CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Exploring the camouflaged and non-coding genome: analysing difficult genomic regions in human disease genetics
Authors	Becerra Rodríguez , Maria de los Ángeles
Publication date	2023
Original Citation	Becerra Rodríguez , M. A. 2023. Exploring the camouflaged and non-coding genome: analysing difficult genomic regions in human disease genetics. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2023, Maria de los Ángeles Becerra Rodríguez. - https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2026-11-15 02:59:24
Item downloaded from	https://hdl.handle.net/10468/14996



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Ollscoil na hÉireann, Corcaigh

National University of Ireland, Cork



Exploring the camouflaged and non-coding genome: analysing difficult genomic regions in human disease genetics

Thesis presented by

María de los Ángeles Becerra Rodríguez

ORCiD: 0000-0001-5818-1308

for the degree of

Doctor of Philosophy

University College Cork

School of Biochemistry and Cell Biology

Head of School/Department: Prof Justin McCarthy

Supervisors: Prof Tom Moore, Prof Pavel Baranov

2023

Acknowledgements

I would like to thank my supervisor Prof. Tom Moore for his guidance, support, and valuable input throughout this research journey. I also wish to express my gratitude to my co-supervisor Prof. Pasha Baranov for his constructive feedback and helpful suggestions. I extend my thanks to Dr. Elliott Rees from Cardiff University. His expertise and advice have played an important role in shaping the direction of this work. I am also grateful to all the participants who generously donated their sample to genetic studies. Their contributions have been essential to the completion of this study.

Thanks to all my family for their love and unwavering support, especially my parents, my sister, and my grandad Miguel. Thanks to all my friends from my three homes (Córdoba, Granada, and Cork) for making it fun. To my colleagues, for the great atmosphere in the lab. To Ana, for making these years feel like I was I bit less away from family. Lastly, to Dáire, for bringing so much love and joy to my life, you make it all worth it.

Gracias mamá por ser mi gran apoyo y a Carmen por entenderme a tantos niveles. Gracias papá por enseñarme tantas cosas y a mi abuelo Miguel, por la historia de tu vida. Esta tesis doctoral va dedicada a vosotros.

Contents

Acknowledgements 2
List of Figures 7
List of Tables 9
Abbreviations 11
Declaration 13
Abstract 14
Chapter 1: General introduction 16
1.1 The dark and camouflaged genome 16
1.1.1 Segmental duplications 16
1.2 The non-coding genome 20
1.2.1 Introns 21
1.2.2 Cis-Candidate Regulatory Regions 22
1.2.3 Non-coding RNA 23
1.2.3.1 Long non-coding RNAs 24
1.3 Genetic architecture of psychiatric disorders 25
1.4 Psychotic disorders are human-specific 27
1.5 Schizophrenia 28
1.5.1 Environmental risk factors 31
1.5.2 Genetic risk factors 32
1.4.2.1 Common variation 32
1.4.2.2 Rare variation 34
1.4.2.2.1 Copy number variants 36
1.5.3 The complex genetic aetiology of schizophrenia 36
1.6 Missing heritability 37
Chapter 2: Contribution of copy number variants in segmental duplications to schizophrenia genetic liability 40
2.1 Summary 40
2.2 Introduction 41
2.2.1 Classifications of CNVs 41
2.2.1.1 Recurrent and non-recurrent CNVs 41
2.2.1.2 Common and rare CNVs 42
2.2.1.3 Genic and non-genic CNVs 43

2.2.2 Detection methods for CNVs 44
2.2.2.1 Microarray 44
2.2.2.2 Next-generation sequencing (NGS) 45
2.2.2.3 Polymerase chain reaction (PCR) 45
2.2.3.4 Karyotype and Fluorescence in situ hybridization (FISH) 46
2.2.3 CNVs and their contribution to schizophrenia genetic risk 47
2.2.3.1 Deletion at 22q11.2 (DiGeorge syndrome) 47
2.2.3.2 Duplication at 16p11.2 48
2.2.3.3 Deletions and duplications at 1q21.1 49
2.2.3.4 Deletions in *NRXN1* (2p16.3) 49
2.2.3.5 Deletions at 3q29 50
2.2.3.6 Duplications at 7q11.23 (Williams-Beuren syndrome) 50
2.2.3.7 Deletions and duplications at 15q11-q13 51
2.2.3.8 Other deletions and duplications at 16p 51
2.2.3.9 The discovery gap 52
2.3 Results 56
2.3.1 Gene-wise analysis 56
2.3.2 Genome-wide analysis of CNV burden 62
2.3.2.1 Gene-set burden analyses 63
2.3.3 Validation of the *DRD5* duplication 67
2.4 Discussion 74
2.5 Methods 78
2.5.1 Samples 78
2.5.2 Exclusion of low-quality probes 78
2.5.3 CNV detection 81
2.5.3.1 CNV quality control 82
2.5.3.2 Sample quality control 82
2.5.4 Gene-wise analysis 86
2.5.5 CNV burden analyses 87
2.5.5.1 Genomic element enrichment 88
2.5.6 Validation 89

Chapter 3: Contribution of constrained non-coding regions in copy number variants to schizophrenia genetic risk 91
3.1 Summary 91
3.2 Introduction 92
3.2.1 Coding and non-coding constraint 92
3.2.2 Study design 95
3.3 Results 99
3.3.1 CNVs in individuals with schizophrenia exhibit higher non-coding constraint than CNVs in controls 99
3.3.2 Sensitivity analyses 103
3.3.2.1 CNVs previously implicated in schizophrenia 103
3.3.2.2 LoF intolerant genes (pLI < 0.9) 103
3.3.3 Constraint in genic and non-genic CNVs 106
3.3.3.1 Coding Z score constraint effect in genic CNVs 107
3.3.3 Other Z score measures 110
3.4 Discussion 111
3.5 Methods 116
3.5.1 CNV calls 116
3.5.2 Annotation of CNV Z score constraint based on genomic features 116
3.5.3 Logistic regression and meta-analysis 117
3.5.3.1 Sensitivity analyses 117
Chapter 4: *In silico* characterization of *PSG8-AS1* long non-coding RNA in the segmentally duplicated *PSG* locus and its involvement in human brain development and disease 120
4.1 Summary 120
4.2 Introduction 121
4.2.1 The pregnancy-specific glycoprotein (*PSG*) locus 121
4.2.2 The non-coding content of the *PSG* locus 122
4.2.3 The human brain is uniquely complex 124
4.2.4 Specific cell types in the brain 126
4.2.5 Brain disease 128
4.3 Results 132
4.3.1 *PSG8-AS1* is an oligodendrocyte-specific lncRNA 132
4.3.2 *PSG8-AS1* emerged in the Catarrhini lineage 138

4.3.3 *PSG8-AS1* may be involved in myelination 144
4.3.4 *PSG8-AS1* as a potential biomarker in glioma 151
4.4 Discussion 156
4.5 Methods 161
4.5.1 Gene expression datasets 161
4.5.2 Transcription factor binding and expression quantitative trait loci (eQTLs) 164
4.5.3 Conservation 164
4.5.4 Weighted gene co-expression network analysis 165
4.5.5 Prediction of binding partners 166
4.5.6 Functional annotation 166
4.5.7 Glioma diagnostic and survival analyses 166
4.5.8 Statistical analyses 167
Chapter 5: General discussion 168
Appendix: Supplementary material 176
Supplementary Figures 176
Chapter 3 176
Chapter 4 177
Supplementary Tables 178
Chapter 2 178
Chapter 3 201
Chapter 4 224
References 255

List of Figures

Figure 1. Visual abstract of chapter 2. 55
Figure 2. Manhattan plot for the gene-wise analysis of duplications 60
Figure 3. Manhattan plot for the gene-wise analysis of deletions. 61
Figure 4. CNV tracks in the *DRD5* duplication locus in 4p16.1 (hg19 genome version). 69
Figure 5. LRR and BAF traces from raw files from MGS study (hg19 genome version). 70
Figure 6. LRR and BAF traces from raw files from ClozUK study (hg19 genome version). 73
Figure 7. Samples quality control histograms for the MGS study. 84
Figure 8. Samples quality control histograms for the ClozUK study. 85
Figure 9. Analysis strategy: example of the assignment of constraint to CNVs conditioning on the genomic annotation. 98
Figure 10. Primary findings. Z score constraint is increased in CNVs in schizophrenia cases compared to CNVs in controls. 102
Figure 11. Sensitivity analyses of non-coding constraint. 105
Figure 12. Comparison of the contribution of non-coding constraint in genic and non-genic CNVs. 109
Figure 13. Expression of PSG8 is low across tissues. 134
Figure 14. Different tissue expression of the *PSG* genes and *PSG8-AS1*. 135
Figure 15. Comparison of single cell-specific expression of *PSG8-AS1* with brain cell marker genes. 136
Figure 16. Localization of cCREs, different sets of transcription factor binding proteins, and eQTLs that potentially regulate *PSG8-AS1*. 137
Figure 17. Best alignments to the human genome (represented by “nets”) in the UCSC browser Mammals Multiz Alignment & Conservation (27 primates) to *PSG8-AS1*. 141
Figure 18. Multi-Alignment of the RNA sequences of the *PSG8-AS1* orthologs 142
Figure 19. Barplots showing *PSG8-AS1* and ortholog expression levels in different primate species and tissues. 143

Figure 20. Expression of *PSG8-AS1* and several brain cell markers across brain developmental stages. 147
Figure 21. Barplot showing False Discovery Rate (FDR) of enriched pathways in the gene set enrichment analysis of the *PSG8-AS1* high-correlated gene expression module..... 148
Figure 22. Alignment coverage density across *PSG8-AS1* in the SK-N-SH neuroblastoma cell line. 149
Figure 23. Prediction of protein interaction with *PSG8-AS1* results. 150
Figure 24. Expression of *PSG8-AS1* in tumours compared to paired non-tumoral tissue. 154
Figure 25. Kaplan-Meier survival curves in patients from the LGG TCGA study stratified by *PSG8-AS1* expression and molecular subtype. 155
Figure 26. Overview of thesis structure. 172

--

Supplementary Figure 1. Length distribution comparison between genic and non-genic CNVs. 176
Supplementary Figure 2. Validation of *PSG8-AS1* expression specificity in oligodendrocytes 177

List of Tables

Table 1. List of CNV loci previously implicated in schizophrenia. 54
Table 2. CNV Burden by number of disrupted great ape specific genes preferentially expressed in progenitors of fetal neocortex. 65
Table 3. CNV Burden by the number of disrupted "human ancestor quickly evolved regions" (HAQERs). 66
Table 4. Number of samples and SNP array probes on each ClozUK dataset. 80
Table 5. Genomic annotations used in non-coding constraint study. 90
Table 6. Hierarchy for the different categories of constraint Z scores windows tested. 119
Table 7. LncRNAs dysregulated in brain disease 129
Table 8. Putative *PSG8-AS1* orthologs obtained from blastn. 142
Table 9. Gene expression datasets accessed through GEO. 163

--

Supplementary Table 1. MGS dataset. Results of the gene-wise analysis of duplications. 178
Supplementary Table 2. ClozUK dataset. Results of the gene-wise analysis of duplications. 182
Supplementary Table 3. Combined dataset. Results of the gene-wise analysis of duplications. 185
Supplementary Table 4. MGS dataset. Results of the gene-wise analysis of deletions. 190
Supplementary Table 5. ClozUK dataset. Results of the gene-wise analysis of deletions. 193
Supplementary Table 6. Combined dataset. Results for the gene-wise analysis of deletions. 194
Supplementary Table 7. Burden by number of CNVs. 198
Supplementary Table 8. Burden by length (Kb) of CNVs. 199
Supplementary Table 9. Burden by number of genes affected by the CNV. 200
Supplementary Table 10. Primary findings of Z score constraint. 201

Supplementary Table 11. Sensitivity analysis results removing previously implicated loci.....205
Supplementary Table 12. Sensitivity analysis results adding the number of LoF intolerant genes affected. 208
Supplementary Table 13. Genic CNVs results. 212
Supplementary Table 14. Non-genic CNVs results. 215
Supplementary Table 15. Genic CNVs results, correcting for the maximum Z score from coding exons.....218
Supplementary Table 16. Sum Z score. 221
Supplementary Table 17. Mean Z score. 222
Supplementary Table 18. Median Z score. 223
Supplementary Table 19. Transcription factor binding to cCREs in proximity to PSG8-AS1 genomic location. 224
Supplementary Table 20. Cis-eQTLs for *PSG8-AS1* in the brain. 228
Supplementary Table 21. Genes in *PSG8-AS1* WGCNA module. 236
Supplementary Table 22. Gene set enrichment analysis performed with DAVID online tool of the *PSG8-AS1* high-correlated gene expression module. 245
Supplementary Table 23. Top 300 *PSG8-AS1* protein binding predictions by catRapiD v2.1.....246
Supplementary Table 24. Gene set enrichment analysis performed with DAVID online tool of the predicted *PSG8-AS1*-binding proteome. 254

Abbreviations

ADHD	Attention defficit and hyperactivity disorder
ANOVA	Analysis of variance
ASD	Autism spectrum disorder
BAF	B alelle frequency
BMI	Body mass index
cCRE	Cis candidate regulatory element
CGH	Comparative genomic hybridization
CNS	Central nervous system
CNV	Copy number variant
CPM	Counts per million
DAVID	Database for Annotation, Visualization and Integrated Discovery
DNA	Deoxyribonucleic acid
DR	Depletion Rank
eQTL	Expression quantitative trait locus
FISH	Fluorescence In Situ Hybridization
FoSTeS	(Replication) Fork stalling and template switching
FPKM	Fragments per kilobase of transcript per million mapped reads
GEO	Gene expression omnibus
GTEx	Genotype-Tissue Expression project
GWAS	Genome-wide association analysis
hiPSC	Human induced pluripotent stem cell
HLA	Human leukocyte antigen
HMM	Hidden Markov model
IDH	Isocitrate dehydrogenase (NADP(+)) 1
IDH-mt	IDH mutated
IDH-wt	IDH wildtype
IPS	Interaction propensity score
LD	Linkage disequilibrium
LGG	Low-grade glioma
lncRNA	Long non-coding ribonucleic acid
LoF	Loss of function
LRR	Log R ratio
MDD	Major depression disorder
MGS	Molecular Genetics of Schizophrenia
miRNA	Micro ribonucleic acid
MMBIR	Microhomology-mediated break-induced replication
MR	Magnetic resonance
mRNA	Messenger ribonucleic acid

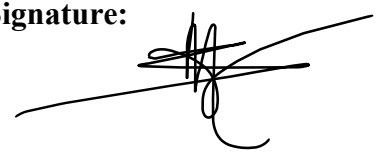
NAHR	Non-allelic homologous recombinantion
ncRNA	Non-coding ribonucleic acid
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
NSC	Neural stem cell
OL	Oligodendrocyte
OPC	Oligodendrocyte precursor cell
OR	Odds ratio
pLI	Probability of loss of function intolerance
pre-mRNA	Pre-messenger ribonucleic acid
PRS	Polygenic risk score
PSG	Pregnancy-specific glycoprotein
PSG8-AS1	Pregnancy-specific glycoprotein 8 antisense 1
REST	RE1-silencing transcription factor
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RPKM	Reads per kilobase per million mapped reads
RPM	Reads per million
rRNA	Ribosomal ribonucleic acid
scRNA-seq	Single cell RNA sequencing
SD	Segmental duplication
snoRNA	Small nucleolar ribonucleic acid
SNP	Single nucleotide polymorphism
snRNA	Small nuclear ribonucleic acid
snRNA-seq	single nucleus RNA sequencing
SNV	Single nucleotide variant
TCGA	The cancer genome atlas
TGF-β	Transforming growth factor β
TMM	Trimmed mean of M values
TOM	Topological overlap matrix
TPM	Transcripts per million
tRNA	Transfer ribonucleic acid
WBS	Williams-Beuren syndrome
WES	Whole exome sequencing
WGCNA	Weighted gene correlation network analysis
WGS	Whole genome sequencing

Declaration

I hereby declare that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism and intellectual property.

Name and date: María de los Ángeles Becerra Rodríguez, June 2023

Signature:



Abstract

The main objective of this thesis is to highlight the importance of investigating camouflaged regions, specifically segmental duplications, and non-coding regions, in the human genome. These regions, often overlooked due to their complexity, hold immense potential for uncovering novel insights into disease genetics. In pursuit of this objective, this thesis first focused on the study of camouflaged and non-coding regions in the context of schizophrenia. Schizophrenia is a complex psychiatric disorder specific to humans characterized by a combination of altered cognitive function, distorted perception, and disrupted social behaviour. Understanding the genetic underpinnings of schizophrenia is crucial for advancing our knowledge of its aetiology and developing more effective diagnostic and therapeutic approaches. Through comprehensive genomic analyses, novel insights were gained, which identified a novel duplication in a locus affecting a dopamine receptor implicated in neurotransmission. Additionally, small deletions in constrained non-coding regulatory regions were implicated in schizophrenia for the first time. Moreover, this thesis characterizes a long non-coding RNA (lncRNA) originating from the segmentally duplicated *Pregnancy-Specific Glycoprotein* locus. The lncRNA was expressed exclusively in oligodendrocytes, implicating it in the regulation of myelination processes in the brain. This lncRNA is human-specific, further emphasizing the biological relevance of camouflaged and non-coding regions in the context of human evolution. Throughout this thesis, a human-specific perspective was adopted, recognizing the unique genomic features that shape our species. By expanding our knowledge of difficult genomic regions, such as camouflaged and non-coding regions, this thesis aims to close the gap

in the missing heritability problem and to gain a comprehensive understanding of the genetic architecture underlying complex disorders and uniquely human traits like schizophrenia.

