>G	chr6:108977663	T>C	chr6:96432627	A>G
chr1:177271313	A>G	chr1:98536724	T>C	chr18:53567855	A>G	chr6:111451467	A>G	chr6:96462496	T>G
chr1:177271797	T>C	chr1:98537363	T>G	chr18:77630600	C>G	chr6:111501915	A>G	chr6:96464060	T>C
chr1:177275650	A>G	chr1:98541836	C>G	chr18:77631679	A>C	chr6:111834954	T>C	chr6:96464628	T>G
chr1:177276006	T>C	chr1:98545702	A>G	chr18:77632194	A>G	chr6:111845570	A>T	chr7:100278657	A>G
chr1:177280121	T>C	chr1:98552832	A>T	chr19:11849736	T>C	chr6:111863195	C>G	chr7:104083702	A>G
chr1:177760918	A>G	chr1:98687691	T>C	chr19:19383755	A>G	chr6:111872482	T>C	chr7:104083719	A>C
chr1:177766021	C>G	chr10:101907673	T>C	chr19:19391402	T>C	chr6:111895728	A>G	chr7:104583772	A>G
chr1:190887345	A>G	chr10:104238558	C>G	chr19:19431423	T>G	chr6:111895860	T>C	chr7:104583843	A>G
chr1:190891896	T>G	chr10:104359350	T>C	chr19:19431963	T>G	chr6:111897865	A>G	chr7:104584410	T>C
chr1:190893756	T>C	chr10:104384029	A>G	chr19:19453521	T>G	chr6:111914700	A>G	chr7:104585516	C>G
chr1:190895558	A>T	chr10:104384337	A>G	chr19:19453560	A>C	chr6:111915564	C>G	chr7:104585760	A>C
chr1:190900600	A>G	chr10:104386934	T>C	chr19:19471805	A>G	chr6:111918869	A>G	chr7:104611876	T>C
chr1:190903602	T>C	chr10:104387735	T>C	chr19:19475088	A>G	chr6:114685980	A>G	chr7:104616744	A>G
chr1:190905760	A>T	chr10:104388710	T>C	chr19:19475717	C>G	chr6:114711095	T>C	chr7:104618318	T>C
chr1:190912648	T>C	chr10:104390303	A>G	chr19:19476365	A>G	chr6:1164975014	A>G	chr7:104624626	T>C
chr1:190913774	C>G	chr10:104393489	A>G	chr19:19482882	A>G	chr6:25174070	T>C	chr7:104643470	T>C
chr1:190923380	T>C	chr10:104393506	A>G	chr19:19484295	A>G	chr6:25450026	A>C	chr7:104652671	A>G

chr1:190924775	A>G	chr10:104401218	A>C	chr19:19495954	A>G	chr6:25486014	T>C	chr7:104653265	A>G
chr1:190927218	A>G	chr10:104401486	A>T	chr19:19497669	C>G	chr6:25686033	T>G	chr7:104768619	A>T
chr1:190932733	T>C	chr10:104401632	T>C	chr19:19518316	A>T	chr6:25890656	A>G	chr7:104771857	T>C
chr1:190932926	T>C	chr10:104401721	A>G	chr19:19572108	T>G	chr6:25943234	A>C	chr7:104842643	A>G
chr1:190940627	A>G	chr10:104412308	T>C	chr19:19572785	A>G	chr6:25962951	A>C	chr7:104842960	T>C
chr1:190942870	A>T	chr10:104419465	T>G	chr19:19597055	A>G	chr6:25999149	T>C	chr7:104844328	T>C
chr1:190946081	A>G	chr10:104429639	A>G	chr19:19597240	T>C	chr6:26099175	A>G	chr7:104886669	T>C
chr1:190947945	A>T	chr10:104430251	A>G	chr19:19640524	T>C	chr6:26017542	T>C	chr7:104978010	T>C
chr1:190949551	A>T	chr10:104431271	T>C	chr19:19643028	A>C	chr6:26021872	T>C	chr7:104986644	A>T
chr1:190962403	T>C	chr10:104436641	T>C	chr19:19768474	T>G	chr6:26022244	A>G	chr7:104988848	A>G
chr1:190962591	A>G	chr10:104464675	T>C	chr19:2686313	A>G	chr6:26022648	A>C	chr7:105028370	T>C
chr1:190965433	T>C	chr10:104464763	A>G	chr19:2695661	T>C	chr6:26026163	C>G	chr7:110737149	T>C
chr1:190965934	T>C	chr10:104504564	T>C	chr19:2696955	A>G	chr6:26030044	A>G	chr7:110848329	A>T
chr1:190967918	T>C	chr10:104536360	A>G	chr19:2719883	T>C	chr6:26030492	A>C	chr7:110865216	A>G
chr1:190970614	T>G	chr10:104556054	C>G	chr19:2721592	A>G	chr6:26031811	T>G	chr7:110879070	T>G
chr1:190973549	A>G	chr10:104572081	T>C	chr19:2721991	A>C	chr6:26044373	T>C	chr7:110879251	T>C
chr1:190975424	T>C	chr10:104572276	A>G	chr19:2722050	A>C	chr6:26044864	T>C	chr7:110879514	A>G
chr1:190975703	T>G	chr10:104573936	T>C	chr19:2722397	T>C	chr6:26045905	C>G	chr7:110879776	A>G
chr1:190976674	T>G	chr10:104574063	A>G	chr19:38322166	T>C	chr6:26055368	A>G	chr7:110880559	T>C
chr1:190979629	T>C	chr10:104574329	A>T	chr19:3884411	T>G	chr6:26104630	T>C	chr7:110896492	T>C
chr1:190982864	T>G	chr10:104574422	A>G	chr19:38895314	A>G	chr6:26119245	T>C	chr7:110904226	A>G
chr1:190983667	T>G	chr10:104574562	T>G	chr19:3897669	C>G	chr6:26122648	C>G	chr7:110907308	T>C
chr1:190994234	A>G	chr10:104574642	A>T	chr19:38997713	T>C	chr6:26122957	A>G	chr7:110915339	T>C
chr1:190995271	T>C	chr10:104578531	C>G	chr19:38997843	C>G	chr6:26124303	T>C	chr7:110932739	A>G
chr1:190995967	T>C	chr10:104591393	T>G	chr19:38997934	T>G	chr6:26157762	T>C	chr7:111126914	A>T
chr1:190998629	A>T	chr10:104594507	A>G	chr19:3901922	T>C	chr6:26158079	A>C	chr7:111172634	T>C
chr1:190998630	C>G	chr10:104596396	T>C	chr19:4969053	A>G	chr6:26172219	T>C	chr7:121942674	A>C
chr1:190999674	T>C	chr10:104596924	A>C	chr19:50084454	T>C	chr6:26189356	C>G	chr7:121943642	A>G
chr1:191003483	T>C	chr10:104596981	A>G	chr19:50084596	C>G	chr6:26196593	A>G	chr7:121950965	A>G
chr1:191004544	A>T	chr10:104597152	A>G	chr19:50087673	T>C	chr6:26199903	T>C	chr7:137079949	A>C
chr1:191005717	T>C	chr10:104613355	A>G	chr19:50087876	T>C	chr6:26226032	C>G	chr7:137080046	T>C
chr1:191008270	T>C	chr10:104613361	A>G	chr19:50093572	A>G	chr6:26233387	A>G	chr7:137085250	A>G
chr1:191008381	T>G	chr10:104614350	T>C	chr19:50095961	T>C	chr6:26272548	T>G	chr7:1858618	T>C
chr1:191008459	T>G	chr10:1046228873	A>G	chr19:50096144	T>C	chr6:26272829	T>C	chr7:1858898	T>C
chr1:191008849	T>C	chr10:104629011	T>C	chr19:50096421	C>G	chr6:26286468	T>C	chr7:1881813	T>C