Chapter 1: General introduction

1.1 The dark and camouflaged genome

Over the past two decades, researchers have grappled with the daunting complexity of the human genome. One major challenge arises from the presence of "dark" regions within the genome, which cannot be adequately assembled or aligned using standard high-throughput short-read sequencing technologies. These dark regions, comprising both "dark by depth" (with few or no mappable reads) and "dark by mapping quality" (reads aligned with low mapping quality), pose obstacles to identifying mutations relevant to human health and disease. Dark by depth regions are inherently difficult to sequence due to their chemical characteristics, such as high GC content, rendering them devoid of sequencing reads (Ebbert et al., 2018; Zheng-Bradley et al., 2017). Alternatively, bioinformatic challenges arise from duplicated genomic regions, giving rise to "dark by mapping quality" or "camouflaged" regions. Camouflaged regions encompass segmental duplications (SDs) and tandem DNA repeats (i.e., centromeres, telomeres, and other short tandem repeats) (Ebbert et al., 2019). Short tandem repeats consist of repeat units of 2-20 bp and can be either interspersed (usually from a transposable element origin) or repeated locally and their copy number variation has been linked to disease, such as Parkinson's disease and autism spectrum disorder (ASD) (Kõks et al., 2022; Thakur et al., 2021; Trost et al., 2020).

1.1.1 Segmental duplications

Segmental duplications, also termed low copy repeats (LCRs), are long (often more than 100 Kb) genomic regions with highly identical sequences [>90% and >1 Kb]

(Bailey et al., 2001). These duplications arise from errors or variations in DNA recombination or transposition mechanisms and contribute to gene duplication, genomic instability, and the evolution of new traits in primates (Bailey and Eichler, 2006; Cantsilieris et al., 2020; Florio et al., 2018; Samonte and Eichler, 2002; Stankiewicz et al., 2004). Segmental duplications constitute 7% of the 3.2 billion bp human genome and promote copy number variation implicated in disease (Table 1) (Bailey and Eichler, 2006; ; Marshall et al., 2017; Rees et al., 2014b; Vollger et al., 2022).

The formation of segmental duplications involves the unequal crossing-over during nonallelic homologous recombination. Homologous recombination occurs between sister chromatids as a system to correct replication errors such as double strand breaks or stalled replication forks. It also contributes to chromosome shuffling during meiosis. When sequences are not correctly aligned (i.e. alignment happens between sequences that are not in allelic positions), loci are unequally crossed over resulting duplication of the intervening DNA segment in one strand and deletion in the other (Liu et al., 2012; Malhotra and Sebat, 2012; Parks et al., 2015). Expansion of gene families with homologous functions occurs through this process (Dornburg et al., 2022). However, over time, independent evolution of gene paralogs or DNA segments with no previous function accumulate enough mutations to give rise to completely new functions or altered regulation (Florio et al., 2018; Mangan et al., 2022; Stankiewicz et al., 2004).

Overall, segmental duplications contribute to rapid gene evolution and innovation, which can be an important driver of species diversity and adaptation. Interestingly, segmental duplications are enriched in great apes compared to other primates,

particularly in the genomes of humans, chimpanzees and gorillas (Bailey and Eichler, 2006; Florio et al., 2018; Samonte and Eichler, 2002; Stankiewicz et al., 2004). They are especially abundant in regions near centromeres and telomeres and are thought to play a role in the evolution of primate-specific traits such as increased brain size and improved cognition (Erady et al., 2022; Florio et al., 2018; Vollger et al., 2022).

Segmental duplications are thus fixed in the genome of different species, mostly as gene families (Dornburg et al., 2022). Their rapid evolution underpins the production of novel transcripts important for human brain development (Florio et al., 2018; Vollger et al., 2022). Up to 15 coding genes have been described to be human-specific, with no orthologs in other primate species, that are preferentially expressed in the progenitors of developing neocortex, a brain region important for higher cognitive functions like language and memory (Berg et al., 2021; DeFelipe et al., 2002; Florio et al., 2015; Petersen, 2007). All 15 genes are located to segmental duplications (Florio et al., 2018). In addition, non-coding "human ancestor quickly evolved regions" (HAQERs) that shaped hominin-specific enhancers active in the developing cerebral cortex are enriched in recent segmental duplications (Mangan et al., 2022).

Human brain-specific genes that emerged as a result of segmental duplications may not be related to the genes that comprise the segmentally duplicated gene family. For example, the segmentally duplicated *GOLGA* gene family on chromosome 15 contains one copy of the human-specific gene *ARHGAP11B* (Florio et al., 2016). *ARHGAP11B* encodes Rho GTPase-activating protein 11B, which regulates the Rho family of small GTPases. *ARHGAP11B* is highly expressed in the developing neocortex and is thought to be involved in the expansion of the neocortex during human embryonic development,

as well as the regulation of neural circuit connectivity (Fischer et al., 2022; Florio et al., 2015; Heide et al., 2020; Schmidt and Polleux, 2022).

In contrast, the segmentally duplicated *GOLGA* gene family encodes for the golgin subfamily of coiled-coil proteins associated with the Golgi apparatus (Jiang et al., 2007; Pujana et al., 2001; Zody et al., 2006). These genes are expressed in all tissues and may play roles in membrane traffic and Golgi structure, but their precise molecular function is unknown. The *GOLGA* repeats have specifically expanded in primates, resulting in 40 copies spread across human chromosome 15 (Sudmant et al., 2013).

Segmental duplications subsequently promote additional meiotic unequal crossover events, which contribute to common and recurrent copy number variants (CNVs) that are associated with psychiatric disorders such as autism and schizophrenia (Table 1) (Baba et al., 2019; Cooper et al., 2011; Kirov et al., 2009; Marshall et al., 2017; Mulle et al., 2014, 2010; Quintero-Rivera et al., 2010; Rees et al., 2014c, 2014b; Sanders et al., 2011; Yoon and Mao, 2021), and other complex genetic diseases and traits (Ebbert et al., 2019; Hujoel et al., 2022). The *GOLGA* locus has been linked to evolutionary rearrangements in primates as well as disease susceptibility, such as Prader-Willi syndrome, 15q13 microdeletions, and 15q24 microdeletions (Amos-Landgraf et al., 1999; Antonacci et al., 2014; Burgess et al., 2014; Cooper et al., 2011; Doelken et al., 2013; El-Hattab et al., 2009; Maggiolini et al., 2019; Pujana et al., 2002; Wat et al., 2010).

Current techniques, however, have fallen short in analysing the genetic and transcriptomic diversity of human segmental duplications (Prodanov and Bansal, 2020; Zhao et al., 2021). Segmental duplications constitute a component of the “dark genome”

termed the “camouflaged genome”, whereby short reads or array probes cannot be confidently aligned to a unique genomic location (Ebbert et al., 2019). The fact that many genomic elements (genes and enhancers) essential for human brain development are contained within segmental duplications implies that many variants possibly implicated in the genetic liability to psychiatric disease could be awaiting discovery (Ebbert et al., 2019; Florio et al., 2018; Mangan et al., 2022). This is extremely important in rare CNV studies, where variants covered by segmental duplications in more than 50% of their length have been systematically excluded (Marshall et al., 2017).

1.2 The non-coding genome

Non-coding DNA refers to the part of the genome that does not code for proteins. Some non-coding DNA sequences act as regulatory elements interacting with transcription factors or are transcribed to RNA to control the expression of coding genes in many different ways. Accounting for 98% of the human genome, non-coding DNA contributes the majority of the signal in GWAS studies (Barešić et al., 2020; Bartonicek et al., 2017; Trubetskoy et al., 2022). However, the contribution of non-coding variants to disease biology is more challenging to understand than that of coding variants. For example, loss of function (LoF) variants in coding genes, where the resulting protein is prematurely truncated or becomes dysfunctional due to missense mutations, will increase the chance of pathogenicity (Lek et al., 2016). While similar variants do not exist in the non-coding genome, the alteration of a non-coding regulatory region could affect the expression of a nearby gene (cis-acting) or a distant gene (trans-acting),

resulting in equivalent downstream effects to that of a loss-of-function (LoF) variant (Moore et al., 2020; Q. Wang et al., 2019; Wittkopp, 2005).

Modern evolutionary genetic studies indicate that complex organisms contain extensive non-coding regions in their genomes (Jo and Choi, 2015; Mattick, 2001; Mattick et al., 2023). The generation of novel coding genes might have been superseded by enhanced sophistication of the regulatory machinery of existing coding genes, underpinning increased biological complexity of the species (Jo and Choi, 2015; Mattick et al., 2023). Consistent with this, measures of constraint against variation in a population, instead of phylogenetic conservation scores, have emerged as a way to estimate the biological importance of a DNA sequence in a specific species (Chen and The gnomAD consortium, 2022; di Iulio et al., 2018; Gussow et al., 2017; Halldorsson et al., 2022; Karczewski et al., 2020; Lek et al., 2016; Pollard et al., 2010; Vitsios et al., 2021). At the same time, in recent years, tremendous progress has been made in the field of the "non-coding and regulatory genome", and many different functional modes have been described.

1.2.1 Introns

All eukaryotic genomes carry genes structured into exons and introns. When a coding gene is transcribed, exonic regions are joined together to form the mature messenger RNA (mRNA) that will be translated into protein, while introns are removed by a molecular complex called the spliceosome (Jo and Choi, 2015). Introns are longer than exons (25% vs. 2% of the human genome) and therefore should convey a huge energetic disadvantage. To overcome this, introns have provided an adaptative advantage to eukaryotes, displaying many different functional roles in gene regulation, such as the

control of alternative splicing, nonsense-mediated decay and chromatin organization (Gehring and Roignant, 2021; Jo and Choi, 2015; Peccarelli and Kebaara, 2014; Schwartz et al., 2009).

1.2.2 Cis-Candidate Regulatory Regions

Other functional non-coding regions include DNA sequences that recruit transcription factors and structural proteins to allow transcription of other DNA sequences. Regions of open chromatin, where histone modification of active enhancer elements (H3K27Ac and H3K4Me1) and/or transcription factor binding can be detected are termed cis-Candidate Regulatory Regions (cCREs). Different types of cCREs are summarized by the ENCODE Candidate Regulatory Elements resource (promoters, enhancers and CTCF-bound) (Moore et al., 2020).

Promoters are DNA sequences extending 250 bp immediately up- and downstream of a gene Transcription Start Site (TSS) where relevant proteins (such as RNA polymerase and transcription factors) bind to initiate transcription. In contrast, enhancers mediate temporal and tissue-specific regulation of transcription via long-distance (dELS) or proximal (pELS) interactions with promoter regions (Moore et al., 2020).

Functional alteration, expression quantitative trait loci (eQTL) density, and chromatin interactions have been used to link cCREs to specific genes (Akbarian et al., 2015; Lonsdale et al., 2013; Moore et al., 2020; Schoenfelder and Fraser, 2019; Schöpflin et al., 2022). However, some cCREs do not have a gene-specific effect and, instead, ensure optimal chromatin architecture for gene regulation. Regulatory regions interact with each other inside topologically associating domains (TAD), which constitute the fundamental units of three-dimensional nuclear organization. Binding sites for the zinc

finger protein CCCTC-binding factor (CTCF), or CTCF-bound elements, reside predominantly at TAD boundaries and help in the restriction of interactions between enhancers and promoters (Nanni et al., 2020).

1.2.3 Non-coding RNA

Non-coding RNA (ncRNA) genes produce functional RNA that is not translated into proteins. Some are also structured into exons and introns in the genome (Carninci et al., 2005; Kapranov et al., 2007; Kellis et al., 2014; Mattick, 2001; Mattick et al., 2023). There are many different types of ncRNAs, which can be divided into two categories: housekeeping ncRNAs and regulatory RNAs (Mattick, 2001). In eukaryotes, housekeeping RNAs are widely conserved and required for basic cell processes (Necsulea and Kaessmann, 2014). Ribosomal RNAs (rRNAs) and transfer RNA (tRNA) act during protein synthesis by serving as scaffolds of ribosomes and transferring specific amino acids to the nascent peptide, respectively (Giegé, 2008; Sloan et al., 2017). Other housekeeping RNAs, such as small nucleolar RNAs (snoRNAs), help in the modification of rRNAs and tRNAs, while small nuclear RNAs (snRNAs) assist in the process of intron-splicing from pre-messenger RNAs (pre-mRNAs) (Huang et al., 2022; Karijolich and Yu, 2010). Regulatory RNAs, on the other hand, have varying levels of conservation, with many being species-specific (Necsulea and Kaessmann, 2014). MicroRNAs (miRNAs) target complementary sequences in mature messenger RNAs (mRNAs) to prevent their translation into protein (Bartel, 2018; Jonas and Izaurralde, 2015). Approximately one hundred miRNAs are highly conserved in sequence and expression levels across mammalian species, but others are lineage-specific (Necsulea and Kaessmann, 2014). Other examples of regulatory RNAs are Piwi-Interacting RNAs (piRNAs), which modulate DNA transposable element

activity, and circular RNAs (circRNAs), whose functions are still unclear but are thought to interact with proteins and other types of RNA to regulate their biogenesis and activity (Ozata et al., 2019; Xu and Zhang, 2021). These last two RNA elements are believed to have low conservation across species (Necsulea and Kaessmann, 2014).

1.2.3.1 Long non-coding RNAs

Long non-coding RNAs (lncRNAs) are linear RNAs longer than 200 nucleotides that regulate gene expression by many different mechanisms. LncRNAs bind proteins, RNA and DNA and are involved in every step of the DNA to protein process (Yao et al., 2019). The vast majority of lncRNAs are cell and developmental-stage specific, and they exhibit very rapid turnover across species (Derrien et al., 2012; Necsulea and Kaessmann, 2014). Whereas some genes annotated as pseudogenes may act in the same way as lncRNAs (N. Wang et al., 2019), several studies have documented evidence that many transcripts thought to be lncRNAs may, indeed, encode small functional peptides (Anderson et al., 2015; Nelson et al., 2016). For example, the small peptide DWORF, previously annotated as lncRNA LOC100507537, has been shown to perform an essential role in myocyte contractility (Nelson et al., 2016).

Depending on their subcellular localization, lncRNAs can directly participate in a variety of processes. For example, the lncRNA *XIST* silences the transcription of an entire X chromosome in the nucleus for gene dosage compensation in females (Hall and Lawrence, 2010). Many other lncRNAs contribute to chromatin remodelling and accessibility or interact with transcription factors and RNA polymerases to modulate the generation of new transcripts (Boque-Sastre et al., 2015; Calo et al., 2015; Chu et al., 2011; Jain et al., 2016; Rinn et al., 2007). The activity of lncRNAs in the nucleus

also influences RNA splicing and nuclear retention (Sasaki et al., 2009; Valgardsdottir et al., 2008; Yamazaki et al., 2018). In the cytoplasm, lncRNAs are involved in the regulation of translation and post-translational modifications of proteins, modulating mRNA stability and recruiting protein modifying enzymes (Gong and Maquat, 2011; Ingolia et al., 2014; Poliseno et al., 2010; Wang et al., 2014; Zamore et al., 1997).

The investigation of camouflaged and non-coding regions represents a significant challenge in the field of disease genetics. These regions, due to their inherent complexities, have been systematically underexplored, limiting our understanding of their contributions to complex disorders, such as psychiatric diseases. However, given the pressing need to unravel the intricate genetic architecture underlying these disorders, it is imperative that we shift our attention towards these difficult regions.

1.3 Genetic architecture of psychiatric disorders

Psychiatric disorders encompass a diverse range of mental conditions characterized by atypical patterns of cognition, emotion, and behaviour (Eaton et al., 2008). They signify an enormous societal and individual burden since, second only to musculoskeletal disorders (arthritis, low back pains, gout etc.), psychiatric disorders are one of the main contributors to years lived with disability worldwide (Eaton et al., 2008; GBD 2019 Diseases and Injuries Collaborators, 2020). In the context of genetics, they are categorized as “complex disease”, meaning that, as opposed to Mendelian disease, thousands of genomic variants in conjunction with environmental factors lead to the development of psychiatric traits (Jostins and Barrett, 2011; P. H. Lee et al., 2019; Romero et al., 2022). The pathophysiology of psychiatric disorders remains intractable in most cases. Approaches that have been successful for other areas of complex disease,

such diabetes type II or Crohn's disease, have failed to define the molecular disruptions and structural pathology that predispose individuals to psychiatric disease (Gettler et al., 2021; Hahn et al., 2022; Smoller et al., 2019; Sullivan et al., 2012; Voight et al., 2010).

The substantial impact of genetic variation on the susceptibility to a wide spectrum of psychiatric disorders has been firmly established by family/twin studies and large-scale genomic studies. Autism spectrum disorder (ASD), schizophrenia, bipolar disorder, and attention-deficit hyperactivity disorder (ADHD) reach estimates of around 80% heritability, followed by anxiety disorder, anorexia nervosa (less than 60%), and major depressive disorder (MDD, less than 40%) (Smoller et al., 2019). Through common variant-level analysis of major psychiatric disorders comprising more than 725,000 cases and controls, noteworthy evidence of pleiotropy has emerged, exemplified by the fact that 75% of the genome-wide significant and independent loci associated with one disorder are implicated in two or more other disorders (P. H. Lee et al., 2019). Schizophrenia and bipolar disorder are the most genetically correlated disorders (r_g = 0.68). Through genetic correlations including all genotyped single nucleotide polymorphisms (SNPs), four clusters have been suggested: disorders with psychotic experiences (schizophrenia and bipolar disorder), disorders with compulsive behaviors (anorexia nervosa, Tourette syndrome and obsessive-compulsive disorder), disorders of early neurodevelopmental onset (ASD and ADHD) and mood and internalizing disorders (MDD and anxiety disorder) (Grotzinger et al., 2022; Romero et al., 2022). All these psychiatric disorders have been delineated with specific and different diagnostic criteria; however, it is a widespread accepted fact that an individual receiving treatment for a mental disorder is at increased susceptibility of experiencing another

psychiatric trait (AL-Asadi et al., 2015; Plana-Ripoll et al., 2019). Additionally, rare variants with greater effect sizes also contribute to the genetic susceptibility to psychiatric disorders, particularly for disorders with significant fitness consequences such as ASD and schizophrenia (Rees et al., 2021).