chr1:191010445	A>C	chr10:104629465	T>G	chr19:50096859	C>G	chr6:26286744	T>C	chr7:1882227	A>C
chr1:191011045	T>C	chr10:104669621	T>C	chr19:50097547	A>G	chr6:26305053	T>C	chr7:1882470	T>C
chr1:191011109	T>C	chr10:104677126	T>C	chr19:50097784	T>C	chr6:26305623	T>C	chr7:1892565	T>C
chr1:191011296	T>G	chr10:104748459	A>C	chr19:50168871	A>G	chr6:26306286	T>C	chr7:1893311	C>G
chr1:191011348	A>G	chr10:104748718	A>G	chr19:50168927	A>C	chr6:26312170	C>G	chr7:1894022	A>T
chr1:191011811	A>T	chr10:104749725	A>G	chr19:50169020	T>C	chr6:26313348	T>C	chr7:1894116	T>C
chr1:191012376	A>G	chr10:104814162	T>C	chr19:50180266	C>G	chr6:26322768	C>G	chr7:1900208	T>C
chr1:191012788	A>C	chr10:104829469	T>C	chr2:110267912	T>C	chr6:26322861	T>C	chr7:1903100	T>C
chr1:191014487	T>C	chr10:104844872	T>C	chr2:110271723	A>G	chr6:26323132	C>G	chr7:1904709	T>C
chr1:191014574	A>G	chr10:104871204	A>G	chr2:110320075	A>G	chr6:26327905	C>G	chr7:1912057	A>G
chr1:191015427	T>G	chr10:104871279	T>G	chr2:134844407	T>C	chr6:26327956	T>C	chr7:1912222	A>G
chr1:191016342	T>C	chr10:104871361	A>G	chr2:145161902	A>G	chr6:26327965	T>G	chr7:1921429	C>G
chr1:191016750	A>G	chr10:104872547	A>T	chr2:145196196	T>G	chr6:26328366	A>G	chr7:1923695	T>C
chr1:191018978	T>C	chr10:104883337	A>G	chr2:145202375	A>C	chr6:26364628	A>C	chr7:1935245	T>C
chr1:191020093	C>G	chr10:104886188	C>G	chr2:145214607	T>G	chr6:26364930	T>C	chr7:1935273	T>C
chr1:191021828	A>T	chr10:104886374	A>G	chr2:147612734	T>C	chr6:26365147	T>C	chr7:1938655	T>C
chr1:191035029	A>G	chr10:104906211	T>C	chr2:162802993	T>G	chr6:26365346	C>G	chr7:1952582	T>G
chr1:191036146	T>C	chr10:104913653	A>G	chr2:162805019	A>T	chr6:26365586	T>C	chr7:1953056	T>C
chr1:191044104	A>G	chr10:104913870	T>C	chr2:162810159	T>G	chr6:26365679	T>G	chr7:1953521	T>C
chr1:191045040	A>G	chr10:104913940	T>C	chr2:162811186	A>G	chr6:26365759	T>C	chr7:1953523	T>C
chr1:191049583	A>G	chr10:104927634	A>G	chr2:162818621	T>C	chr6:26365766	T>C	chr7:1953571	A>G
chr1:191049835	A>G	chr10:104928914	A>T	chr2:162825174	T>C	chr6:26365913	T>G	chr7:1953606	T>C
chr1:191050262	T>C	chr10:104929191	T>C	chr2:162856148	T>C	chr6:26379537	T>C	chr7:1953650	C>G
chr1:191051478	T>C	chr10:104943986	A>T	chr2:162891848	T>C	chr6:26380032	A>G	chr7:1968953	T>C
chr1:191056694	A>G	chr10:104943993	A>G	chr2:172543913	T>G	chr6:26402717	T>C	chr7:1979741	A>G
chr1:191059435	A>G	chr10:104944744	C>G	chr2:172544483	A>G	chr6:26421345	T>C	chr7:1979750	A>G
chr1:191061596	A>G	chr10:104945213	A>G	chr2:172579954	A>G	chr6:26440429	A>T	chr7:1982181	A>G
chr1:191068048	T>C	chr10:104945463	A>G	chr2:172620152	T>C	chr6:26474044	T>C	chr7:1986738	A>G
chr1:191070621	C>G	chr10:104945751	T>C	chr2:172780132	T>C	chr6:26474571	T>C	chr7:2019875	A>G
chr1:191073058	T>C	chr10:104945823	A>G	chr2:175191038	A>C	chr6:26476782	A>T	chr7:2041432	A>G
chr1:191076906	A>C	chr10:104952499	T>C	chr2:180325356	A>G	chr6:26477143	A>T	chr7:2041638	A>G
chr1:191082452	A>G	chr10:104953044	A>G	chr2:185462041	T>C	chr6:26477779	A>G	chr7:2042448	T>C
chr1:191084822	A>G	chr10:104953094	T>C	chr2:185489221	C>G	chr6:26478252	T>C	chr7:2042547	C>G
chr1:191087345	A>T	chr10:104953547	T>C	chr2:185547462	T>C	chr6:26501768	T>C	chr7:2044096	C>G
chr1:191087633	A>G	chr10:104966171	T>C	chr2:185604982	T>C	chr6:26537801	A>G	chr7:2046328	A>G
chr1:191088099	A>C	chr10:104967679	C>G	chr2:185624638	T>C	chr6:26569135	T>C	chr7:2046830	T>C
chr1:191088135	A>G	chr10:104968050	T>G	chr2:185767854	T>C	chr6:26896482	T>C	chr7:2049854	A>G
chr1:191088294	T>C	chr10:106569207	A>G	chr2:185768559	A>C	chr6:26923976	T>C	chr7:2050401	T>C
chr1:191088498	A>G	chr10:106569253	A>G	chr2:185777207	T>G	chr6:26924050	A>G	chr7:2103637	A>C
chr1:191089307	A>C	chr10:123912580	A>G	chr2:186044627	A>G	chr6:26987900	A>T	chr7:2111184	T>C
chr1:191090355	T>C	chr10:18601928	T>C	chr2:189750807	T>C	chr6:27034181	A>G	chr7:2112506	A>G
chr1:191090626	T>C	chr10:18727999	A>G	chr2:193751359	T>C	chr6:27034253	T>G	chr7:2112548	T>C
chr1:191091298	T>C	chr10:18763194	T>C	chr2:193751516	T>C	chr6:27103580	A>G	chr7:2120758	A>G
chr1:201012597	A>G	chr10:18763343	C>G	chr2:193848340	A>C	chr6:27103654	A>G	chr7:2138109	T>C
chr1:201020360	A>G	chr10:18763386	T>C	chr2:193917139	A>T	chr6:27114052	T>G	chr7:2138296	A>G
chr1:204599295	A>G	chr10:3819714	T>C	chr2:193917638	C>G	chr6:27115786	A>G	chr7:2139286	T>C
chr1:204599461	A>G	chr10:3820787	T>C	chr2:194344883	T>C	chr6:27143883	T>C	chr7:2139457	T>C
chr1:204968787	T>C	chr10:3821561	A>G	chr2:194344987	T>C	chr6:27145311	T>C	chr7:2139990	A>G

chr1:205015284	T>C	chr10:88112939	A>G	chr2:194377270	T>C	chr6:27145341	T>C	chr7:2140312	T>C
chr1:205015699	T>C	chr11:112263434	C>G	chr2:194377529	A>G	chr6:27219491	C>G	chr7:2140330	A>G
chr1:205015799	A>T	chr11:112263503	T>C	chr2:194377887	A>T	chr6:27356527	T>C	chr7:2141239	A>G
chr1:205035455	A>G	chr11:112263531	A>G	chr2:194428327	A>G	chr6:27446359	A>T	chr7:2143208	A>C
chr1:205037354	A>G	chr11:113344912	T>C	chr2:194430541	T>C	chr6:27446566	T>C	chr7:2144378	T>C
chr1:205039295	A>C	chr11:113377488	A>C	chr2:194432127	T>C	chr6:27560805	T>C	chr7:2145629	T>C