1.4 Psychotic disorders are human-specific

Psychotic experiences are characterized by the loss of contact with reality. They originated with the evolution of human species. In comparison with some components of other psychiatric disorders, including hyperactivity, limited social interaction, or anxiety, which can be modeled in other species, psychotic symptoms have only been described in humans (Levchenko et al., 2023; Scheepers et al., 2018). Interestingly, common genetic variants associated with psychotic disorders (schizophrenia and bipolar disorder) can predict creativity (Power et al., 2015). Although a precise definition of the concept of creativity is difficult to determine, similar trends have been observed in numerous epidemiological studies (Herbert, 1959; Jamison, 1989; Juda, 1949; Karlsson, 1970; Ludwig, 1992; Nancy C. Andreasen, 1987; Post, 1994). The capacity of artistic creations and the communication of abstract ideas became gradually more prominent in our Hominin relatives (Aubert et al., 2019; Dl et al., 2018; Gonen-Yaacovi et al., 2013; Hennessey and Amabile, 2010; Jaubert et al., 2016; Joordens et al., 2015). Thus, psychotic symptoms have been interpreted as reflecting the abilities of imagination and abstract thinking, which conferred significant evolutionary advantages to the human species, but are occasionally prone to deregulation (Scheepers et al., 2018).

Genetic variants across the frequency spectrum implicated in schizophrenia exhibit a significant overlap with human accelerated regions (Bhattacharyya et al., 2022; Caseras et al., 2021; Erady et al., 2022). These genomic regions represent sequences that were conserved across all other primate, mammalian, or vertebrate species, but evolved rapidly in human species, and may be responsible for traits exclusive to humans. Of note, human accelerated regions are observed predominantly in the non-coding genome (Whalen and Pollard, 2022). Due to the meaningful overlap between the human-specific nature of schizophrenia, the expansion of segmental duplications (camouflaged regions), and accelerated evolution of non-coding regions, the research themes of this thesis will largely focus on these concepts.

1.5 Schizophrenia

Schizophrenia is a common and severe psychiatric disorder that affects how a person thinks, feels, and behaves. With a lifetime risk of approximately 0.7%, it is correlated with substantial overall health deterioration (linked to smoking, substance abuse and obesity) and social impairment, increased risk of suicide and reduced life expectancy of approximately 15 years (Charlson et al., 2018; Hjorthøj et al., 2017; McGrath et al., 2008; Pompili et al., 2007; Wiersma et al., 2000). It is estimated that around 1 in every 3 people with schizophrenia attempts suicide at some point in their lifetime (Pompili et al., 2007). The prevalence of schizophrenia has risen from 13.1 million cases in 1990 to 23.1 million cases in 2019, with the greatest increases seen in Eastern Sub-Saharan Africa and North Africa/Middle East, owing to rapid population expansion in those regions (Charlson et al., 2018; GBD 2019 Diseases and Injuries Collaborators, 2020).

Schizophrenia usually develops in the late teenage years or early adulthood and affects men and women almost equally (male:female ratio of 1.4), with an incidence of 16.7 per 100,000 individuals in 2019 (Charlson et al., 2018; GBD 2019 Diseases and Injuries Collaborators, 2020; McGrath et al., 2008). People with schizophrenia experience a range of symptoms, categorized as positive symptoms (delusions, hallucinations, disordered thinking, and abnormal behavior) and negative symptoms (social isolation and apathy). These symptoms can be very distressing and disruptive and can make it difficult for people with schizophrenia to function in their daily lives, leading to cognitive impairment in many cases (Dollfus and Lyne, 2017; Tandon et al., 2013).

Delusions are false beliefs that are not based in reality, such as being convinced of being spied on by relatives or other people or believing that someone possesses special powers or abilities. Hallucinations are false perceptions that can affect any of the senses, such as hearing voices or seeing things that are not there. Disordered thinking hinders the ability of people with schizophrenia to organize their thoughts or communicate effectively, while abnormal behavior can manifest as agitation, catatonia, or other unusual movements or actions. According to the latest edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) the presence of at least two of the core symptoms (at least one of these should be delusions, hallucinations, or disorganized speech) over a period of six months, with at least one month of active symptoms, is required for the diagnosis of schizophrenia (Tandon et al., 2013).

The neurochemical imbalances that underpin schizophrenia's clinical features are not fully understood. Previous studies have suggested that presynaptic dopamine constitutes the common pathway that is dysregulated as a result of environmental and

genetic factors (Howes and Kapur, 2009). More recent findings integrate the alteration of synaptic plasticity and the disruption of inhibitory modulation of synaptic signaling through GABAergic neurotransmission (Pocklington et al., 2015). In addition, the classical complement cascade, which harbours the strongest genetic association with schizophrenia through common variants, was implicated with the excessive 'pruning' of synapses observed in brains of affected individuals (Ripke et al., 2014; Sekar et al., 2016). The abnormal processes specific to schizophrenia are believed to initiate at early stages of brain development, years before the age of illness onset (Grotzinger et al., 2022; Owen et al., 2011; Rapoport et al., 2012). The latest mechanistic hypothesis, which accounts for major risk factors and clinical characteristics, proposes that people at risk of schizophrenia are born with more stress-vulnerable synapses, which may result in the disruption of synaptic activity in pyramidal neurons (negative symptoms) and disinhibition of dopamine pathways (positive symptoms) later in life depending on the exposure to triggering factors (Howes and Onwordi, 2023).

Treatment for schizophrenia typically involves a combination of antipsychotic medications, that almost exclusively address positive symptoms, and psychotherapy (Li et al., 2016; Patel et al., 2014; Sanyal and Van Tol, 1997). The principal pharmacological mechanism of antipsychotic medications is their antagonistic action on D2 dopamine receptors (Besnard et al., 2012; Li et al., 2016; Sanyal and Van Tol, 1997). Despite antipsychotic treatment, around 30% of individuals with schizophrenia continue to show symptoms and suffer significant impairment. The only remaining treatment option for resistant patients is clozapine (Meltzer, 1997). Although it is unclear which characteristics are precisely responsible for its efficacy in treatment-resistant schizophrenia, the pharmacological profile of clozapine comprises relatively

low affinity for D2 receptors and a much higher affinity for D4 dopamine receptors as well as for subtypes of serotonin receptors (Besnard et al., 2012; Li et al., 2016; Sanyal and Van Tol, 1997).

Schizophrenia is a chronic condition, and some people may experience persistent symptoms and require ongoing care. The persistence of symptoms, poor adherence to the treatment plan, lack of awareness and substance abuse are predictors of disease relapse (Higashi et al., 2013; Kelly and Daley, 2013; Schoeler et al., 2017). The main pharmacological interventions currently in use were established in the 1950s and 1970s, indicating a severe lack of progress in therapeutic development (Meltzer, 1997; Patel et al., 2014; Sanyal and Van Tol, 1997). With proper treatment and support, many people with schizophrenia are able to manage their symptoms and lead fulfilling lives. In fact, approximately 1 in 7 schizophrenia cases recover, based on clinical and social functioning indicators (Jääskeläinen et al., 2013). By deepening our understanding of the molecular aetiology, the scientific community could greatly improve the quality of life of those at risk of experiencing persistent symptoms of schizophrenia.

1.5.1 Environmental risk factors

Adverse environmental factors increase the chances of triggering schizophrenia symptoms. Depending on when in life environmental factors have more influence, they can be grouped into early development, proximal, and onset factors (Stilo and Murray, 2019). Pregnancy complications and advanced parental age constitute factors that act early in development (Abel et al., 2010; Cannon et al., 2002; Dalman et al., 2001; Kotlicka-Antczak et al., 2014; Rubio-Abadal et al., 2015). Migration, especially from developing countries, living in urban areas and, in general, social adversity (in

childhood defined as sexual, physical, or psychological abuse, bullying and/or parental death), are included in the group of proximal factors (Agid et al., 1999; Arseneault et al., 2011; Fisher et al., 2013; Gianfrancesco et al., 2019; Morgan et al., 2008, 2007; Stilo et al., 2013; Trotta et al., 2013; Varese et al., 2012). Remarkably close in time with the age of onset, the other primary environmental factors are drug abuse and trauma. Individuals that suffer the first episode of psychosis are likely to have lived alone and have been socially isolated and economically disadvantaged, not only at onset but also 5 years prior (Morgan et al., 2008; Stilo et al., 2013).

1.5.2 Genetic risk factors

The heritability of schizophrenia has been estimated to be around 60-80% in family/twin studies, indicating that genetic factors explain a large proportion of the variation in susceptibility to the disorder (Owen et al., 2016; Owen and Williams, 2021). However, identifying the specific genetic variants that contribute to schizophrenia has proved to be challenging.

1.4.2.1 Common variation

Common genetic variation, which is defined as variation that is present in at least 1-5% of the population, can explain up to 24% of the genetic liability (Trubetskoy et al., 2022). Genome-wide association studies (GWAS) have identified a large number of genetic variants that are associated with increased risk for schizophrenia, revealing its highly polygenic nature (Trubetskoy et al., 2022). These studies directly genotyped thousands of common SNPs with probe arrays. Because recombination of DNA happens in blocks of genomic segments (haplotype blocks), certain SNP alleles in proximity to each other occur together more often than can be accounted for by chance

(Halldorsson et al., 2019). This phenomenon, known as linkage disequilibrium (LD), allows for the imputation from reference samples that have undergone whole-genome sequencing of millions of SNPs not targeted by SNP arrays (Dias et al., 2022). Not only single nucleotide variants, but also common copy number variants (CNVs), can be captured in GWAS because CNVs lead to different allele frequencies (although the probe coverage in regions with common CNV tends to be lower) (Campbell et al., 2011; J. Chen et al., 2016; Craddock et al., 2010; Ebbert et al., 2019). In the last 15 years, GWAS efforts have revolutionized the research on genetic risk to schizophrenia.

In European populations, the first GWAS was performed in 2009 and reported 3 independent genome-wide significant loci using data from less than 50,000 people (Purcell et al., 2009). By 2014, 108 loci had been identified due to a 3-fold increase in sample size compared to the 2009 GWAS (Ripke et al., 2014). The most recent large-scale GWAS, which used data from more than 300,000 individuals, identified associations with schizophrenia at 287 distinct genomic loci that survived the threshold for genome-wide significance (Trubetskoy et al., 2022). After fine-mapping, a process by which GWAS signals are refined to identify variants that are truly causal to the phenotype, 120 genes were found to be likely implicated (Schaid et al., 2018; Trubetskoy et al., 2022). Synaptic function, which is believed to be disrupted in schizophrenia, was found to be the most affected pathway in the prioritized genes (Trubetskoy et al., 2022). Previously, common variation linked to schizophrenia had been shown to be enriched in genes intolerant to loss of function variants (LoF-intolerant genes) (Lek et al., 2016; Pardiñas et al., 2018). The accumulation of variants with small phenotypic effects at common allele frequencies in constrained genes may be possible because the strong selection against highly deleterious variants reduces the

number of available haplotypes (Pardiñas et al., 2018). Some of the genes known to contribute substantially to genetic risk via rare variants, such as *GRIN2A* and *SP4*, were also prioritized candidate genes from fine-mapping of SNPs reaching genome-wide significance in GWAS (Singh et al., 2022; Trubetskoy et al., 2022).

The only other population where large-scale GWAS has been performed is in East Asian ancestry individuals. With only the top 3 loci shared with European ancestry, they reported an additional 16 loci seemingly specific to the East-Asian population (Lam et al., 2019). Other genetic studies in African American, Indian and Latin American populations, albeit not large-scale, have also underlined the differences in the genetic architecture of schizophrenia between populations (Bigdeli et al., 2020; Gulsuner et al., 2020; Periyasamy et al., 2019; Suarez et al., 2006).

Polygenic risk score (PRS) has become a powerful tool for predicting genetic liability. PRS measures the cumulative contributions of common variants associated with a trait. Different p value thresholds (not only genome-wide significance) can be used for calculating PRS, with higher PRS indicating higher genetic liability (Lewis and Vassos, 2020). Alone, PRS can explain up to 7.7% of heritability of schizophrenia (Trubetskoy et al., 2022). While the predictive power of PRS has been applied to other common diseases in the clinic, PRS in schizophrenia are currently not reliable enough for diagnostic purposes (Lewis and Vassos, 2020).

1.4.2.2 Rare variation

Rare variants are commonly defined as those with less than a 1% allele frequency. Both single nucleotide variants (SNVs), which include insertion or deletion of a few

nucleotides (indels), and copy number variants (CNVs, those that affect from 50 to millions of base pairs) have been studied in rare variant analyses. Depending on the design of the study, rare variants can be identified in a case/control comparison, similar to GWAS, or in pedigree or family trio studies. The latter allow for the identification of *de novo* variants, which are present in the affected offspring but absent in the parents (Bergen et al., 2019). Because schizophrenia leads to a reduction in fecundity (McGrath et al., 1999), it has been hypothesized that rare and *de novo* variants play an important role in schizophrenia genetic aetiology (Howrigan et al., 2020).

The huge amount of computer memory that whole genome sequencing experiments (WGS) require hinders the design of studies with sufficient sample size to detect significant associations (Pervez et al., 2022). Consequently, SNV analyses in schizophrenia have systematically excluded non-coding regions. A whole exome sequencing (WES) case/control study in the European population that combined data from over 24,000 schizophrenia cases and 97,000 controls, identified 22 genes harbouring damaging SNVs that achieved substantial levels of significance (false discovery rate < 0.05), with 10 reaching genome-wide significance ($p < 2 \times 10{-}6$) (Singh et al., 2022). The most significant gene was *SETD1A,* which codes for a component of a histone methyltransferase protein complex involved in gene regulation from early development (Singh et al., 2022; Wang et al., 2021). Not only European ancestry studies, but also South African and Taiwan ancestries, have shown that damaging SNVs are enriched in conserved and LoF-intolerant genes in schizophrenia (Gulsuner et al., 2020; Howrigan et al., 2020). The most recent family trio study, including data from 3,444 schizophrenia parent-proband trios, lacked sufficient power

to detect exome-wide significant variants, but found enrichment of damaging *de novo* SNVs in LoF-intolerant genes in probands (Rees et al., 2020).

1.4.2.2.1 Copy number variants

The same data generated by SNP arrays in GWAS can be utilized to detect CNVs. The largest CNV study to date, in over 21,000 schizophrenia cases and 20,000 controls found eight loci that together explain 0.85% of the genetic liability to the disorder (Marshall et al., 2017). The CNV locus with the highest penetrance was a 1-3 Mb deletion in 22q11.2, which confers a 20-fold increased risk, with nearly 25% of carriers developing schizophrenia (Bassett and Chow, 2008; Marshall et al., 2017). Additionally, CNV burden, measured by the length of genes affected, was increased in schizophrenia, while no contribution to genetic risk was found from CNVs that do not disrupt genes, or in other words, are exclusively non-coding (Marshall et al., 2017). Moreover, schizophrenia CNVs are enriched in LoF-intolerant genes, with the signal concentrated among deletions, while the rest of genes affected by CNVs may not contribute to genetic aetiology (Singh et al., 2017).

1.5.3 The complex genetic aetiology of schizophrenia

Schizophrenia is a polygenic disorder that is often comorbid with other psychiatric conditions (Kessler et al., 2005). This fact is reflected in the high correlation of common genetic liability between schizophrenia and bipolar disorder (r_g = 0.68), and the substantial genetic correlation of schizophrenia with other psychiatric disorders (r_g ranging from 0.22 to 0.34) (Grotzinger et al., 2022). Genetic correlations of common variants between psychiatric conditions are replicated in other non-European populations (Trubetskoy et al., 2022). In addition, rare variants are shared between

psychiatric and neurodevelopmental disorders (Niarchou et al., 2019; Singh et al., 2022, 2017). For example, although the 22q11.2 deletion has been reported as most highly prevalent in schizophrenia, individuals with ASD, ADHD, and mood and anxiety disorders also exhibit high frequency of the 22q11.2 deletion (Schneider et al., 2014). This pattern is also observed for other CNV loci associated with schizophrenia (Niarchou et al., 2019; Rees et al., 2014b; Rees and Kirov, 2021; Thygesen et al., 2018).

Schizophrenia exhibits lower CNV penetrance compared to other neurodevelopmental disorders, with rare CNV and common SNP burden interacting in genetic risk (Bergen et al., 2019; Kirov et al., 2014; Tansey et al., 2016; Vassos et al., 2010). Additionally, gene-set enrichment analysis revealed that the burden of *de novo* variants in schizophrenia is significantly higher in genes previously implicated in multiple neurodevelopmental disorders (Singh et al., 2022). It is also believed that genetic and environmental factors may interact to increase the risk of schizophrenia. For example, some genetic variants may increase susceptibility to environmental stressors such as cannabis use, while others may protect against them (Hillmer et al., 2021). The use of genetic diagnosis only for schizophrenia might therefore still be challenging because pleiotropic effects that confer a risk for general psychiatric psychopathology may need to interact with specific genetic factors and environmental stressors to cause the disease.

1.6 Missing heritability

Narrow sense heritability is the portion of phenotypic variance in a population attributable to additive genetic factors. Early studies of human genetics suggested that genetic variants could explain a large proportion of the phenotypic spectra of highly heritable complex traits and diseases. The “missing heritability” problem arose when

genetic studies failed to identify these variants in psychiatric disorders (Owen and Williams, 2021). Remarkably, heritability of schizophrenia is estimated to be approximately 80% from twin/family studies, in contrast to the heritability explained by SNPs (24%) and CNVs (0.85%) (Marshall et al., 2017; Owen et al., 2016; Trubetskoy et al., 2022). Leaving other confounders aside (such as the shared environment in family studies), the disparity between estimates is not only attributed to the lack of statistical power due to small sample sizes, but also to the inability to identify or interpret variants across the genome (Owen and Williams, 2021).

The heritability explained by CNVs is strikingly small despite the tremendous efforts made to date, which analysed more than 40,000 individuals (Marshall et al., 2017). During the CNV calling process, many events are filtered out to ensure the avoidance of noisy signals that arise, for example, in segmental duplications (Ebbert et al., 2019). Moreover, CNVs affect numerous genes and non-coding regions and, on many occasions, it is difficult to identify the critical region that underpins the genetic risk to schizophrenia (Chen and The gnomAD consortium, 2022).

The succeeding chapters of this thesis are dedicated to addressing the long-standing scientific neglect surrounding camouflaged and non-coding regions, particularly in the context of schizophrenia. The aims of chapters 2 and 3 is to shed more light on their potential genetic contributions to the pathogenesis of schizophrenia. In chapter 4, the focus shifts towards the investigation of the functional role in the brain of a recently evolved non-coding RNA within a rapidly evolving segmental duplication locus – the *Pregnancy Specific Glycoprotein* (*PSG*) locus. Through these integrated efforts, this thesis seeks to close the cycle of scientific neglect and pave the way for significant

advancements in our understanding of the genetic architecture and functional aspects of schizophrenia and brain disorders.

Chapter 2: Contribution of copy number variants in segmental duplications to schizophrenia genetic liability

2.1 Summary

Schizophrenia is a common neurodevelopmental disorder characterized by delusions, hallucinations, and cognitive impairment. Due to the limited progress made in identifying genetic variants associated with schizophrenia, the genetic architecture of the disorder remains largely unknown. Large-scale analyses have identified several CNV loci associated with schizophrenia, but much of the genetic risk from CNVs for schizophrenia remains to be unaccounted for. To address this gap in knowledge, an association study was conducted in a dataset of 8,261 schizophrenia cases and 16,410 controls. Segmental duplications are a source of CNVs and recently evolved genomic elements involved in unique human brain traits. This work focuses on CNVs that overlap with a segmental duplication by more than 50% of its length (Figure 1), which have been excluded previously. A new ultrarare duplication locus affecting the dopamine receptor D5 gene (*DRD5*) was identified in 9 cases and 0 controls. This discovery sheds new light on the genetic basis of schizophrenia and supports the hypothesis that dopaminergic pathways may be important in the development of the disorder. Additionally, this work underscores the importance of including previously neglected regions of the genome in GWAS and provides a foundation for future studies aimed at understanding the genetic susceptibility to other psychiatric disorders.

2.2 Introduction

2.2.1 Classifications of CNVs

Copy-number variants (CNVs) are structural genomic alterations that involve the duplication and deletion of from 50 bp to megabase-scale genomic segments throughout the human genome (Hujoel et al., 2022; Zarrei et al., 2015). CNVs have been shown to contribute to a wide range of disorders, including neuropsychiatric diseases (Hujoel et al., 2022; Malhotra and Sebat, 2012; Marshall et al., 2017; Rees et al., 2014c; Rees and Kirov, 2021; Yoon and Mao, 2021). Their phenotypic impact is not only limited to LoF events affecting protein-coding genes. CNVs can also have unique functional consequences not comparable with those of single nucleotide polymorphisms (SNPs). For example, duplications can increase gene dosage, while deletions can eliminate regulatory elements (Malhotra and Sebat, 2012).

2.2.1.1 Recurrent and non-recurrent CNVs

Several different ways to categorize CNVs beyond duplications and deletions exist. They can be classified as recurrent or non-recurrent based on the boundaries of the CNV. Recurrent CNVs have boundaries that are identical or nearly identical among different individuals, whereas non-recurrent CNVs have different boundaries even when they encompass the same genomic region. Recurrent CNVs typically arise through non-allelic homologous recombination (NAHR) of segmental duplications. In contrast, non-recurrent CNVs are typically formed through less frequent mechanisms such as non-homologous end joining (NHEJ), fork stalling and template switching (FoSTeS), and microhomology-mediated break-induced replication (MMBIR) (Arlt et al., 2012; Lee et al., 2007; Liu et al., 2012, 2011).

Recurrent CNVs are more likely to have been identified and characterized because of their stability across different individuals, whereas non-recurrent CNVs may be more challenging to detect and analyze due to their variable boundaries (Liu et al., 2012, 2011). However, non-recurrent CNVs can also be important contributors to disease and have been implicated in various genomic disorders (Arlt et al., 2012; Ishizuka et al., 2020; Kirov et al., 2009; Lowther et al., 2017; Schaaf et al., 2012). CNV events associated with disease are believed to contain a "critical" region, which refers to the smallest genomic segment within a CNV that is necessary and sufficient to cause a particular phenotype. While CNV loci might have recurrent or non-recurrent boundaries, identifying the critical region is important for understanding the molecular basis of a phenotype and developing targeted interventions or therapies for individuals with the CNV (Riggs et al., 2020).

2.2.1.2 Common and rare CNVs

Another classification system of CNVs is based on their frequency in the population. Common CNVs are generally defined as those that occur with a frequency of more than 1% in the general population. These CNVs are contained in regions of the genome that are more susceptible to genomic rearrangements, such as segmental duplications and therefore, have recurrent boundaries (Kim et al., 2008). In contrast, rare CNVs are those that occur with a frequency of less than 1% in the general population. These CNVs are often larger in size and are typically associated with more severe pathogenicity. Rare CNVs can also be *de novo*; meaning they are not inherited from either parent but arise spontaneously during gametogenesis or early embryonic development, with either recurrent or non-recurrent boundaries (Li et al., 2020; Thygesen et al., 2018).

Notably, the distinction between common and rare CNVs is not always straightforward, as the frequency cutoff used to define rare CNVs can vary depending on the specific population being studied and the genomic regions being examined. Additionally, some CNVs may be classified as rare in one population but common in another (Aguirre et al., 2019; Kato et al., 2010).

2.2.1.3 Genic and non-genic CNVs

Genic CNVs and non-genic CNVs refer to the location of the variant boundaries with respect to the genomic regions containing protein-coding genes. Genic CNVs involve gains or losses of entire coding genes or exons. These variants have the potential to directly impact protein expression or mRNA splicing. It is believed that the genetic contribution of CNVs to schizophrenia liability is concentrated in genic CNVs (Marshall et al., 2017).

Non-genic CNVs, on the other hand, occur in intergenic regions, regulatory regions such as enhancers or promoters; or in regions containing non-coding RNAs, repetitive elements, or structural elements such as centromeres or telomeres. These variants may affect the expression of nearby genes by altering the chromatin structure or the distance between the gene and its regulatory elements, without directly affecting the gene sequence. Although they are believed to be less important than genic CNVs, non-genic CNVs have also been linked to various diseases, such as cancer and developmental disorders, through their effects on gene expression regulation (Han et al., 2020; Spielmann and Klopocki, 2013; Zhang and Lupski, 2015).

2.2.2 Detection methods for CNVs

Several technologies can be used to detect CNVs. The choice of the appropriate method should depend on the research question, the sample size, the available resources, and the level of resolution required.

2.2.2.1 Microarray

These methods use microarray platforms to detect CNVs and rely on the differential hybridization of probes representing different regions of the genome to identify CNVs. Comparative genomic hybridization (CGH) involves the comparison of two DNA samples labelled with different fluorophores, a test sample (often from a patient with a disease) and a reference sample (typically from a healthy individual), which are then co-hybridized on the microarray. The resolution of CGH is restricted to alterations of approximately 5-10 Mb (Weiss et al., 1999). On the other hand, SNP genotyping arrays can be used to detect CNVs with a resolution ranging from hundreds of base pairs to several megabases depending on the probe coverage, and they can also provide information on the zygosity, boundaries, and size of CNVs, without the requirement of a test and reference sample pair (Colella et al., 2007; Oldridge et al., 2010; Wang et al., 2007). Because the same data generated for GWAS can be employed for the detection of CNVs, this is the most common technique in large-scale CNV studies. In most cases, probe data is processed with PennCNV, which utilizes a hidden Markov model (HMM) based logarithm that identifies CNVs from fluorescence intensity signals and allele frequencies (Wang et al., 2007).

2.2.2.2 Next-generation sequencing (NGS)

NGS-based methods for CNV detection rely on the read depth and/or split-reads analysis of short reads generated from whole-genome sequencing or whole-exome sequencing (Coutelier et al., 2022; Moreno-Cabrera et al., 2020). The read depth approach involves counting the number of reads that map to each genomic region and comparing it to the expected coverage based on the genome-wide average. CNVs can be detected as regions with increased or decreased read depth compared to the reference (Abyzov et al., 2011; Boeva et al., 2011; Fowler, 2022; Jiang et al., 2015; Johansson et al., 2016; Povysil et al., 2017; Rajagopalan et al., 2020; Talevich et al., 2016). Split-reads analysis, on the other hand, involves identifying reads that span the breakpoint of a CNV and align partially to two different genomic regions. CNVs can be inferred by detecting clusters of split reads that overlap within a given region (X. Chen et al., 2016; Rausch et al., 2012). While NGS methods overcome the limitation of resolution and coverage biases from SNP arrays, they demand high costs and large storage systems, which limit their use in large scale studies (Coutelier et al., 2022; Moreno-Cabrera et al., 2020).

2.2.2.3 Polymerase chain reaction (PCR)

These methods use the PCR technique to amplify specific regions of the genome and detect changes in copy number. They can be used for targeted detection of CNVs or validation of candidate loci from large-scale studies. One commonly used method is quantitative PCR (qPCR), which measures the relative copy number of a specific genomic region by comparing the amount of PCR product from the target region to that of a reference region. Typically, the polymerization reaction is engineered to produce

fluorescence when the target sequence is being amplified. This method is simple, inexpensive, and can be used with a small amount of starting material, but it has limited resolution and may not accurately detect CNVs that are smaller than the size of the PCR amplicon (Ma and Chung, 2014; Ponchel et al., 2003).

Another PCR-based method for CNV detection is digital droplet PCR (ddPCR), which partitions the sample into thousands of small qPCR reactions to achieve an absolute quantification of the target sequence. This method can detect CNVs with higher precision and resolution than simple qPCR and can be used with low input DNA, but it is generally more expensive and time-consuming (Taylor et al., 2017).

2.2.3.4 Karyotype and Fluorescence in situ hybridization (FISH)

A karyotype is an ordered depiction of the chromosomes of an individual, organized by size, banding pattern, and centromere position (Tjio and Levan, 1956). Karyotyping can detect large-scale chromosomal abnormalities such as aneuploidy (an abnormal number of chromosomes) and structural rearrangements such as chromosomal translocations that result in chromosomal arm aneuploidy. It constituted one of the first diagnostic tools for genetic abnormalities such as Down syndrome, Turner syndrome, and Klinefelter syndrome (Gekas et al., 2011; Jacobs and Strong, 1959; Sybert and McCauley, 2004). Karyotyping is also utilized in cancer diagnosis and research to uncover chromosomal abnormalities linked to tumour formation or progression (Baudoin and Bloomfield, 2021). FISH uses fluorescent probes that hybridize to specific DNA sequences and can be used to visualize the location and copy number of a specific gene or genomic region in the chromosome. This method has higher

resolution than karyotyping and can detect smaller CNVs or changes in specific genes (du Manoir et al., 1993; Kallioniemi et al., 1996).

2.2.3 CNVs and their contribution to schizophrenia genetic risk

CNVs have been a major focus of research in schizophrenia due to their potential impact on genes involved in important neurodevelopmental processes such as neurotransmitter signaling, synapse formation, and neuronal migration (Girirajan et al., 2012, 2011; Marshall et al., 2017). In neuropsychiatric disorders, much attention has been given to rare CNVs, as it is believed that these variants have a greater effect size on disease risk compared to common variants (Kirov et al., 2014; Legge et al., 2021; Rees and Kirov, 2021). CNV loci previously implicated in schizophrenia are discussed below and summarized in Table 1.

2.2.3.1 Deletion at 22q11.2 (DiGeorge syndrome)

The 22q11.2 deletion has been extensively studied in schizophrenia. The deletion encompasses a 3 Mb region with recurrent boundaries overlapping segmental duplications and affecting more than 60 genes. The "critical" region has been narrowed down to the hemizygous loss of at least 30 of these genes. Despite this, the specific genes that contribute to schizophrenia susceptibility remain unclear. This is further complicated by the fact that several candidate genes, including *PRODH*, *PIK4CA* and *ZDHHC8*, are in regions of segmental duplication (Marshall et al., 2017; Qin et al., 2020; Rees et al., 2014c).

Individuals with 22q11.2 deletion syndrome have a significantly increased risk of developing schizophrenia compared to the general population (20-30 odds ratio), with

approximately 0.29% of schizophrenia patients having this deletion (Marshall et al., 2017; Rees et al., 2014c). Of note, protective effects have been attributed to duplications of this locus (Rees et al., 2014a). Schizophrenia in 22q11.2 deletion syndrome shares core symptoms, treatment response, and MRI brain anomalies with schizophrenia in the general population. However, individuals with 22q11.2 deletion syndrome may have distinguishable physical traits (such as mild facial dysmorphic features), different auxiliary clinical characteristics (congenital heart defects), and may have cognitive deficits diagnosed as other psychiatric disorders in childhood, such as ADHD, ASD, and generalized anxiety disorder (McDonald-McGinn et al., 2015; Qin et al., 2020).

2.2.3.2 Duplication at 16p11.2

The 16p11.2 duplication spans a region of 600 Kb enclosed by recurrent boundaries, affecting at least 30 genes. The 16p11.2 locus has attracted attention due to its bidirectional link to certain neurodevelopmental and psychiatric diseases. Deletions at 16p11.2 increase the risk of developing ASD and intellectual disabilities, but not schizophrenia (Cooper et al., 2011; Marshall et al., 2017; Rees et al., 2014c). In contrast, a high risk of developing schizophrenia has been shown in carriers with 16p11.2 duplication (odds ratios 8 to 24-fold), but only a small proportion have intellectual disability and they do not generally show autistic traits. Therefore, phenotypic effects of the 16p11.2 duplication may be more specific to schizophrenia than to intellectual disability, unlike the 22q11 deletion (Chang et al., 2017; Kirov et al., 2014; McCarthy et al., 2009; Niarchou et al., 2019). Although many genes in this locus have potential roles in neurodevelopment, no independent genes seem to act alone in its phenotypic effects, and instead, disruption of the hippocampal-orbitofrontal and

hippocampal-amygdala pathways has emerged as the most probable explanation for the involvement of the 16p11.2 duplication in schizophrenia (Bristow et al., 2020).

2.2.3.3 Deletions and duplications at 1q21.1

The 1q21.1 deletion and duplication have also been linked to increased risk of schizophrenia (deletions odds ratio = 6.8 and duplication odds ratio = 2.3), as well as other neuropsychiatric disorders (Marshall et al., 2017; Yoon and Mao, 2021). The most common form of CNVs in the 1q21.1 locus affects a region of approximately 1 Mb, with adjacent segmental duplications at both sides that constitute the recurrent boundaries. No single gene or critical region has been identified to explain the pathogenic effect. In addition, the clinical representation is very broad, as carriers of the 1q21.1 deletion or duplication are predisposed to other psychiatric and neurodevelopmental disorders such as ASD, anxiety and mood disorders, intellectual disabilities, ADHD, sleep disorders, and speech and motor delays; as well as physical abnormalities in several organs and microencephaly (Yoon and Mao, 2021).

2.2.3.4 Deletions in *NRXN1* (2p16.3)

Non-recurrent deletions of different sizes disrupting exons of the *NXRN1* gene directly implicate this gene in the development of schizophrenia (Kirov et al., 2009; Marshall et al., 2017). The *NXRN1* gene has been described to play an essential role in the synapses of the brainstem and neocortex (Zeng et al., 2013). Variants of *NRXN1* increase the liability of developing schizophrenia (odds ratio = 9), as well as cognitive impairment, ASD, substance dependence, and developmental delay (Ishizuka et al., 2020; Lowther et al., 2017; Schaaf et al., 2012).

2.2.3.5 Deletions at 3q29

The 1.6 Mb recurrent deletion at 3q29 is estimated to occur in approximately 1 in 30,000 individuals in the general population, but in approximately 1 in 100 individuals with schizophrenia, representing, at least, a 30-fold increase in risk. Being so rare, the association between 3q29 deletions and schizophrenia is not as statistically significant as some other CNVs, such as the 22q11.2 deletion or the 16p11.2 duplication, but it has been consistently reported in multiple studies (Baba et al., 2019; Marshall et al., 2017; Mulle et al., 2010; Quintero-Rivera et al., 2010). Individuals with the 3q29 deletion may also exhibit a range of physical and developmental abnormalities, including microcephaly, intellectual disability, and cardiac and skeletal anomalies, and other psychiatric disorders, such as bipolar disorder and ASD (Chirita Emandi et al., 2019; Clayton-Smith et al., 2010; Li et al., 2009; Quintero-Rivera et al., 2010; Willatt et al., 2005). While the specific genes affected by the 3q29 deletion have not yet been clearly identified, it is thought that their disruption may affect neural development and function, leading to an increased susceptibility to schizophrenia (Dasouki et al., 2011).