chr1:205041542	T>C	chr11:113424042	T>C	chr2:194437206	A>T	chr6:27586220	T>C	chr7:2145640	C>G
chr1:205042583	T>C	chr11:113425430	A>G	chr2:194449488	T>C	chr6:27640246	A>G	chr7:2145720	C>G
chr1:205044842	T>C	chr11:113436072	A>G	chr2:194449877	A>G	chr6:27720712	C>G	chr7:2149967	A>G
chr1:205046894	T>C	chr11:113448762	A>C	chr2:194456118	C>G	chr6:27722064	T>C	chr7:2151633	T>C
chr1:205048482	A>G	chr11:123395864	A>G	chr2:194492351	T>C	chr6:27729706	A>T	chr7:2155250	C>G
chr1:205070178	T>C	chr11:123395987	T>C	chr2:194524270	A>G	chr6:27730064	T>C	chr7:2155251	C>G
chr1:205073842	A>G	chr11:123396674	C>G	chr2:194524775	A>C	chr6:27730082	T>C	chr7:2156338	A>T
chr1:205074019	A>G	chr11:124606285	T>C	chr2:194525437	T>C	chr6:27730334	T>C	chr7:2159437	T>C
chr1:205077033	T>C	chr11:124606484	T>C	chr2:194658126	A>G	chr6:27730463	A>G	chr7:2159594	C>G
chr1:205078660	T>C	chr11:124628619	T>C	chr2:198152657	T>C	chr6:27740300	A>C	chr7:2159746	T>C
chr1:205079915	A>G	chr11:124633529	T>C	chr2:198152786	T>C	chr6:27740666	T>C	chr7:2159817	A>G
chr1:205080963	T>C	chr11:124633640	A>G	chr2:198154547	A>G	chr6:27741825	T>C	chr7:2160931	T>C
chr1:205081583	T>C	chr11:130732100	A>T	chr2:198158151	C>G	chr6:27742035	A>G	chr7:2166072	T>C
chr1:205084120	A>G	chr11:130732739	T>G	chr2:198158171	T>G	chr6:27759115	C>G	chr7:2166233	T>C
chr1:205085358	T>C	chr11:130733185	T>C	chr2:198160501	T>C	chr6:27774824	A>G	chr7:2166319	T>G
chr1:205087683	A>T	chr11:130733365	A>G	chr2:198170688	A>G	chr6:27775028	T>C	chr7:2166514	T>C
chr1:205141232	A>G	chr11:130733393	T>G	chr2:198171058	T>G	chr6:27775674	A>G	chr7:2168495	T>C
chr1:205144014	T>C	chr11:130735482	T>C	chr2:198171098	T>C	chr6:27775697	C>G	chr7:2180753	A>C
chr1:205145005	T>C	chr11:130735691	A>T	chr2:198172148	T>G	chr6:27779506	T>C	chr7:2180885	A>G
chr1:205146460	A>G	chr11:130738568	A>C	chr2:198172336	T>C	chr6:27781872	C>G	chr7:2181628	A>G
chr1:205146726	A>G	chr11:130738946	A>T	chr2:198175179	C>G	chr6:27782031	T>G	chr7:2183219	T>C
chr1:205148592	T>C	chr11:130741761	A>G	chr2:198175297	C>G	chr6:27782528	T>C	chr7:2183702	A>G
chr1:205151293	T>C	chr11:130741940	A>G	chr2:198205840	A>C	chr6:27782535	T>C	chr7:2184060	T>C
chr1:205153454	A>G	chr11:130742155	T>G	chr2:198243749	A>T	chr6:27782538	A>T	chr7:2184505	A>G
chr1:205157100	T>C	chr11:130742630	A>G	chr2:198317783	T>C	chr6:27783359	A>G	chr7:2184673	A>G
chr1:205163057	A>G	chr11:130742788	T>C	chr2:198356065	T>C	chr6:27791865	T>G	chr7:2190100	A>G
chr1:205164580	T>C	chr11:130743388	T>C	chr2:198362018	T>C	chr6:27792640	T>C	chr7:2203692	C>G
chr1:205179130	A>G	chr11:130744767	T>G	chr2:198362524	T>C	chr6:277974716	A>C	chr7:2206110	A>G
chr1:207931913	A>G	chr11:130744977	A>G	chr2:198362851	A>C	chr6:27798257	T>C	chr7:2212758	A>C
chr1:207950490	T>C	chr11:130745045	T>C	chr2:198365043	C>G	chr6:27798887	C>G	chr7:24769231	T>C
chr1:207957555	T>C	chr11:130745591	A>G	chr2:198369370	T>C	chr6:27799470	A>G	chr7:24769241	C>G
chr1:207965427	A>G	chr11:130745719	T>C	chr2:198370338	A>G	chr6:27799514	T>C	chr7:24797764	T>C
chr1:207976205	A>T	chr11:130746146	T>C	chr2:198370340	A>G	chr6:27799520	A>T	chr7:24828055	T>C
chr1:207977083	A>G	chr11:130747231	C>G	chr2:198372003	T>C	chr6:27799527	T>C	chr7:2702816	A>G
chr1:207978780	A>G	chr11:130747386	C>G	chr2:198372035	T>C	chr6:27804934	T>C	chr7:82439757	T>C
chr1:207992546	A>G	chr11:130748131	A>G	chr2:198380635	T>C	chr6:27805255	A>C	chr7:82440193	T>C
chr1:207993637	A>G	chr11:130749059	A>G	chr2:198436103	A>G	chr6:27806985	A>G	chr7:82447245	A>C
chr1:207999605	T>C	chr11:130758113	A>G	chr2:198437386	C>G	chr6:27833174	A>G	chr7:86452924	T>C
chr1:208000625	T>C	chr11:130758152	T>G	chr2:198463090	T>G	chr6:27834085	A>G	chr7:86571304	A>C
chr1:208002010	A>T	chr11:130758153	A>T	chr2:198478320	A>G	chr6:27834139	C>G	chr7:86571305	A>G
chr1:208006429	A>T	chr11:130758556	A>G	chr2:198493703	A>T	chr6:27835218	A>G	chr7:86758464	T>G
chr1:208006749	A>G	chr11:130759410	A>G	chr2:198541398	T>C	chr6:27835272	T>C	chr7:86782139	A>G
chr1:208017360	A>C	chr11:130763995	C>G	chr2:198542619	T>C	chr6:27835435	A>G	chr7:86782538	C>G
chr1:208018074	T>C	chr11:130764409	T>C	chr2:198545592	C>G	chr6:27835772	A>G	chr7:86782797	T>C
chr1:214145706	T>C	chr11:130764540	A>T	chr2:198668751	T>C	chr6:27839746	T>C	chr7:86782805	T>C
chr1:214147889	T>C	chr11:130767486	T>G	chr2:198669754	C>G	chr6:27840926	T>C	chr7:86784271	A>C
chr1:214148628	A>G	chr11:130786308	T>C	chr2:198670478	A>T	chr6:27850714	A>T	chr7:86843375	T>C
chr1:214148772	T>G	chr11:130786345	A>C	chr2:198671145	T>C	chr6:27850873	A>G	chr7:86860482	C>G
chr1:214149028	A>T	chr11:132562494	A>G	chr2:198715972	A>G	chr6:27858421	A>G	chr8:10009949	C>G
chr1:214149061	T>G	chr11:133817009	T>C	chr2:198718630	A>G	chr6:27858837	T>C	chr8:10009952	T>G
chr1:214154547	T>G	chr11:133817037	T>C	chr2:198718667	C>G	chr6:27858904	T>C	chr8:10030698	A>G
chr1:214155398	A>C	chr11:133817055	A>G	chr2:198721560	A>T	chr6:27859453	T>C	chr8:10030902	T>G
chr1:214158986	C>G	chr11:133817223	A>G	chr2:198738019	A>G	chr6:27859568	T>C	chr8:10190040	T>C