2.2.3.6 Duplications at 7q11.23 (Williams-Beuren syndrome)

This chromosomal region is involved in several critical developmental processes in the cardiovascular and neurological systems, and recurrent 1.3Mb duplication at this locus is known to cause Williams-Beuren syndrome (WBS), a rare genetic disorder characterized by distinct facial features, developmental delay, ASD and cardiovascular anomalies (Kozel et al., 2021; Sanders et al., 2011; Wang et al., 2023). Recent studies have shown that individuals with a duplication at 7q11.23 have a significantly increased risk of developing schizophrenia compared to the general population, with an estimated

odds ratio of 5.2-fold (Marshall et al., 2017; Mulle et al., 2014). The critical region that contributes to schizophrenia risk is still unclear, and further research is needed to elucidate the underlying mechanisms of this association (Qaiser et al., 2021).

2.2.3.7 Deletions and duplications at 15q11-q13

The segmental duplication enrichment in this genomic region facilitates the generation of recurrent CNVs. Large but ultrarare duplications, involving more than 5.5 Mb, are associated with the Prader-Willi and Angelman syndromes, which are characterized by developmental delay, metabolic and behavioral phenotypes, and confer more than 10-fold increase in the risk of suffering schizophrenia (Marshall et al., 2017). Other smaller and more frequent CNVs at this locus show less prominent odds ratios for schizophrenia (1.8-4.6) (Marshall et al., 2017; Rees et al., 2014c). An interesting feature of this locus is the high concentration of genes that undergo genomic imprinting, a phenomenon that heritably silences the expression of one of the parental alleles of certain genes. Critical regions thus only have an impact on the phenotype if they are disrupted on a specific parental allele; for example, the Angelman syndrome is caused by the loss of the maternally inherited *UBE3A* gene allele (Amos-Landgraf et al., 1999; Ingason et al., 2011a; Kalsner and Chamberlain, 2015).

2.2.3.8 Other deletions and duplications at 16p

With substantially lower odds ratios compared to other loci (1.7-3.3), a recurrent 700 Kb duplication at 16p13.11 and a recurrent 500 Kb deletion at 16p12.1 have been shown to contribute to schizophrenia genetic risk. Both are multigenic and a critical region has not been discovered. Other neurodevelopmental phenotypes such as

intellectual disability, developmental delay and ASD have been associated with 16p13.11 duplication and 16p12.1 deletion carriers (Cai et al., 2022, p. 13; Ingason et al., 2011b; Rees et al., 2014c).

2.2.3.9 The discovery gap

Previously implicated CNVs can only confidently explain up to 0.85% of schizophrenia heritability, in contrast to the estimated heritability of the disorder by family/twin studies (80%) (Marshall et al., 2017; Owen et al., 2016; Owen and Williams, 2021). Even though large-scale, genome-wide CNV analyses have been performed in schizophrenia, they have shown a lack of efficiency for the discovery of new loci, in contrast with common variants efforts with GWAS (Marshall et al., 2017; Rees et al., 2014c; Trubetskoy et al., 2022). One of the reasons new loci are not being discovered might be the systemic disregard of crucial regions, such as segmental duplications, in the CNV field. In previous large-scale studies, CNV events overlapping a segmental duplication by more than 50% of their length have been excluded due to the expected low specificity of SNP probes in repetitive regions (Figure 1). However, segmental duplications are the source of almost all the CNVs implicated in schizophrenia, and they harbour genes and regulatory regions responsible for cognitive traits that are uniquely human (Erady et al., 2022; Florio et al., 2018; Mangan et al., 2022; Marshall et al., 2017). In this chapter, high quality probes are selected from SNP microarray platforms to improve the confidence of CNV calls overlapping a segmental duplication by more than 50% of their length (Figure 1). Associations with schizophrenia are tested only in this set of CNVs, extracted from raw data used in previous large-scale analyses. The method developed in this chapter assumes that enough paralog sequence in the

segmental duplication has evolved independently to contain specific sequence that SNP probes can target.

Table 1. List of CNV loci previously implicated in schizophrenia.

First identified in (Rees et al., 2014c), odds ratio and p-values from the most recent large-scale study (Marshall et al., 2017). * PVAL = meta-analysis P value. ** OD = Meta-analysis odds ratio.

LOCUS	UNIQUE REGION (HG19)	P VAL*	OD**	BOUNDARIES
1q21.1 del	chr1:146527987-147394444	0.00000021	6.8	Segmental duplication
1q21.1 dup	chr1:146527987-147394444	0.022	2.3	Segmental duplication
NRXN1 del	chr2:50145643-51259674	0.000028	4.5	Non-recurrent
3q29 del	chr3:195720167-197354826	0.0002	18	Segmental duplication
WBS dup	chr7:72744915-74142892	0.045	5.2	Segmental duplication
PWS/AS dup	chr15:22805313-28390339	0.0000004	Inf	Segmental duplication
15q11.2 BP1-BP2 del	chr15:22805313-23094530	0.000023	1.8	Segmental duplication
15q13.3 del	chr15:31080645-32462776	0.0015	4.6	Segmental duplication
16p13.11 dup	chr16:15511655-16293689	0.0022	1.7	Segmental duplication
16p12.1 del	chr16:21950135-22431889	0.00034	3.3	Segmental duplication
16p11.2 dup	chr16:29650840-30200773	3.7E-15	11	Segmental duplication
22q11.2 (DiGeorge/VCFS syndrome) del	chr22:19037332-21466726	0	Inf	

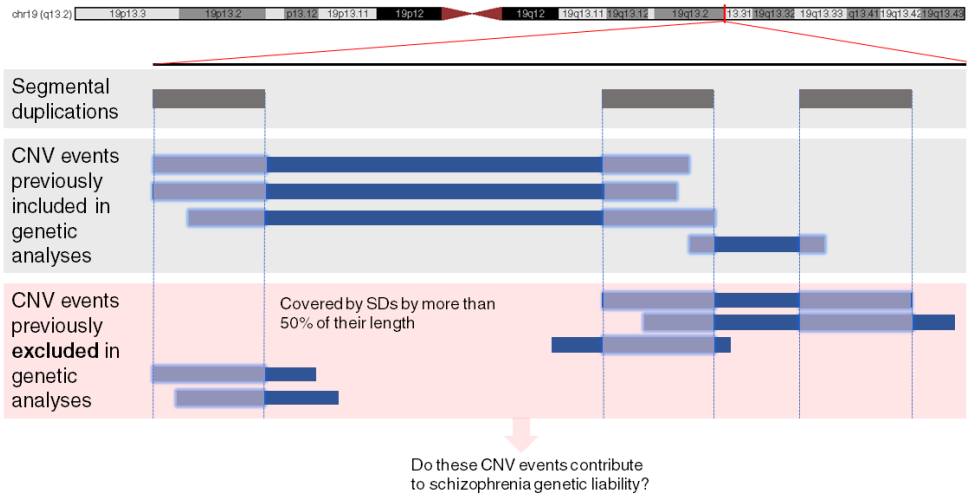


Figure 1. Visual abstract of chapter 2.

Segmental duplications (such as the *PSG* locus on chromosome 19q13.2) are the source of recurrent CNVs included in genetic association analyses and associated with schizophrenia (Table 1). In CNVs previously included in genetic association analyses, segmental duplications are predominantly found only at the boundaries. Therefore, while some of the microarray probes are expected to be low quality at the boundaries, the majority of sequence covered by the CNV is specific, and therefore they are included in analyses. CNVs that are overlapped by segmental duplications by more than 50% of their length are expected to be covered by too many low quality probes and are filtered out before the association test. In this chapter, we focus specifically on the previously excluded CNVs, and develop a method to study them in the context of schizophrenia genetic liability.

2.3 Results

2.3.1 Gene-wise analysis

Association of CNV loci with a phenotype in case/control studies can be performed by assessing which genes are disrupted by a set of overlapping CNV, rather than focusing on the CNVs per se. This approach is more informative because CNV critical regions tend to be in coding genes (Marshall et al., 2017). This analysis employed raw data from the Molecular Genetics of Schizophrenia (MGS) study (2,200 schizophrenia cases and 2,571 controls, Affymetrix Genome-Wide Human SNP Array 6.0, with 1,425,313 probes used) and the ClozUK study (6,061 schizophrenia cases and 13,839 controls, common probes from several Illumina microarrays used [Table 4], with 295,912 probes used). All controls were individuals with no psychiatric diagnoses. First, each protein-coding gene was evaluated for exon-disrupting CNVs in schizophrenia cases and controls in each study independently. The analysis was restricted to protein coding genes in segmental duplications plus adjacent genes disrupted by a set of CNVs that overlapped more than 50% of a segmental duplication (number of genes = 2644). In the MGS dataset, duplications affected 151 genes, while deletions disrupted a total of 105 genes. Only 6 genes in duplications and 5 genes in deletions were enriched in cases or controls with nominal levels of significance (logistic regression $P < 0.05$; Supplementary Table 1, Supplementary Table 4). The most significant gene affected by duplications in MGS was *DPY19L2* on chromosome 12, but it was not captured in the ClozUK study. *DPY19L2* is predominantly expressed in the testis and mutations and deletions in this gene have been found in infertile men (Supplementary Table 1) (Harbuz et al., 2011, p. 1; Zhu et al., 2013). This signal can be seen on chromosome 12 in Figure 2. Genes encoding olfactory receptors (OR4N2, OR4M1, OR4Q3, OR11H2)

in chromosome 14 are affected by deletions only in MGS (Supplementary Table 4). Olfactory receptors are expressed in sensory neurons in the nose and have expanded in mammals by gene duplication to allow for a more complex perception of different odors (Malnic et al., 2004). This signal can be seen in Figure 3. In the ClozUK samples, duplications spanned 74 genes whereas deletions were found in 34 genes. Only 5 genes in the duplications set and 1 gene in the deletions reached nominal significance for the association with either cases or controls (Supplementary Table 2, Supplementary Table 5). None of the genes in either study survived the conservative Bonferroni correction for multiple testing of 2,644 genes ($P < 1.89 \times 10^{-05}$). It was noted that more genes were affected by duplications than deletions in both datasets (MGS 151 versus 105; ClozUK 74 versus 34). The smaller number of disrupted genes in the ClozUK datasets compared to the MGS dataset was counterintuitive, since ClozUK is a more sizable study, and thus, more events affecting a higher number of genes is expected. This observation reflects the difference in the number of probes of the SNP arrays used in each study. The SNP arrays in the ClozUK datasets shared less than 300,000 common high-quality probes, in contrast to MGS, for which more than a million high quality probes could be used for CNV calling, conferring more detection sensitivity to the MGS dataset (See 2.5.2 Exclusion of low-quality probes and Table 4, page 76-77). Moreover, the Illumina arrays used in the ClozUK studies specifically select against probes in segmental duplications. Consequently, many CNVs in segmental duplications may have been overlooked in the ClozUK datasets.

Both datasets were combined to improve statistical power (8,261 schizophrenia cases and 16,410 controls, henceforth referred to as the combined dataset). In this case, 10 genes from 7 different duplication loci exhibited nominal significant enrichment in

either cases or controls (Figure 2, Supplementary Table 3). In deletions, 6 genes were affected by 3 different loci surpassing nominal significance ($P < 0.05$) (Figure 3, Supplementary Table 6). Low overlap in genes affected was observed between studies (Figure 2, Figure 3). After filtering for genes that were affected by CNV events in both datasets, only two candidate genes remained, *DRD5* (duplications, Figure 2) and *TARP* (deletions, Figure 3). However, deletions in *TARP* were more frequent in cases in MGS (0.5% in cases versus 0.04% in controls; Supplementary Table 4) while only controls in the ClozUK study showed such deletions (0% in cases versus 0.03% in controls; Supplementary Table 5). For this reason, deletions in *TARP* were not considered further. The *TARP* gene encodes for an unrearranged version of the T cell receptor γ chain that is expressed in non-lymphoid tissues, especially in the prostate. Variation in this gene has been linked to the concentration of plasminogen activator inhibitor-1 in plasma (Lanktree et al., 2010).

Duplications spanning *DRD5* occurred only in cases in both studies. The *DRD5* duplication frequency in cases in MGS was 0.09%, which is consistent with the frequency of 0.08% observed in ClozUK schizophrenia cases. The gene *DRD5* encodes for the D5 subtype of the dopamine receptor. The dopamine pathway disruption is the main molecular explanation for positive symptoms in schizophrenia (Howes and Kapur, 2009; Howes and Onwordi, 2023). Although the *DRD5* duplication was the best novel candidate, it failed to survive correction for multiple testing (logistic regression $P = 6.20 \times 10^{-4}$, odds ratio (OR) = 2.24×10^{3}, 95% confidence interval (95%CI) = 1.55×10^{1}–1.69×10^{8}) (Figure 2). When all CNVs, including those that are not spanned by segmental duplications by more than 50% of their length were inspected through UCSC Genome Browser, two other duplications were found to disrupt *DRD5* in cases (**Error! R**

eference source not found.). In the future, odds ratios and p values should be updated using all CNVs. Caution should be taken with extremely high odds ratios, since they could be inflated due to low sample size with no observations in controls. Because the *DRD5* duplication appears to be ultrarare (frequency less than 0.1%), this discovery would benefit from a more statistically powerful dataset.

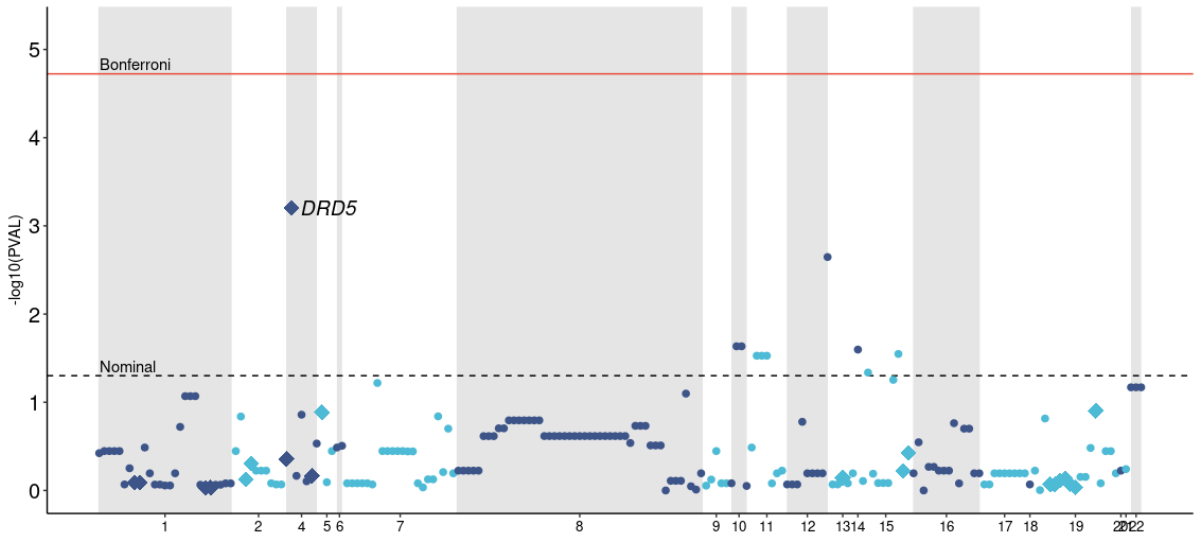


Figure 2. Manhattan plot for the gene-wise analysis of duplications

Loci supported by both MGS and ClozUK are represented by diamonds instead of circles. The two significance thresholds are highlighted (nominal p = 0.05, Bonferroni p = 1.89×10^{-05}). The significant locus supported by both studies on chromosome 4 is highlighted (*DRD5*). On chromosomes 10 to 15, several loci reach nominal significance but are only supported by one study (Supplementary Table 1, Supplementary Table 2, Supplementary Table 3).

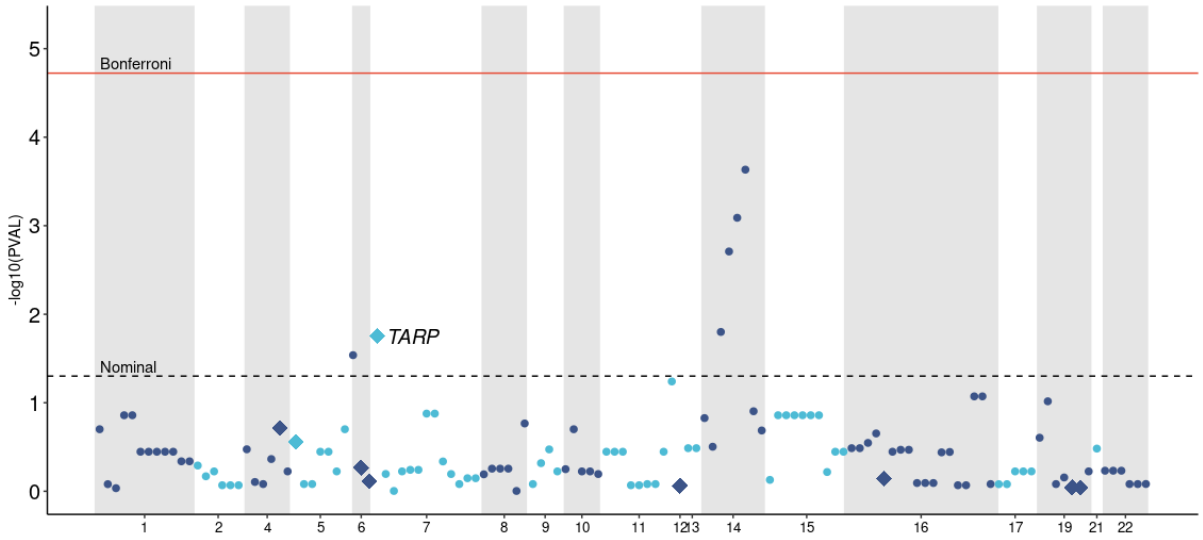


Figure 3. Manhattan plot for the gene-wise analysis of deletions.

Loci supported by both MGS and ClozUK are represented by diamonds instead of circles. The two significance thresholds are highlighted (nominal p = 0.05, Bonferroni p = 1.89×10^{-05}). The gene *TARP* is highlighted because it is supported by both studies. Significant loci on chromosomes 6 and 14 are only found in the ClozUK and MGS studies, respectively (Supplementary Table 4, Supplementary Table 5, Supplementary Table 6).

2.3.2 Genome-wide analysis of CNV burden

Possibly because the sample size was insufficiently powered and the two studies showed low overlap in SNP array probes, genome-wide significant associations at individual loci were not observed. An alternative in these situations is to assess the burden of pools of variants between cases and controls instead of single loci and analyze which general characteristics of CNVs could be contributing to schizophrenia risk (Legge et al., 2021; Marshall et al., 2017). For example, rare CNVs in schizophrenia are more abundant, longer, and affect more genes compared to controls; in other words, CNV burden is elevated in schizophrenia. However, CNVs that overlapped more than 50% of a segmental duplication had been excluded (Marshall et al., 2017). In this study, CNV burden was analyzed in the set of CNVs that overlapped by a segmental duplication by more than 50% of their length (Figure 1). First, CNV burden in each individual was measured by the number of CNV deletions or duplications they exhibited. No significant enrichment was found (combined dataset, CNVs, OR = 0.87, 95% CI = 0.65-1.17, P = 3.67 x 10^{-1}; Supplementary Table 7) Increased burden by length (combined dataset, CNVs, OR = 0.99, 95% CI = 0.99-1.00, P = 4.86 x 10^{-1}; Supplementary Table 8) or the number of genes affected (combined dataset, CNVs, OR = 1.04, 95% CI = 0.92-1.17, P = 5.42 x 10^{-1}; Supplementary Table 9) was also not observed.

When inspecting the two studies separately, MGS showed increased burden by number of genes affected (MGS, CNVs, OR = 1.11, 95% CI = 0.97-1.26, P = 1.31 x 10^{-1}), whereas the ClozUK dataset gave the opposite result (ClozUK, CNVs, OR = 0.68, 95% CI = 0.48-0.98, P = 3.66 x 10^{-2}). Such conflicting results in the two cohorts may reflect the non-overlapping SNP probe coverage between the datasets and the limited number

of probes within segmental duplications, which may have hindered sensitivity, particularly in the ClozUK datasets.

2.3.2.1 Gene-set burden analyses

A primary reason for studying CNVs in segmental duplications is that these genomic regions harbour human-specific genes and open chromatin regions essential for brain development (Florio et al., 2018; Mangan et al., 2022). In addition, genomic regions with accelerated evolution in humans are enriched in variants implicated in enhanced cognition and schizophrenia, which are traits characteristic of humans (Corces et al., 2020; Erady et al., 2022; Xu et al., 2015). In this chapter, it was hypothesized that CNV burden by number of genes affected could be concentrated in great ape specific genes important for brain evolution. To assess this, CNVs were mapped to a previously described gene set of 50 great ape specific genes preferentially expressed in progenitors of fetal neocortex (Florio et al., 2018). To avoid disturbance of the signal from global gene enrichment in the different datasets, correction for the total number of genes affected was added to the analysis. Deletions in great ape specific genes preferentially expressed in progenitors of fetal neocortex were enriched in controls, indicating that these variants could have a protective effect (combined dataset, deletions, OR = 0.10, 95% CI = 0.01-0.61, P = 9.20×10^{-3}; Table 2). Results were consistent in both MGS and ClozUK datasets.

CNVs were also scanned for disruption of “human ancestor quickly evolved regions” (HAQERs, n = 79) that overlap open chromatin genomic windows in the developing human brain (Kundaje et al., 2015; Mangan et al., 2022). When burden by the number of HAQERs disrupted was measured, the signal was significantly protective for

deletions (combined dataset, deletions, OR = 0. 003, 95% CI = 0.00-0.53, P = 2.32 x 10^{-2}; Table 3), similar to great ape specific genes preferentially expressed in progenitors of fetal neocortex. These results suggest that, while human-specific coding and non-coding elements have had an essential role in the evolution of human brain unique traits, they may also take part in the genetic liability of human brain unique susceptibilities, such as schizophrenia. Albeit an exciting hypothesis, it is important to note that the numbers of individuals contributing to the association are low (Table 2, Table 3), which inflates odds ratios and potentiates winner's curse (Xiao and Boehnke, 2009). Therefore, these conclusions should be treated with caution subject to further replication.

Table 2. CNV Burden by number of disrupted great ape specific genes preferentially expressed in progenitors of fetal neocortex.

Results are organized by study and type of CNV (CNV = deletions and duplications together, DUP = duplications, DEL = deletions). OR = Odds ratio, CI1 – 95% lower limit confidence interval, CI2 – 95% upper limit confidence interval, SE = standard error, NCONTROLS = number of controls with variant and average number of disrupted elements of interest per individual between brackets, NCASES = number of cases with variant and average number of disrupted elements of interest per individual between brackets.

	TYPE	OR	CI1	CI2	SE	PVAL	NCONTROLS	NCASES
MGS	CNV	0.35	0.02	3.49	0.89	3.81E-01	6 (1.33)	1 (2)
	DUP	NA	NA	NA	NA	NA	0 (0)	0 (0)
	DEL	0.24	0.01	2.36	0.60	2.27E-01	6 (1.33)	1 (2)
ClozUK	**CNV**	**0.03**	**0.00**	**0.49**	**0.12**	**9.92E-03**	**22 (1)**	**2 (1)**
	DUP	0.04	0.00	11.13	2.84	3.05E-01	3 (1)	0 (0)
	DEL	**0.05**	**0.00**	**0.79**	**0.20**	**3.11E-02**	**19 (1)**	**2 (1)**
Combined	**CNV**	**0.08**	**0.01**	**0.57**	**0.14**	**9.20E-03**	**28 (1.07)**	**3 (1.33)**
	DUP	0.05	0.00	11.57	2.95	3.13E-01	3 (1)	0 (0)
	DEL	**0.10**	**0.01**	**0.63**	**0.16**	**1.28E-02**	**25 (1.08)**	**3 (1.33)**

Table 3. CNV Burden by the number of disrupted "human ancestor quickly evolved regions" (HAQERs).

Results are organized by study and type of CNV (CNV = deletions and duplications together, DUP = duplications, DEL = deletions). OR = Odds ratio, CI1 – 95% lower limit confidence interval, CI2 – 95% upper limit confidence interval, SE = standard error, NCONTROLS = number of controls with variant and average number of disrupted elements of interest per individual between brackets, NCASES = number of cases with variant and average number of disrupted elements of interest per individual between brackets.

	TYPE	OR	CI1	CI2	SE	PVAL	NCONTROLS	NCASES
MGS	**CNV**	**0.004**	**0.000**	**0.737**	**0.19**	**3.55E-02**	**5 (1)**	**0 (0)**
	DUP	0.035	0.000	14.008	3.57	2.97E-01	2 (1)	0 (0)
	DEL	**0.003**	**0.000**	**0.500**	**0.13**	**2.22E-02**	**5 (1)**	**0 (0)**
ClozUK	CNV	0.009	0.000	9.692	2.47	3.31E-01	2 (1.5)	0 (0)
	DUP	0.017	0.000	19.031	4.85	2.55E-01	1 (1)	0 (0)
	DEL	1.801	0.005	68.798	17.55	7.76E-01	1 (2)	0 (0)
Combined	**CNV**	**0.003**	**0.000**	**0.513**	**0.13**	**2.13E-02**	**7 (1.14)**	**0 (0)**
	DUP	0.019	0.000	4.949	1.26	1.83E-01	3 (1)	0 (0)
	DEL	**0.003**	**0.000**	**0.533**	**0.14**	**2.32E-02**	**6 (1.16)**	**0 (0)**

2.3.3 Validation of the *DRD5* duplication

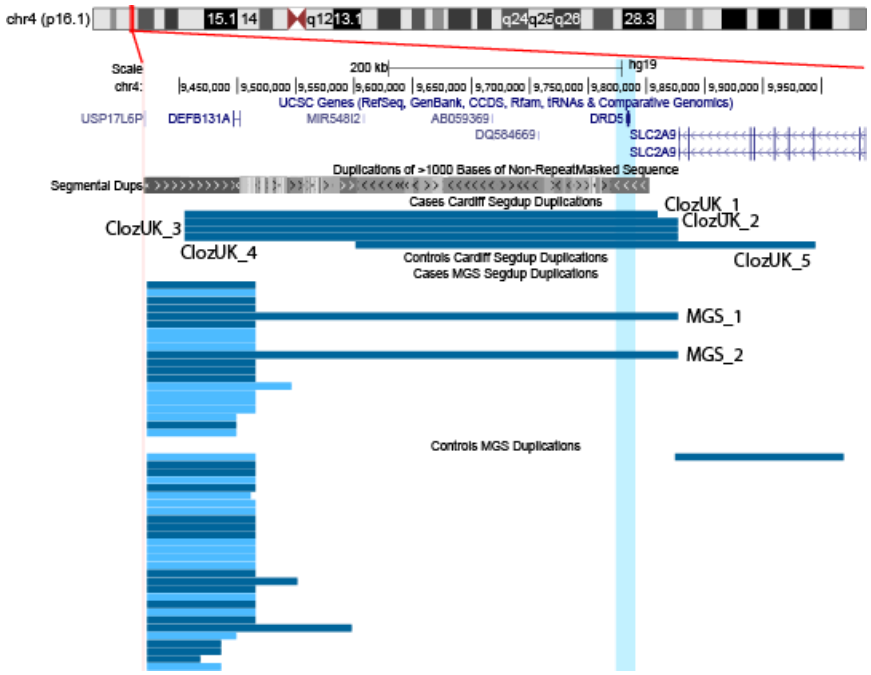
First, *DRD5* duplications were inspected with the UCSC Genome Browser, revealing that they overlap with a significant portion of a segmental duplication, which confirmed that this rare variant would have been excluded from any previous CNV analysis (Figure 4A). A separate locus that is prone to copy number variation was found to overlap approximately 40 Kb with the *DRD5* duplication upstream boundary in the MGS dataset only. Other genes are partially or fully affected by the duplication, including *DEFB131A, MIR546I2, AB059369, DQ584669* and *SLC2A9* (Figure 4A). The protein-coding genes *DEFB131A* and *SLC2A9* have a role in the immunologic response against invading pathogens and the excretion of urate through the kidneys, respectively (Kim et al., 2015; Vitart et al., 2008). Genetic variants in *SLC2A9* and uric acid levels in the blood have been associated with social phobia and anxiety (Lyngdoh et al., 2013). *MIR546I2, AB059369, DQ584669* appear to be non-coding genes with poor experimental characterization. Since *DRD5* takes part in the dopamine pathway and constitutes the only gene inside all possible boundaries of the duplication locus of interest, it is most likely that the *DRD5* gene represents the critical region for the involvement of this CNV locus in schizophrenia pathogenesis. Of note, when examining CNVs that have been included in previous genetic studies (not covered by segmental duplications by more than 50% of their length), two additional duplications in cases disrupt *DRD5* towards their upstream boundary (Figure 4B).

Turning to the SNP coverage in the *DRD5* duplication, it was observed that the Affymetrix array used in MGS provided adequate coverage (up to 90 probes), while only 6 to 24 SNP probes in Illumina arrays from the ClozUK datasets covered the relevant region (Figure 5, Figure 6). The consistent boundaries of the *DRD5* duplication

in MGS and ClozUK studies indicated encouraging evidence for validation (Figure 4A). PennCNV software was utilized to perform CNV calling, which integrates Log R ratios (probe intensity signals, LRR) and B allele frequencies (A/B allele frequencies, BAF), in a hidden Markov model (HMM). Like all statistical models, the HMM used by PennCNV can fail to accurately identify CNVs in certain situations. It is of the utmost importance to review the raw LRR and BAF traces in order to validate CNV calls in cases where the SNP coverage is poor (Wang et al., 2007).

Probe LRR and BAF traces in the MGS dataset clearly followed the trend expected for a duplication event (Figure 5, Figure 6). First, the *DRD5* duplication exhibited a substantial increase in LRR in probes within the locus compared to LRR values of probes beyond the boundaries. In addition, B allele frequencies transitioned from 3 possible states (AA 0.0, AB 0.5, BB 1.0) to 4 states (AAA 0.0, AAB 0.25, ABB 0.75, BBB 1.0). Since many probes in Affymetrix arrays are specific for copy number detection and are only informative for intensity signals (LRR) but do not distinguish between alleles, the B allele frequency trend was supported by considerably less probes (25 probes versus 90 total probes). However, it was still visible (Figure 5). Samples from the ClozUK studies were also confirmed to harbour the *DRD5* duplication through LRR and BAF traces (Figure 6).

A



B

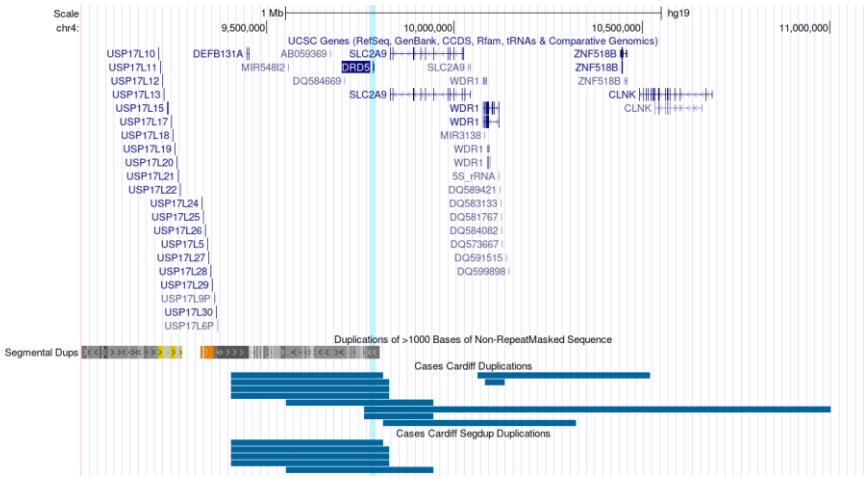
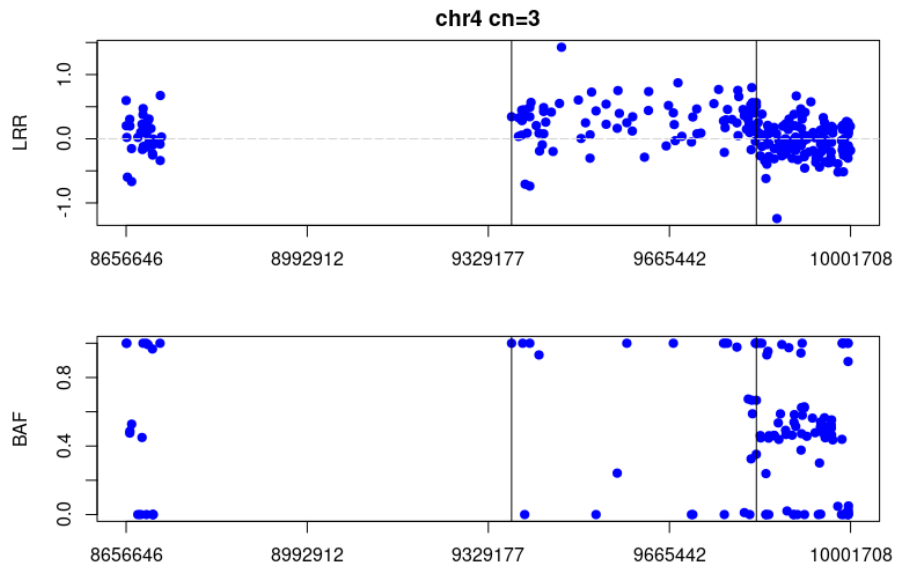


Figure 4. CNV tracks in the *DRD5* duplication locus in 4p16.1 (hg19 genome version).

A) Duplications covered by more than 50% by segmental duplication. CNVs from each individual contributing to the association are named MGS_1, MGS_2, ClozUK_1, ClozUK_2, ClozUK_3, ClozUK_4, ClozUK_5 in the order from top to bottom in their corresponding tracks. B) CNV tracks in the *DRD5* duplication locus at 4p16.1 including CNVs that do not overlap >50% of a segmental duplication in the ClozUK (Cardiff) datasets. Cases Cardiff (ClozUK) Duplications track includes all duplications, whereas Cases Cardiff (ClozUK) Segdup Duplications includes duplications overlapping >50% of a segmental duplication. This shows that two duplications that would have been included in previous analyses affect *DRD5*.