chr1:214160529	A>G	chr11:133817225	T>C	chr2:198743655	T>G	chr6:27862127	A>G	chr8:10193772	C>G
chr1:214160548	A>T	chr11:133817333	A>G	chr2:198754244	T>C	chr6:27862152	T>C	chr8:10220412	A>C
chr1:214162573	A>T	chr11:133817792	T>C	chr2:198776546	A>G	chr6:28048535	A>G	chr8:10226355	T>C
chr1:214162734	T>G	chr11:133818709	T>C	chr2:198777676	A>T	chr6:28048538	A>G	chr8:10253098	A>G
chr1:214163069	T>C	chr11:133819417	T>G	chr2:198789765	A>C	chr6:28048624	T>C	chr8:10255901	A>G
chr1:214163675	A>G	chr11:133821132	A>C	chr2:198798630	A>G	chr6:28048996	T>G	chr8:10256054	A>G
chr1:226790001	T>C	chr11:133821144	A>G	chr2:198805098	T>C	chr6:28058635	A>G	chr8:10257041	T>G
chr1:230272624	A>G	chr11:134296384	T>C	chr2:198825462	A>G	chr6:28058949	C>G	chr8:10260921	A>T
chr1:2369498	T>C	chr11:134296641	A>G	chr2:198839145	A>G	chr6:28059020	A>G	chr8:10267540	A>C
chr1:2372858	T>C	chr11:134297110	A>G	chr2:198896895	A>G	chr6:28059217	C>G	chr8:10268805	A>G
chr1:2372954	A>G	chr11:14405921	A>C	chr2:198902053	C>G	chr6:28072602	A>G	chr8:10283426	A>G
chr1:2373089	T>G	chr11:14406140	T>C	chr2:198920142	T>C	chr6:28075911	A>G	chr8:10283602	A>C
chr1:2373168	A>G	chr11:14406212	A>G	chr2:198920560	A>C	chr6:28092307	C>G	chr8:10283748	T>C
chr1:2374120	T>C	chr11:14406264	A>G	chr2:198921201	A>G	chr6:28092603	T>G	chr8:111476741	A>C
chr1:2374758	A>G	chr11:14408311	T>G	chr2:198923468	C>G	chr6:28092685	T>C	chr8:111571834	T>C
chr1:2375506	T>C	chr11:14408398	T>G	chr2:198940251	T>C	chr6:28104476	T>C	chr8:111598096	T>G
chr1:2375844	T>G	chr11:14408424	T>G	chr2:200051390	T>C	chr6:28104634	T>G	chr8:111601659	A>G
chr1:2379248	A>G	chr11:14408483	T>G	chr2:200051520	A>C	chr6:28105196	T>C	chr8:111601741	T>C
chr1:2379705	A>G	chr11:14408952	T>C	chr2:200054222	A>G	chr6:28108967	T>C	chr8:111972205	T>G
chr1:2381539	T>C	chr11:144089527	A>G	chr2:200072082	A>T	chr6:28109262	T>C	chr8:12794181	A>G
chr1:2381568	A>G	chr11:144092438	A>G	chr2:200126019	T>G	chr6:28110148	T>C	chr8:12810483	C>G
chr1:2383039	A>G	chr11:146352266	C>G	chr2:200126042	A>T	chr6:28129231	T>C	chr8:12810504	T>C
chr1:2383756	A>G	chr11:146355932	C>G	chr2:200126137	T>C	chr6:28129232	T>C	chr8:143297312	T>C
chr1:2384108	T>C	chr11:146361978	C>G	chr2:200185567	C>G	chr6:28129580	T>C	chr8:143297663	T>C

chr1:2387101	T>C	chr11:46368241	A>T	chr2:200194271	T>C	chr6:28129713	A>C	chr8:143335830	A>G
chr1:2390715	C>G	chr11:46368559	T>G	chr2:200314206	T>G	chr6:28129789	T>G	chr8:143364744	A>G
chr1:2390840	T>C	chr11:46370767	A>G	chr2:200677625	T>C	chr6:28129831	T>C	chr8:143365481	C>G
chr1:2391252	A>C	chr11:46373311	T>C	chr2:200695901	A>G	chr6:28130089	T>C	chr8:143370923	A>T
chr1:239208470	A>C	chr11:46376094	A>C	chr2:200697539	A>T	chr6:28175549	C>G	chr8:143388290	T>C
chr1:239209134	A>T	chr11:46377770	T>C	chr2:200714766	T>C	chr6:28193102	A>C	chr8:143399860	T>G
chr1:239210058	A>C	chr11:46378032	A>G	chr2:200715388	T>G	chr6:28193786	T>G	chr8:143720544	A>G
chr1:239211820	T>C	chr11:46390680	A>C	chr2:200716119	A>C	chr6:28219270	A>G	chr8:17049377	A>C
chr1:2392648	C>G	chr11:46391602	T>G	chr2:200775125	T>C	chr6:28219661	A>G	chr8:17050344	A>G
chr1:2397319	A>G	chr11:46402852	A>G	chr2:200787719	T>C	chr6:28219826	A>G	chr8:17052736	A>G
chr1:2398376	A>G	chr11:46438715	A>C	chr2:200787804	A>G	chr6:28234597	T>C	chr8:17053502	A>G
chr1:2398763	T>C	chr11:46484685	A>G	chr2:200819794	A>G	chr6:28235176	A>C	chr8:17053515	A>T
chr1:2399119	A>G	chr11:46486059	A>G	chr2:200820505	T>G	chr6:282302784	A>G	chr8:18422923	T>C
chr1:2399149	A>G	chr11:46506885	T>C	chr2:200838613	A>G	chr6:282303464	C>G	chr8:26231597	A>G
chr1:2402499	A>G	chr11:46513249	T>C	chr2:200850402	A>G	chr6:282304384	A>C	chr8:26250047	A>G
chr1:243376757	T>C	chr11:46582072	A>C	chr2:200915395	T>C	chr6:282318331	T>G	chr8:27437573	T>C
chr1:243377707	C>G	chr11:46622392	T>C	chr2:200915716	T>C	chr6:282321993	T>C	chr8:27453579	A>G
chr1:243401997	A>C	chr11:46639582	A>G	chr2:200916004	A>G	chr6:282322120	C>G	chr8:27453762	T>C
chr1:243418063	A>C	chr11:46657430	T>G	chr2:200916230	T>C	chr6:28232296	A>G	chr8:27454682	A>G
chr1:243418182	T>C	chr11:46684524	A>T	chr2:200916239	T>C	chr6:282323702	A>G	chr8:34126948	T>C
chr1:243419429	A>T	chr11:46684677	T>G	chr2:200916403	T>C	chr6:282323938	T>C	chr8:38126359	C>G
chr1:243420388	A>G	chr11:46686260	A>G	chr2:200941367	T>C	chr6:28234378	T>C	chr8:38137530	A>G
chr1:243421404	A>G	chr11:46691737	A>G	chr2:201051150	A>G	chr6:28234397	T>C	chr8:38152592	C>G
chr1:243428152	T>G	chr11:46693169	A>G	chr2:201139529	C>G	chr6:28235201	C>G	chr8:38153571	A>T
chr1:243430818	T>C	chr11:46695483	A>G	chr2:201171129	A>G	chr6:28235308	T>C	chr8:38161469	T>C
chr1:243431654	T>C	chr11:46697142	T>C	chr2:201172999	T>C	chr6:28351728	T>C	chr8:38176187	T>G
chr1:243433141	C>G	chr11:46705974	A>T	chr2:201173048	A>G	chr6:28365515	T>C	chr8:38201232	T>C
chr1:243433654	A>G	chr11:46922596	T>C	chr2:201173100	C>G	chr6:28366151	A>G	chr8:38206242	T>C
chr1:243437266	T>C	chr11:46923107	T>C	chr2:201173620	A>G	chr6:28367355	A>G	chr8:38240008	A>G
chr1:243440047	T>C	chr11:46927877	T>C	chr2:201182065	T>C	chr6:28367623	C>G	chr8:38240392	T>C
chr1:243443304	C>G	chr11:46928455	A>G	chr2:201221350	A>T	chr6:28367658	A>G	chr8:38240498	A>G
chr1:243444176	T>C	chr11:46934369	T>C	chr2:201222403	A>T	chr6:28367663	T>C	chr8:38243128	A>G
chr1:243446994	T>C	chr11:46935185	C>G	chr2:201244261	A>C	chr6:28367683	T>G	chr8:38243227	A>T
chr1:243449881	A>G	chr11:46958299	T>G	chr2:201245272	A>G	chr6:28411220	T>G	chr8:38243865	T>G
chr1:243450955	T>G	chr11:47013441	A>G	chr2:201313988	T>C	chr6:28411244	A>G	chr8:38258154	A>G
chr1:243451405	A>C	chr11:47120066	T>C	chr2:201314326	T>C	chr6:28411303	A>G	chr8:38259035	A>C
chr1:243451853	A>T	chr11:47198638	T>C	chr2:2020041928	T>C	chr6:28414967	T>G	chr8:38259481	T>C
chr1:243451974	T>C	chr11:47207362	A>G	chr2:2020042627	A>C	chr6:28457818	A>G	chr8:38269514	T>C
chr1:243462417	T>G	chr11:574044779	A>G	chr2:2020042675	A>G	chr6:28521570	A>C	chr8:38272542	C>G
chr1:243464416	A>G	chr11:57434122	A>C	chr2:225450161	C>G	chr6:28522695	A>G	chr8:38272582	A>C
chr1:243458922	A>G	chr11:57435211	T>C	chr2:22553120	T>C	chr6:28554543	T>G	chr8:38286811	C>G
chr1:243462367	A>G	chr11:57435296	A>G	chr2:22673506	T>C	chr6:28557133	A>T	chr8:38287555	T>C
chr1:243462417	T>G	chr11:57480623	A>C	chr2:22674222	T>C	chr6:28566522	T>C	chr8:38288566	C>G
chr1:243464416	A>G	chr11:57507983	T>C	chr2:22674261	C>G	chr6:28586282	A>T	chr8:38290424	A>G
chr1:243467236	A>G	chr11:57508678	C>G	chr2:22723287	C>G	chr6:28602631	A>G	chr8:38298647	T>C
chr1:243467689	T>G	chr11:57547000	T>C	chr2:229272784	A>G	chr6:28624894	A>T	chr8:38299624	T>C
chr1:243468694	A>C	chr11:57560452	A>C	chr2:229297803	T>C	chr6:28641957	A>G	chr8:4177231	A>T
chr1:243469029	A>G	chr11:63651036	A>T	chr2:229297832	A>T	chr6:28662521	T>C	chr8:4188832	C>G
chr1:243471192	A>G	chr11:63785348	A>G	chr2:229308795	A>G	chr6:28725804	C>G	chr8:4188891	A>G
chr1:243471949	A>C	chr11:63792768	A>G	chr2:229308855	T>C	chr6:28757489	T>C	chr8:4189745	C>G
chr1:243472187	A>C	chr11:64684283	T>C	chr2:229308856	A>G	chr6:28757555	C>G	chr8:60544006	A>C
chr1:243472664	A>G	chr11:64684990	T>C	chr2:233554499	T>C	chr6:28775322	A>G	chr8:60554402	T>C
chr1:243474536	A>C	chr11:64685314	T>C	chr2:233561743	C>G	chr6:28775564	A>G	chr8:60554538	A>C
chr1:243483279	A>T	chr11:64685482	A>G	chr2:23562197	T>C	chr6:28775982	C>G	chr8:60556509	A>C
chr1:243483368	T>C	chr11:64691910	T>G	chr2:23562731	C>G	chr6:28805638	T>G	chr8:60556826	A>C
chr1:243485126	T>G	chr11:64692312	A>G	chr2:23631062	T>C	chr6:28806218	A>G	chr8:60557136	T>C
chr1:243487861	A>G	chr11:64692356	T>C	chr2:23640871	T>G	chr6:28806303	C>G	chr8:60591355	A>T
chr1:243488186	C>G	chr11:65378028	T>C	chr2:23641010	A>T	chr6:28806334	T>C	chr8:60591582	A>T
chr1:243492567	A>G	chr11:65380248	T>C	chr2:23641083	C>G	chr6:28831611	T>C	chr8:60667231	A>G
chr1:243493907	A>T	chr11:65382564	T>C	chr2:23641924	T>C	chr6:28833627	A>T	chr8:60667638	A>G
chr1:243498108	A>G	chr11:65383755	T>G	chr2:23642230	A>T	chr6:28834142	C>G	chr8:60667640	A>T
chr1:243498312	T>C	chr11:81204480	T>C	chr2:23660704	T>G	chr6:28855805	A>C	chr8:60667755	T>G
chr1:243499744	C>G	chr11:83107194	T>C	chr2:237318440	A>G	chr6:28863264	C>G	chr8:60668254	T>C
chr1:243499758	T>C	chr11:83134888	T>C	chr2:237318877	T>C	chr6:28863369	A>G	chr8:60875600	T>C
chr1:243499947	T>C	chr11:83135151	T>C	chr2:237341256	C>G	chr6:28864739	C>G	chr8:60880161	A>G
chr1:243500397	A>G	chr11:83135392	A>G	chr2:237367037	A>G	chr6:28864849	A>G	chr8:60880646	C>G
chr1:243500994	T>C	chr11:83187113	A>G	chr2:23792565	A>G	chr6:28891176	T>C	chr8:8093511	A>G
chr1:243501047	A>C	chr11:83221390	A>C	chr2:23792972	A>C	chr6:28891839	A>G	chr8:8093578	T>C
chr1:243501056	A>C	chr11:83231605	A>G	chr2:23793211	T>C	chr6:28892919	A>T	chr8:8093592	T>G
chr1:243501763	A>C	chr11:83233298	A>G	chr2:237937644	T>C	chr6:28908214	A>T	chr8:8093967	A>C
chr1:243502623	A>G	chr12:103361112	C>G	chr2:23726719	A>C	chr6:28911802	A>G	chr8:8094114	A>C
chr1:243503764	A>T	chr12:103601638	A>G	chr2:23726720	T>G	chr6:28949960	T>C	chr8:8094116	C>G
chr1:243505047	A>G	chr12:103601765	A>G	chr2:25352200	T>C	chr6:28956416	A>G	chr8:8105359	T>G
chr1:243505841	A>G	chr12:110493222	A>G	chr2:27974971	A>G	chr6:28982272	T>C	chr8:8105405	T>C
chr1:243505950	T>G	chr12:110662327	T>C	chr2:27975394	T>C	chr6:28982823	T>C	chr8:8130538	A>T
chr1:243513047	A>G	chr12:110705800	A>G	chr2:28019367	A>G	chr6:28983274	C>G	chr8:8140901	T>C
chr1:243518691	A>G	chr12:11052465	A>T	chr2:28019856	T>C	chr6:29415464	A>C	chr8:8146953	T>G
chr1:243522906	A>G	chr12:111116655	T>C	chr2:28019966	T>C	chr6:29521445	T>C	chr8:8168222	A>G
chr1:243537729	T>G	chr12:123450765	C>G	chr2:28020157	T>C	chr6:29623739	A>G	chr8:8168394	A>G
chr1:243541698	A>T	chr12:123451018	C>G	chr2:28021845	C>G	chr6:29629344	T>C	chr8:8168413	A>C
chr1:243544827	C>G	chr12:123469647	T>G	chr2:28022675	A>G	chr6:29670250	A>G	chr8:8171595	A>G
chr1:243547345	T>C	chr12:123516917	C>G	chr2:28023120	T>C	chr6:29670261	A>G	chr8:8182270	C>G
chr1:243555219	A>G	chr12:123541606	T>C	chr2:28023284	T>C	chr6:29670556	A>G	chr8:8241772	T>C

chr1:243589156	T>G	chr12:123550813	A>G	chr2:37576136	A>G	chr6:29670573	T>C	chr8:8260445	T>C