A



B

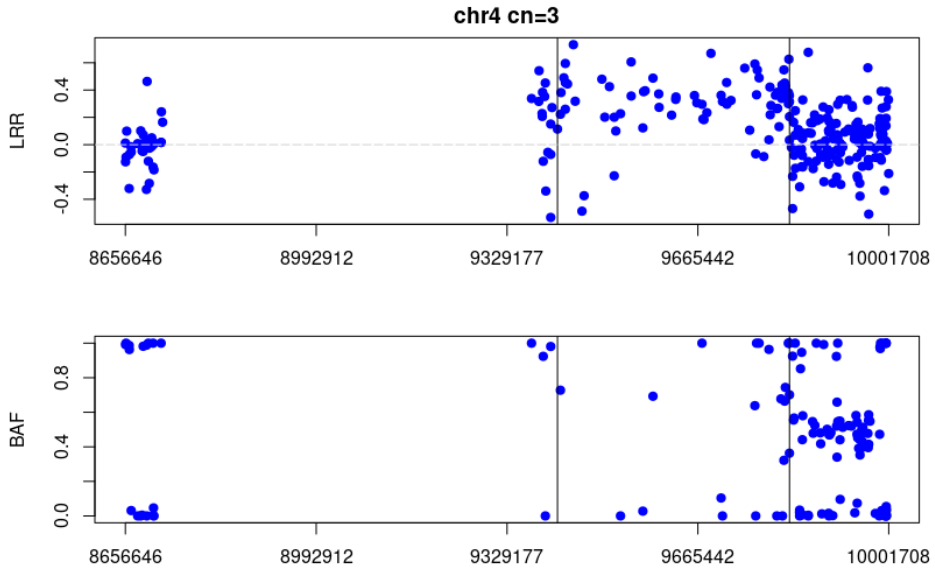
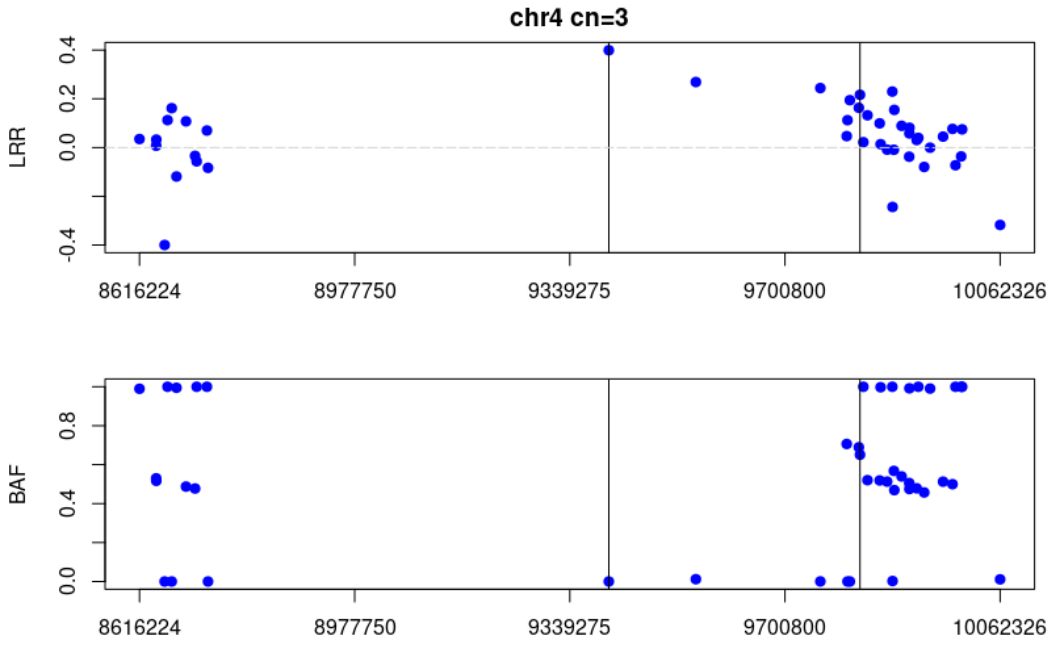


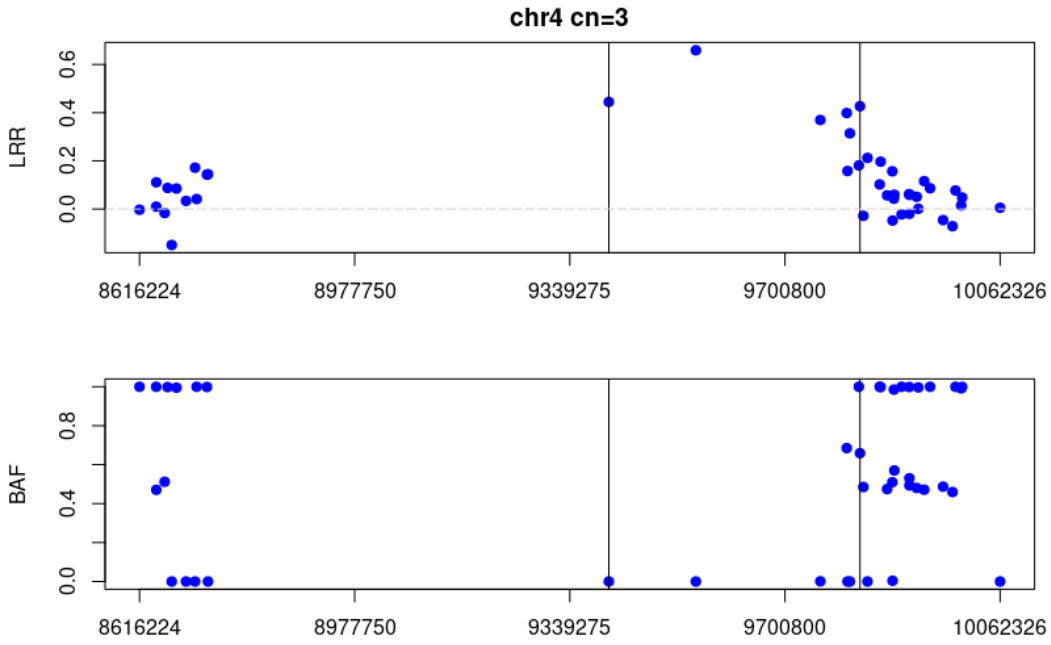
Figure 5. LRR and BAF traces from raw files from MGS study (hg19 genome version).

A) MGS_1; B) MGS_2. Black lines delineate duplication boundaries. Each dot represents the signal for either fluorescence intensity or LRR (top plot) of B allele frequency or BAF (bottom plot). Both the increase in fluorescence intensity and the transition from 3 possible states (AA/AB/BB) to 4 (AAA/AAB/ABB/BBB) of the B allele frequency confirms a duplication inside the boundaries.

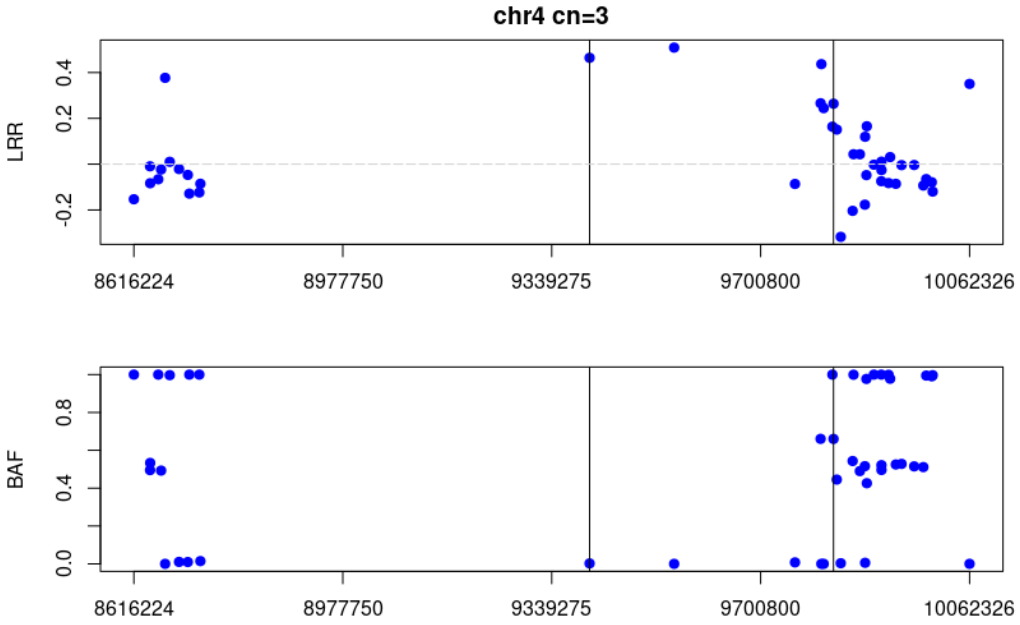
A


chr4 cn=3
LRR
0.4
0.2
0.0
-0.4
8616224
8977750
9339275
9700800
10062326
BAF
0.8
0.4
0.0
8616224
8977750
9339275
9700800
10062326

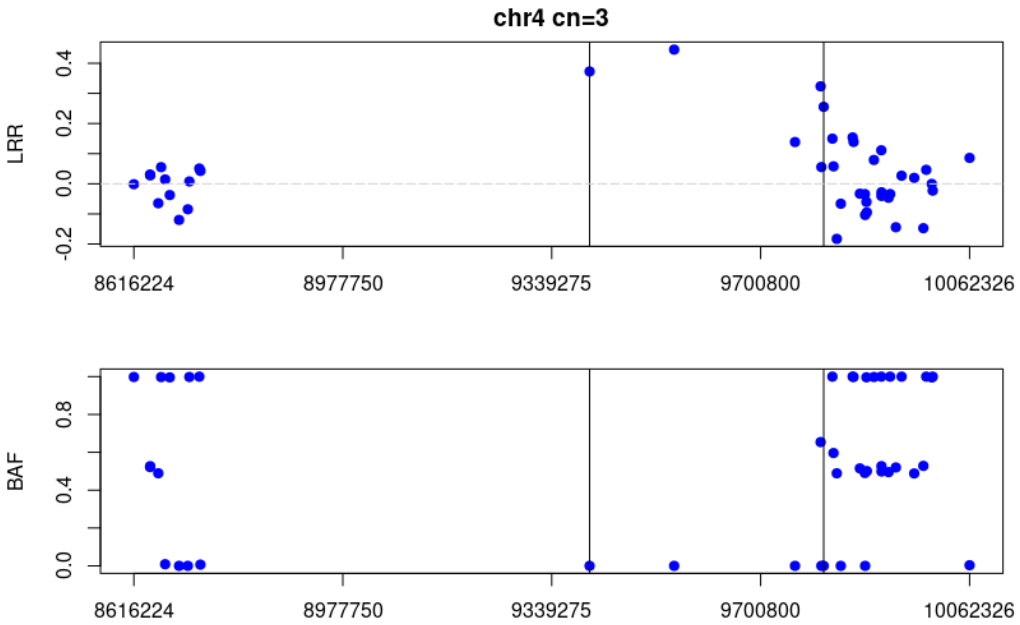
B


chr4 cn=3
LRR
0.6
0.4
0.2
0.0
8616224
8977750
9339275
9700800
10062326
BAF
0.8
0.4
0.0
8616224
8977750
9339275
9700800
10062326

C


chr4 cn=3
LRR
0.4
0.2
-0.2
8616224
8977750
9339275
9700800
10062326
BAF
0.8
0.4
0.0
8616224
8977750
9339275
9700800
10062326

D


chr4 cn=3
LRR
0.4
0.2
0.0
-0.2
8616224
8977750
9339275
9700800
10062326
BAF
0.8
0.4
0.0
8616224
8977750
9339275
9700800
10062326

E

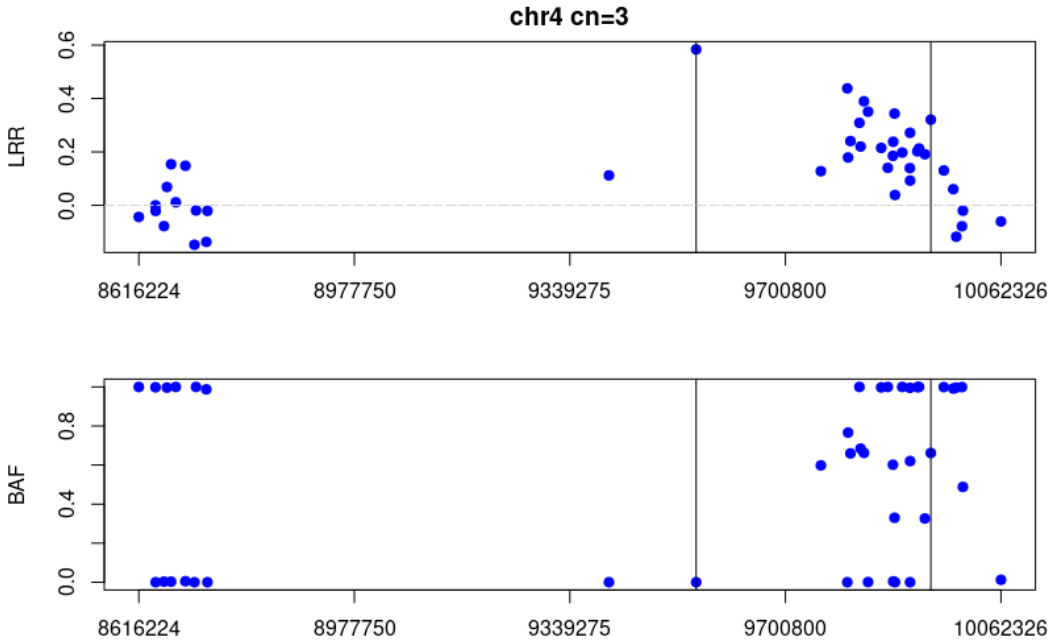


Figure 6. LRR and BAF traces from raw files from ClozUK study (hg19 genome version).

A) ClozUK_1; B) ClozUK_2; C) ClozUK_3; D) ClozUK_4; E) ClozUK_5. Black lines delineate duplication boundaries. Each dot represents the signal for either fluorescence intensity or LRR (top plot) of B allele frequency or BAF (bottom plot). Both the increase in fluorescence intensity and the transition from 3 possible states (AA/AB/BB) to 4 (AAA/AAB/ABB/BBB) of the B allele frequency confirms a duplication inside the boundaries.

2.4 Discussion

The results of this study revealed that an ultrarare (< 0.1%) 500 Kb recurrent duplication harbouring the gene *DRD5* is associated with an increased risk of schizophrenia. Of note, this duplication overlaps a segmental duplication by more than 50% of its length, a type of genomic region that had not previously been analyzed in relation to schizophrenia risk. This finding highlights the importance of examining less-studied regions of the genome to report novel genetic risk factors for complex disorders such as schizophrenia.

While other genes are affected by the different duplication events, such as *SLC2A9,* which has been associated with social phobia and anxiety (Lyngdoh et al., 2013), *DRD5* is the only gene in the locus that is affected by all the duplication events. *DRD5* gene encodes the G protein-coupled dopamine receptor D5 (Sunahara et al., 1991). Dopamine is a neurotransmitter that plays a crucial role in many brain functions (motivation, pleasure, reward), and constitutes the common pathway behind schizophrenia pathophysiology and treatment, especially for positive symptoms (Besnard et al., 2012; Howes and Kapur, 2009; Howes and Onwordi, 2023; Li et al., 2016; Sanyal and Van Tol, 1997). Five different dopamine receptors exist in humans (D1-D5), which can be classified between D2-like receptors (encoded by genes *DRD2, DRD3* and *DRD4*) and D1-like receptors (encoded by genes *DRD1* and *DRD5*) (Missale et al., 1998). One of the main differences between the two types is their ligand-binding affinity – D2-like receptors have a 10- to 100- fold increase in the affinity for dopamine compared to D1-like receptors. In addition, D1-like receptors are generally categorized as positive regulators, while D2-like receptors tend to have inhibitory roles, however, their positive and/or inhibitory effects are tuned by the formation of heterocomplexes

between each other and other receptors (Beaulieu et al., 2005; Hamid et al., 2016). Genetic case/control studies have implicated variants affecting *DRD1, DRD2, DRD3* and *DRD4* in not only schizophrenia susceptibility and differential treatment response, but also ADHD and bipolar disorder genetic risk (Cengiz et al., 2023; Hwang et al., 2012; Ishizuka et al., 2020; Kegeles et al., 2010; Lee et al., 2011; Liu et al., 2014; Talkowski et al., 2008; Zhang et al., 2015). In contrast, no robust evidence has been reported for *DRD5* through this type of analysis in any psychiatric disorder (Hwang et al., 2012; Lee et al., 2011; Zhao et al., 2014).

In a pedigree study that examined high-burden rare CNVs, the duplication in the 4p16.1 region that encompasses *DRD5* was identified in six individuals from one family of European ancestry, where four of them were diagnosed with schizophrenia or schizoaffective disorder. The remaining two individuals who carried the same duplication did not show any psychiatric disorders; however, their age at onset was still below the average for the family. To validate this variant, they used three different techniques due to its location in a segmental duplication (Van Den Bossche et al., 2013). The research performed in this chapter is thus the first investigation to report a significant association between the *DRD5*-containing 4p16.1 duplication and schizophrenia in a large-scale case/control study.

The primary constraints of this study originated from technical limitations in the probe intensity data of SNP arrays. CNVs that were smaller than the SNP array resolution were unavailable and many CNVs are expected to have been overlooked due to inadequate coverage of SNPs. As a result, the power to detect individual CNV loci that overlapped a segmental duplication by more than 50% of its length is insufficient,

which is why the total CNV burden was also investigated. However, pooling variants together to analyze CNV burden from segmental duplications did not result in more positive results.