chr1:243598234	A>G	chr12:123560289	T>C	chr2:48573895	C>G	chr6:29670652	A>G	chr8:8301794	A>G
chr1:243598670	A>G	chr12:123568619	A>C	chr2:48684796	A>G	chr6:29673882	T>C	chr8:89269492	A>T
chr1:243608967	A>G	chr12:123569375	A>G	chr2:48707841	C>G	chr6:29716300	A>G	chr8:89302448	A>G
chr1:243609927	T>C	chr12:123590165	T>C	chr2:5462987	T>G	chr6:29716390	T>G	chr8:89337036	T>C
chr1:243620356	A>C	chr12:123591596	T>G	chr2:57948635	A>G	chr6:29716589	T>C	chr8:89338155	T>C
chr1:243639218	T>C	chr12:123593382	T>C	chr2:57952870	A>C	chr6:29716847	T>C	chr8:89339721	T>G
chr1:243639859	A>G	chr12:123593485	A>T	chr2:57972834	A>C	chr6:29716901	T>G	chr8:89340162	C>G
chr1:243640881	A>G	chr12:123595163	A>G	chr2:57975524	T>G	chr6:29720997	A>G	chr8:89390692	A>G
chr1:243645203	C>G	chr12:123627905	T>C	chr2:57975714	A>G	chr6:29759750	A>G	chr8:89390946	A>G
chr1:243654278	A>G	chr12:123632276	A>G	chr2:57987201	C>G	chr6:29759811	C>G	chr8:89428098	A>T
chr1:243655452	A>G	chr12:123632367	C>G	chr2:57987593	T>C	chr6:29759823	T>C	chr8:89468284	T>C
chr1:243655897	T>C	chr12:123632930	A>G	chr2:57988194	A>T	chr6:29759824	T>C	chr8:89469693	A>G
chr1:243663893	T>C	chr12:123633057	T>C	chr2:58110969	A>C	chr6:29759876	A>G	chr8:89518058	C>G
chr1:243664642	A>G	chr12:123633382	T>C	chr2:58112266	A>T	chr6:29759885	A>G	chr8:89544326	A>C
chr1:243664857	T>C	chr12:123633424	A>G	chr2:58113995	T>C	chr6:29759923	T>C	chr8:89555353	T>C
chr1:243669012	T>C	chr12:123634043	T>C	chr2:58117926	C>G	chr6:29759996	T>C	chr8:89566662	T>G
chr1:243669350	C>G	chr12:123634122	A>G	chr2:58119122	A>G	chr6:29760111	C>G	chr8:89566677	T>G
chr1:243672000	C>G	chr12:123635098	T>C	chr2:58119574	T>C	chr6:29760188	A>G	chr8:89566799	C>G
chr1:243672074	T>C	chr12:123636841	A>G	chr2:58119582	T>G	chr6:29760510	A>G	chr8:89566903	A>G
chr1:243672125	A>G	chr12:123648881	T>C	chr2:58119657	C>G	chr6:29855062	A>G	chr8:89606888	T>C
chr1:243689543	A>C	chr12:123665695	A>G	chr2:58120553	A>T	chr6:29856194	C>G	chr8:89644431	T>C
chr1:243690491	T>C	chr12:123716930	A>T	chr2:58121391	T>C	chr6:29856303	T>C	chr8:89691016	T>G
chr1:243703982	T>C	chr12:123717216	A>G	chr2:58133522	A>G	chr6:29856439	T>G	chr8:89760311	A>G
chr1:243712455	A>C	chr12:123717354	A>C	chr2:58133686	T>G	chr6:29856445	T>C	chr8:89760620	A>G
chr1:243717592	A>T	chr12:123717362	A>G	chr2:58135010	C>G	chr6:29893922	A>T	chr8:9777555	T>C
chr1:243717682	A>G	chr12:123718301	T>C	chr2:58135011	A>G	chr6:29893960	T>C	chr8:9966586	A>G
chr1:243723214	A>C	chr12:123726834	T>C	chr2:58135023	T>C	chr6:29893982	T>C	chr9:122546769	T>G
chr1:243732197	A>G	chr12:123727443	A>G	chr2:58135872	A>G	chr6:29894007	A>T	chr9:131580744	T>C
chr1:243734296	A>G	chr12:123735937	A>G	chr2:58136220	C>G	chr6:29894211	T>C	chr9:131591832	A>G
chr1:243743379	C>G	chr12:123736084	T>G	chr2:58138581	A>G	chr6:29894392	T>G	chr9:26713335	T>G
chr1:243744321	T>C	chr12:123738494	T>C	chr2:58145060	A>G	chr6:29894584	A>C	chr9:26757269	A>C
chr1:243744904	A>G	chr12:123738678	A>G	chr2:58149116	A>G	chr6:29894600	T>G	chr9:37079661	T>G
chr1:243746634	T>C	chr12:123738876	T>C	chr2:58149158	T>C	chr6:29894863	A>G	chr9:7171606	T>G
chr1:243747358	A>G	chr12:123739111	A>G	chr2:58200748	A>G	chr6:29895032	C>G	chr9:7172498	A>C
chr1:243748184	C>G	chr12:123743445	A>G	chr2:58380658	C>G	chr6:29895099	A>C	chr9:77342815	C>G
chr1:243748629	T>C	chr12:123743447	A>G	chr2:58420885	A>G	chr6:29909797	T>C	chr9:77342863	A>T
chr1:243753910	T>C	chr12:123743674	T>C	chr2:59153054	A>T	chr6:29910663	A>G	chr9:77358745	T>C
chr1:243759210	T>C	chr12:123743883	C>G	chr2:73164008	T>C	chr6:29910731	A>G	chr9:84890893	T>C
chr1:243762299	T>C	chr12:123743927	A>G	chr2:73165234	T>C	chr6:29910910	T>C	chr9:84935618	T>C
chr1:243767493	A>G	chr12:123745178	C>G	chr2:73211516	A>T	chr6:29911222	T>C	chr9:84935895	T>C
chr1:243768906	T>C	chr12:123745359	A>G	chr2:73403620	T>C	chr6:29911246	T>C	chr9:85064583	A>G
chr1:243770727	A>G	chr12:123745809	T>G	chr2:73460517	T>G	chr6:30042209	T>G	chr9:85082899	T>C
chr1:243776117	T>C	chr12:123750073	T>C	chr2:73461553	T>C	chr6:30063652	C>G	chr9:96213146	A>T
chr1:243779250	C>G	chr12:123750228	T>C	chr2:73652628	A>G	chr6:30063674	C>G	chr9:96213529	A>G

## 8.4 Fragments of the ValLib

The following control fragments were included: betaActin1, betaActin2, betaActin3, betaActin4, betaActin5, betaActin6, betaActin7, betaActin8, betaActin9, SV40.promoter.enhancer26, CMV.enhancer.promoter27, CMV.enhancer.promoter28, UBC.human16, UBC.human17, UBC.human18, UBC.human19, UBC.human20, UBC.human21, UBC.human22, GAPDH.010, GAPDH23, GAPDH24, GAPDH25, GAPDH.111, GAPDH.212, GAPDH.313, GAPDH.414, GAPDH.515, X.133576343.133576503.Neg.4.032, Random.133, Random.234, Random.335, Random.436, Random.537, Random.638, Random.739, Random.840, Random.941, Random.1042, Random.1143, Random.1244, Random.1345, Random.1446, Random.1547, Random.1648, Random.1749, Random.1850, Random.1951, Random.2052, Random.2153, Random.2254, Random.2355, Random.2456, 3.181680650.181680810.Neg.7.1329, 6.45611355.45611515.Neg.5.330, and 6.45612315.45612475.Neg.5.931. Fragments were 170 nt in length.

Chr.	Start	End	Var. Pos.	Ref.	Alt.
chr1	150335395	150335565	86	CG	C
chr1	16352547	16352717	85	C	T
chr1	182423146	182423316	85	G	A
chr1	205015284	205015454	85	T	C
chr1	2372321	2372491	85	C	T
chr1	2375844	2376014	85	T	G
chr1	2387101	2387271	85	C	T
chr1	2392648	2392818	85	G	C
chr1	240233471	240233641	85	A	G
chr1	243419429	243419599	85	T	A
chr1	243433141	243433311	85	C	G
chr1	30456802	30456972	87	G	GTCCT
chr1	3058249	3058419	85	A	G

Chr.	Start	End	Var. Pos.	Ref.	Alt.
chr18	53067184	53067354	85	G	A
chr18	53068680	53068850	85	T	C
chr18	53070168	53070338	85	C	T
chr18	53071766	53071936	85	A	G
chr18	53074618	53074788	85	A	G
chr18	53076736	53076906	85	C	T
chr18	53076790	53076960	85	G	A
chr18	53085273	53085443	85	T	C
chr18	53085392	53085562	85	C	T
chr18	53086622	53086792	87	A	AC
chr18	53087984	53088154	85	T	A
chr18	53091037	53091207	86	T	TC
chr18	53125474	53125644	85	G	C

chr1	36637708	36637878	85	A	G
chr1	38096994	38097164	85	T	C
chr1	50579237	50579407	85	A	C
chr1	69240478	69240648	85	C	A
chr1	8431607	8431777	85	C	T
chr1	8468278	8468448	85	A	G
chr1	98465528	98465698	85	G	A
chr10	100011970	100012140	85	G	A
chr10	104594507	104594677	85	A	G
chr10	104630412	104630582	85	A	T
chr10	104680137	104680307	85	T	A
chr10	104748009	104748179	85	G	A
chr10	104814162	104814332	85	T	C
chr10	104844872	104845042	85	T	C
chr10	104941589	104941759	85	C	T
chr10	104952499	104952669	85	C	T
chr10	104962011	104962181	90	CA	C
chr10	104963893	104964063	85	C	T
chr10	111957646	111957816	85	G	T
chr10	11444458	11444628	85	C	G
chr10	119991443	119991613	85	T	C
chr10	120307179	120307349	85	G	A
chr10	2971405	2971575	85	C	T
chr10	31158591	31158761	85	A	G
chr10	50471012	50471182	85	T	A
chr10	61136285	61136455	85	C	T
chr10	74151574	74151744	85	C	T
chr11	130719075	130719245	92	A	ACTCC
chr11	130745591	130745761	85	A	G
chr11	132044308	132044478	85	C	T
chr11	132271023	132271193	85	G	A
chr11	133817055	133817225	85	A	G
chr11	133822133	133822303	85	G	A
chr11	133822569	133822739	85	A	G
chr11	20683150	20683320	85	A	C
chr11	29279371	29279541	85	T	G
chr11	30355707	30355877	85	A	C
chr11	46368241	46368411	85	A	T
chr11	57434122	57434292	85	C	A
chr11	57481155	57481325	85	G	A
chr11	64684990	64685160	85	C	T
chr11	65382564	65382734	85	T	C
chr11	86841950	86842120	85	T	C
chr12	123450765	123450935	85	G	C
chr12	123591596	123591766	85	G	T
chr12	123634043	123634213	85	T	C
chr12	123716513	123716683	85	C	T
chr12	123743883	123744053	85	G	C
chr12	124480866	124481036	85	G	A
chr12	2288405	2288575	85	T	C
chr12	2288945	2289115	85	T	A
chr12	2298724	2298894	85	T	C
chr12	2298974	2299144	86	A	AG
chr12	2300207	2300377	85	A	G
chr12	2315228	2315398	85	G	A
chr12	2333671	2333841	85	G	A
chr12	2339250	2339420	85	G	A
chr12	2339252	2339422	87	GC	G
chr12	2339253	2339423	85	C	G
chr12	2350452	2350622	85	A	G
chr12	29935396	29935566	85	C	G
chr12	57323523	57323693	85	A	G
chr12	57331741	57331911	85	G	A
chr12	57369910	57370080	85	G	A
chr12	57428353	57428523	85	A	G
chr12	57438658	57438828	85	T	G
chr12	57486647	57486817	85	A	G
chr12	57487729	57487899	85	T	G
chr12	57487814	57487984	85	A	C
chr12	57487814	57487984	85	A	C
chr12	57490100	57490270	85	T	C
chr12	57547749	57547919	85	T	C
chr12	57569648	57569648	85	C	T
chr12	57622371	57622541	85	G	T
chr12	57682956	57683126	85	C	T
chr12	57778221	57778391	85	C	T
chr12	57824165	57824335	85	G	T
chr12	74307402	74307572	85	A	C
chr12	99462554	99462724	103	CCCTAGT	C
chr13	37807152	37807322	85	T	C
chr13	95713908	95714078	85	A	G
chr14	103801561	103801731	87	A	AAC
chr14	103850906	103851076	85	G	C
chr14	103980639	103980809	89	A	AT
chr14	103984537	103984707	85	T	C
chr14	104008159	104008329	85	C	T
chr14	104021806	104021976	85	G	A
chr14	104193099	104193269	85	C	G

chr18	53125938	53126108	85	T	C
chr18	53126330	53126500	85	T	A
chr18	53144837	53145007	85	C	T
chr18	53213887	53214057	85	G	A
chr18	53215142	53215312	85	T	A
chr18	53215739	53215909	85	A	T
chr18	53251751	53251921	89	A	AAG
chr18	53252285	53252455	85	G	A
chr18	53253431	53253601	85	G	C
chr18	53445930	53446100	85	T	C
chr18	53449667	53449837	85	A	G
chr18	53542763	53542933	86	CT	C
chr18	53567855	53568025	85	G	A
chr19	19431423	19431593	85	T	G
chr19	33834680	33834850	85	T	C
chr19	43454099	43454269	85	A	T
chr19	50083561	50083731	93	AAGAG	A
chr19	50087876	50088046	85	T	C
chr2	01062603	01062773	85	T	G
chr2	18366999	18367169	85	T	C
chr2	188271119	18827289	85	T	C
chr2	33776904	33777074	85	A	G
chr2	98317783	98317953	85	C	T
chr2	98718667	98718837	85	G	C
chr2	99150081	99150251	85	A	G
chr2	00775125	775295	85	T	C
chr2	00916230	916400	85	T	C
chr2	00941367	941537	85	T	C
chr2	01024512	1024682	85	T	C
chr2	01057491	1057661	85	G	A
chr2	11934370	11934540	85	A	G
chr2	25448997	25449167	92	T	TC
chr2	2738509	2738679	85	A	G
chr2	3726720	3726890	85	G	T
chr2	39564772	39564942	85	T	G
chr2	8135009	8135179	86	TC	T
chr2	4849873	4850043	85	C	T
chr2	1040532	1040702	85	A	C
chr2	20732968	20733138	85	G	A
chr2	20828826	20828996	85	G	A
chr2	24748432	24748602	85	A	G
chr2	37434133	37434303	85	G	A
chr2	37437262	37437432	88	A	AC
chr2	37502637	37502807	85	T	G
chr2	37503317	37503487	85	T	A
chr2	39764702	39764872	85	T	C
chr2	43682799	43682969	85	C	T
chr2	58252407	58252577	85	T	C
chr2	58252537	58252707	85	C	T
chr2	40495139	40495309	85	A	C
chr2	42680201	42680371	85	T	C
chr2	46521225	46521395	85	T	C
chr3	109935229	109935399	85	A	T
chr3	117640660	117640830	85	C	T
chr3	143932344	143932514	85	G	T
chr3	151007310	151007480	85	C	T
chr3	152700642	152700812	85	T	C
chr3	170839276	170839446	85	G	A
chr3	17221017	17221187	85	T	C
chr3	36846396	36846566	85	C	A
chr3	52273421	52273591	85	G	A
chr3	52349204	52349374	85	A	G
chr3	52721305	52721475	85	G	A
chr3	52739520	52739690	85	T	C
chr3	86935862	86936032	85	A	G
chr3	87228835	87229005	85	T	C
chr4	108427928	108428098	85	A	T
chr4	11853430	11853600	85	C	T
chr4	119847772	119847942	85	A	G
chr4	127243178	127243348	85	G	A
chr4	129737089	129737259	85	G	A
chr4	139569264	139569434	85	C	T
chr4	14215834	14216004	85	G	C
chr4	153315591	153315761	85	A	G
chr4	3242272	3242442	85	C	A
chr4	45138459	45138629	85	A	C
chr4	57925402	57925572	85	T	C
chr5	125208685	125208855	85	C	G
chr5	137945751	137945921	85	C	G
chr5	137946500	137946670	85	A	G
chr5	137947196	137947366	91	G	GA
chr5	139064267	139064437	85	C	G
chr5	139064835	139065005	85	C	T
chr5	139071632	139071802	85	G	A
chr5	140186657	140186827	85	A	G
chr5	147126495	147126665	85	A	G
chr5	152056234	152056404	85	C	T
chr5	152871184	152871354	85	C	A
chr5	16397739	163977909	85	T	A

chr14	33715780	33715950	85	A	G
chr14	72448232	72448402	85	G	T
chr14	80001655	80001825	85	G	A
chr15	50405175	50405345	85	T	C
chr15	78731808	78731978	85	T	A
chr15	85196244	85196414	85	G	A
chr16	19067660	19067830	85	G	A
chr16	29931593	29931763	85	T	C
chr16	29931899	29932069	85	G	A
chr16	29932064	29932234	85	C	T
chr16	29932691	29932861	85	T	C
chr16	29934050	29934220	85	A	G
chr16	29936654	29936824	85	A	G
chr16	29937652	29937822	85	G	A
chr16	29938820	29938990	85	A	G
chr16	29972690	29972860	85	C	T
chr16	29974167	29974337	85	G	A
chr16	29984839	29985009	85	C	G
chr16	29986205	29986375	85	C	G
chr16	29986525	29986695	85	G	A
chr16	29988349	29988519	85	G	A
chr16	29988941	29989111	85	A	G
chr16	4258005	4258175	85	C	G
chr16	4460049	4460219	85	A	G
chr16	4573040	4573210	85	A	G
chr16	4573938	4574108	85	A	G
chr16	969918	970088	85	C	T
chr16	9914095	9914265	85	C	T
chr16	9914133	9914303	85	G	T
chr16	9914401	9914571	85	C	G
chr16	9914419	9914589	85	C	G
chr16	9915013	9915183	85	A	G
chr16	9916204	9916374	85	C	G
chr17	17661802	17661972	85	A	G
chr17	17696531	17696701	85	G	C
chr17	17741497	17741667	87	CCACG	C
chr17	17748013	17748183	85	A	G
chr17	17748645	17748815	85	T	C
chr17	17840823	17840993	85	G	T
chr17	17866897	17867067	85	A	G
chr17	17876296	17876466	85	C	G
chr17	19153175	19153345	85	G	C
chr17	19174874	19175044	85	T	C
chr17	19217645	19217815	86	G	GC
chr17	19223224	19223394	85	T	G
chr17	19224397	19224567	85	G	C
chr17	2084305	2084475	85	A	G
chr17	55728224	55728394	85	G	A
chr17	73797796	73797966	85	T	A
chr18	3523628	3523798	85	G	A
chr18	36610569	36610739	85	T	C
chr18	42907006	42907176	101	T	TCCAG
chr18	52968230	52968400	85	A	G
chr18	52987161	52987331	85	T	A
chr18	52999283	52999453	86	T	TG
chr18	53014632	53014802	85	C	G
chr18	53014633	53014803	85	C	T
chr18	53023170	53023340	85	G	A
chr18	53052923	53053093	85	C	T
chr18	53053398	53053568	85	G	A
chr18	53060415	53060585	85	T	C
chr18	53063676	53063846	85	G	T
chr18	53063719	53063889	85	G	A
chr18	53065892	53066062	85	A	G

chr5	17655528	17655698	85	G	A
chr5	18334297	18334467	85	G	C
chr5	25808995	25809165	85	G	A
chr5	53300865	53301035	88	A	AG
chr5	55104045	55104215	85	G	A
chr5	60736544	60736714	85	C	T
chr5	76820085	76820255	85	A	G
chr6	111845570	111845740	85	T	A
chr6	166515366	166515536	85	T	C
chr6	28324378	28324548	85	C	T
chr6	28367683	28367853	85	T	G
chr6	28602631	28602801	85	G	A
chr6	28726815	28726985	85	C	T
chr6	28831021	28831191	85	T	G
chr6	28884888	28885058	86	CG	C
chr6	31783507	31783677	85	G	C
chr6	31797587	31797757	85	C	G
chr6	31839309	31839479	85	C	T
chr6	33540209	33540379	85	A	G
chr6	33710933	33711103	85	G	A
chr6	43395101	43395271	85	G	A
chr6	56573656	56573826	85	T	C
chr6	67287611	67287781	85	A	C
chr6	69943057	69943227	94	A	AT
chr6	96464628	96464798	85	G	T
chr7	104616744	104616914	85	G	A
chr7	114850994	114851164	85	C	A
chr7	155926943	155927113	85	C	T
chr7	1907009	1907179	85	G	A
chr7	2041432	2041602	85	A	G
chr7	2046830	2047000	85	C	T
chr7	2048220	2048390	85	G	C
chr7	2048335	2048505	85	A	G
chr7	2155250	2155420	85	C	G
chr7	2156338	2156508	85	A	T
chr7	2570045	2570215	85	C	A
chr7	44613611	44613781	87	CG	C
chr7	44620836	44621006	85	C	A
chr7	70175095	70175265	85	C	G
chr7	81454202	81454372	85	G	A
chr7	86861201	86861371	85	C	G
chr7	99537733	99537903	85	G	A
chr8	10268805	10268975	85	G	A
chr8	103567267	103567437	85	G	T
chr8	107391990	107392160	85	C	T
chr8	125756504	125756674	93	A	ATG
chr8	135923269	135923439	85	G	A
chr8	18669340	18669510	85	C	T
chr8	34459546	34459716	85	G	A
chr8	38240008	38240178	85	A	G
chr8	51763446	51763616	85	C	T
chr8	53440040	53440210	85	T	G
chr8	78457481	78457651	85	T	A
chr9	120421608	120421778	85	C	G
chr9	123179525	123179695	85	G	A
chr9	130052614	130052784	85	G	A
chr9	131591832	131592002	85	G	A
chr9	16287793	16287963	85	T	C
chr9	77342863	77343033	85	A	T
chr9	92116545	92116715	85	T	C
chrX	153626649	153626819	85	A	C
chrX	5071192	5071362	85	T	C
chrX	68377126	68377296	85	T	G