Protective effects (OR < 1) were observed for variants affecting human-specific elements, especially deletions. While suggestive, the number of CNV events supporting this finding is very low. Protective ORs could be a consequence of the number of controls in this study being considerably higher (16,410 controls compared to 8,261 schizophrenia cases). Ideally, a balanced dataset would be used to validate this finding. Overall, the low SNP coverage limitation could be resolved with the development of bioinformatic tools that can detect variants in segmental duplications from whole-exome or whole-genome sequencing data. This is an encouraging area for future research, especially in challenging loci (Išerić et al., 2022; Prodanov and Bansal, 2020).

Another limitation of this research is that, although the *DRD5* duplication is present in both studies with different SNP coverages (MGS/ClozUK), it remains to be externally validated. Experimental work, such as Quantitative Polymerase Chain Reaction (qPCR) or Droplet Digital PCR (ddPCR) will be necessary to conclusively replicate this finding. After validation of the *DRD5*-containing 4p16.1 duplication, characterization of the gene expression effects should be examined, as another possibility is that the critical region in the duplication corresponds to a cis- or trans-acting non-coding regulatory region outside *DRD5.*

Heritability estimates were not calculated for the *DRD5* duplication due to its extremely low frequency. However, this discovery offers new opportunities for closing the heritability knowledge gap in schizophrenia, providing insights into difficult genomic

regions that may control the genetic mechanisms underlying this disorder. Overall, it is expected that other structural variants that are not captured by current genotyping platforms or bioinformatic analyses, including structural variants in the sex chromosomes, small CNVs, inversions, balanced chromosomal rearrangements and short tandem repeats are also an essential part of the solution to the missing heritability problem. To what extent each type of variant is important will only be addressable when these techniques are adequately developed and suitably powered datasets are generated (Brandler et al., 2016; Gymrek et al., 2016; Marshall et al., 2017; Sekar et al., 2016; Sudmant et al., 2013).

The methods used in this study could be extended to evaluate genetic susceptibility to additional neuropsychiatric diseases. The inclusion of variants identified across difficult genomic regions could pave the way for elucidating the complex genetic architecture of other psychiatric disorders, advancing our understanding of the underlying pathophysiology, and ultimately facilitating the development of targeted therapeutics.

2.5 Methods

2.5.1 Samples

Data from two different studies contributed to this CNV analysis. First, the Molecular Genetics of Schizophrenia (MGS) study, which was accessed through the dbGap repository upon request (https://www.ncbi.nlm.nih.gov/gap/, phs000167.v1.p1, phs000021.v3.p2), and provided 2,389 schizophrenia cases and 2,821 controls of European ancestry and 1,042 cases and 1,009 controls African American ancestry. Additionally, data from the ClozUK study was obtained through collaboration with the Division of Psychological Medicine and Clinical Neurosciences in Cardiff University, which stored the data locally. The ClozUK study included a total of 7,401 schizophrenia cases and 15,698 controls combined from different datasets (Table 4). European ancestry samples consisted in 6,112 schizophrenia cases and 14,347 controls.

A detailed description of clinical and phenotypic characteristics of the participants of different studies are available from previous publications (Manolio et al., 2007; Marshall et al., 2017; Rees et al., 2014c; Sanders et al., 2008; Suarez et al., 2006). In all studies, genotyping was performed on DNA obtained from blood samples. Samples of European ancestry in both studies comprised the "discovery sample". Independent datasets to constitute a "replication sample" were not available or had generated data that failed to be suitable for the analysis conducted in this study (e.g., large-scale exome SNP array platforms).

2.5.2 Exclusion of low-quality probes

SNP arrays contain thousands or millions of SNP probes that can be used to genotype samples by measuring the hybridization signal intensity for each SNP probe. Sequence

specificity of the probe refers to how well the probe sequence matches the intended target DNA sequence. Cross-hybridization can occur when a probe sequence has partial or complete homology with a non-target region of the genome, resulting in a hybridization signal that is not specific to the intended target DNA, which may lead to overestimation of CNV frequency and false positives. Therefore, probe sequence specificity constitutes an essential factor in SNP array platforms for detecting CNV events. It is believed that probe sequence specificity is systematically low in segmental duplications, which is why CNVs overlapping a segmental duplication by at least 50% of its length have been excluded in previous large-scale analyses (Marshall et al., 2017).

In this study, a BLAT-based alignment tool (*pblat*) is used to detect SNP probe off-targets in each of the dataset's arrays (Table 4). Probe array sequences were obtained from the Geo Platforms database (https://www.ncbi.nlm.nih.gov/geo). BLAT is suitable for this purpose since it is specifically optimized for searches of regions with at least 95% similarity over a stretch of at least 25 bases, which is the approximate length of a SNP probe (Script 1) (Kent, 2002).

(Script 1) pblat hg37.bit probe.fa -repMatch=2253 -stepSize=5 -minScore=20 -minIdentity=0 -threads=8 -out out.psl

A probe was deemed low quality and filtered out whenever more than one target was identified. The number of SNP probes before and after this quality control step in the ClozUK datasets is shown in Table 4. For the MGS dataset, both schizophrenia cases and controls were genotyped in Affymetrix Genome-Wide Human SNP Array 6.0 (Geo Accession GPL6801), which comprised 1,878,055 probes before filtering and 1,425,313 probes after filtering.

Table 4. Number of samples and SNP array probes on each ClozUK dataset.

	DATASET	N SAMPLES	ARRAY (GEO NUMBER)	N PROBES TOTAL	N UNIQUE PROBES	N COMMON UNIQUE PROBES
CASES	CLOZUK-1	2496	HumanOmniExpress-12v1 (GPL18900)	730,525	403,934	295,912 probes
	CLOZUK-2	989	HumanOmniExpress-12v1-1_B (GPL18900)	719,660	398,413	
	finalCogs	169	HumanOmniExpress-12v1-1_B (GPL18900)	719,665	398,417	
	CLOZUK_B2	3660	HumanOmniExpress-12v1-1_B (GPL18900)	951,117	392,524	
	CLOZUK_B4	87	HumanOmniExpress-12v1-1_B (GPL18900)	951,117	392,524	
CONTROLS	WTCCC2	5619	Illumina 1.2M (GPL6984)	1,238,733	706,255	
	Kora	1869	Illumina HumanOmni2.5 (GPL23136)	2,443,179	1,323,571	
	Melanoma	2598	Illumina HumanOmni1_Quad_v1-0-B (GPL8882)	951,117	430,618	
	Mammography	974	Illumina HumanOmni1_Quad_v1-0_B (GPL8882)	1,134,514	645,687	
	COPD_controls	998	Illumina HumanOmni1-Quad_v1-0_B (GPL8882)	1,134,514	645,687	
	CD_controls	3640	Illumina HumanOmni2.5-4v1-H (GPL23136)	2,443,177	1,323,570	

2.5.3 CNV detection

Log R Ratios (LRR) and B-allele frequencies (BAF) were generated for the MGS dataset using Affymetrix Analysis Power Tools (v2.11). All samples had high genotyping quality metric (> 0.86). For the ClozUK datasets, LRR and BAF data were obtained through collaboration with Cardiff University. PennCNV software (v1.0.5) was used to call CNVs using LRR and BAF data, following a standard protocol that accounted for GC content. DNA sequences with higher GC content tend to have higher hybridization signals than sequences with lower GC content. Therefore, if GC content is not accounted for, regions of the genome with high GC content may be incorrectly identified as having copy number gains, while regions with low GC content may be incorrectly identified as having copy number losses. This is particularly important in this study, whose aim is to identify rare CNVs associated with diseases, because even a few false positive or false negative CNV calls could lead to spurious associations with disease risk. To ensure that CNVs were called consistently across all microarrays used in the ClozUK datasets, only the consensus set of 295,912 common probes present in all microarrays were used (Table 4). Detected CNV events in an individual were merged if the distance between them was less than 50% of their combined length, using a custom R script. This step is necessary because, albeit rare, the HMM algorithm in PennCNV may artificially fragment large CNVs into adjacent smaller events. Due to possible aneuploidy, samples were excluded if more than 10% of any chromosome was copy number variable.

2.5.3.1 CNV quality control

CNVs were subjected to further filtering using PennCNV. Specifically, CNVs that were less than 10 kb in length, covered by less than 5 probes in the ClozUK dataset and less than 10 probes in the MGS dataset, or overlapped with telomeres by more than 50% of their length or with immunoglobulin genes and centromeres by more than 10% of their length were excluded. Annotations of telomeres, centromeres and immunoglobulin regions were supplied by PennCNV. Finally, CNV loci that had a frequency greater than 1% were excluded using PLINK (v1.07) to assess rare events, and only CNVs that overlapped more than 50% of a segmental duplication were kept for further analyses (Table 5).

2.5.3.2 Sample quality control

Samples that exhibited significant deviations in LRR standard deviation, BAF drift, wave factor, or total number of CNVs were excluded from the analysis. Log R Ratio represents the log-transformed intensity ratio between a DNA sample and its baseline signal. LRR standard deviation is calculated from values across all probes in a sample. High LRR standard deviation indicates that signal intensity values are too variable. BAF drift refers to the deviation of BAF values from the expected 0.5 value for heterozygous SNPs. A high level of BAF drift indicates that the BAF values are unreliable, and that genotyping data in the sample may not be accurate. High LRR standard deviation or high BAF drift happen due to factors such as poor sample quality, sample processing issues, or technical noise in the data. The wave factor is calculated as the standard deviation of the second derivative of the LRR values across the genome (in other words, it represents how much the signal oscillates around the mean). A high

wave factor indicates that the LRR values are affected by systematic biases, such as GC content.

Samples with high LRR standard deviation, high BAF drift or high wave factor are typically excluded from CNV calling analysis because they are more likely to produce false positive or false negative CNV calls. Additionally, samples that exhibited a high number of CNV events relative to other samples in the dataset were also excluded, since it indicates poor quality data or technical issues that could affect CNV calling accuracy such as DNA fragmentation. Values were considered outliers when they surpassed the mean plus 3 standard deviations (Figure 7, Figure 8)

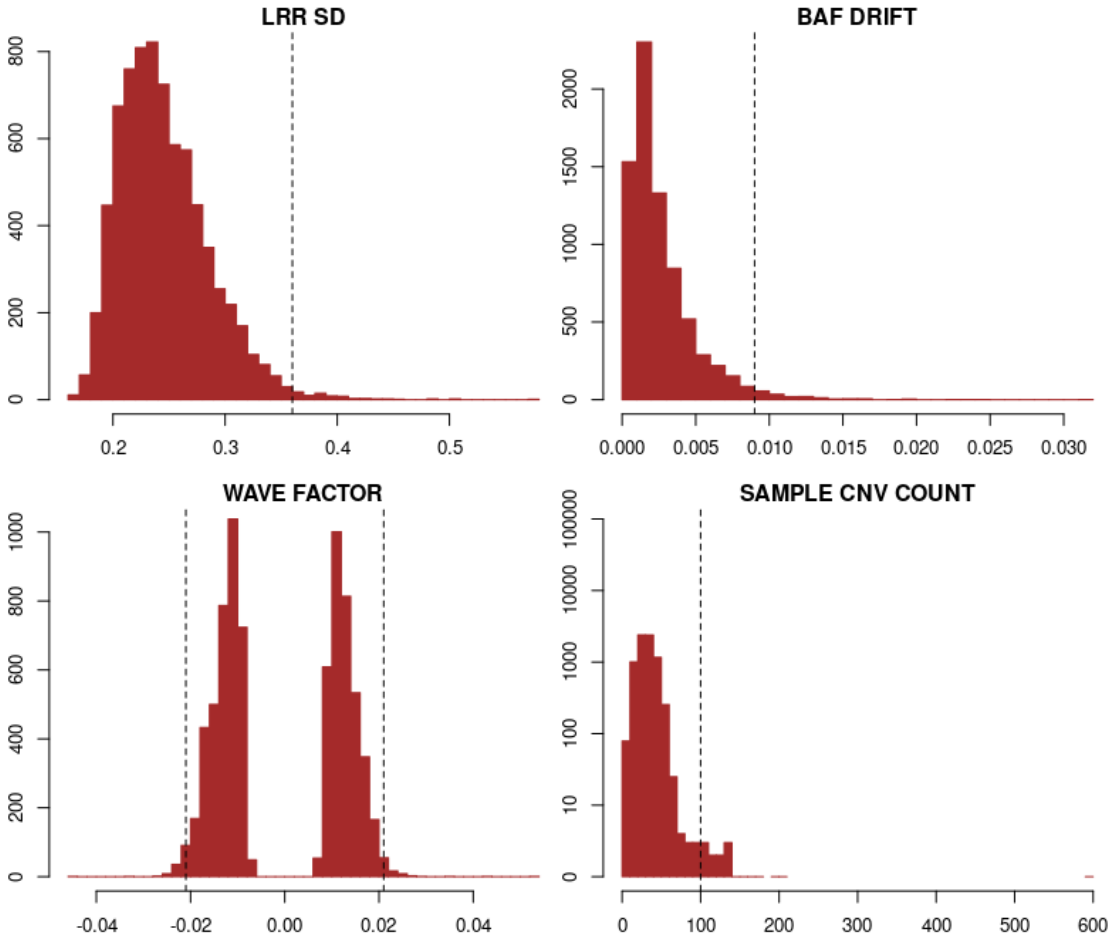


Figure 7. Samples quality control histograms for the MGS study.

Dashed lines indicate outlier thresholds. Samples beyond these values were excluded. Sample CNV count histogram y axis is in logarithmic scale for better visualization.

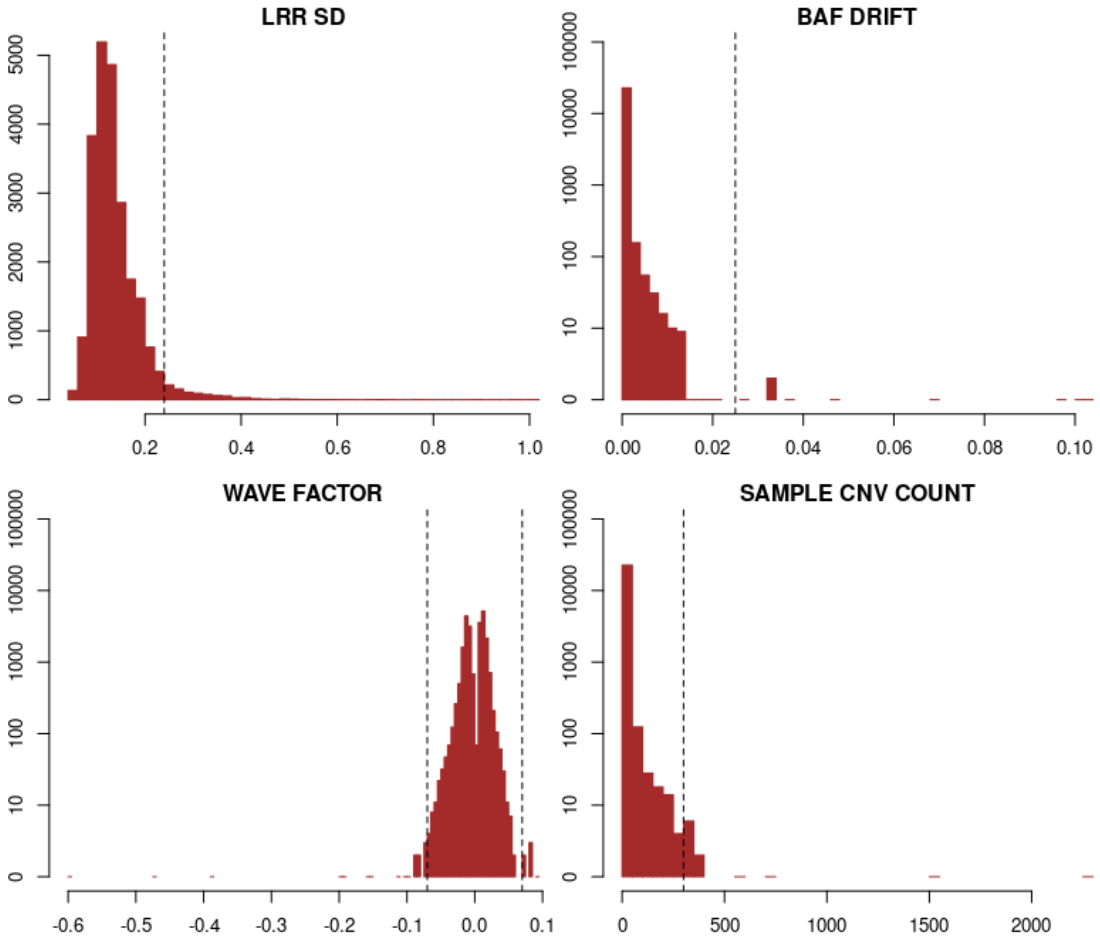


Figure 8. Samples quality control histograms for the ClozUK study.

Dashed lines indicate outlier thresholds. Samples beyond these values were excluded. BAF drift, wave factor and sample CNV count histogram y axes are in logarithmic scale for better visualization.

The numbers of samples that passed quality control in MGS were 2,200 cases and 2,571 controls, whereas in the ClozUK study were 6,061 cases and 13,839 controls. All genomic coordinates reported in this study refer to the UCSC build 37, hg19.

2.5.4 Gene-wise analysis

In order to identify new risk loci, a gene-based approach was employed in this study. Using GENCODE gene coordinates (v42lift37, Table 5), each coding gene in the genome was assessed for exon-disrupting CNVs. Deletions and duplications were analyzed separately with PLINK. Initially, genes were evaluated individually in each study (MGS/ClozUK) using logistic regression with sex and CNV length as covariates. To increase statistical power, the studies were later combined, and logistic regression was performed again, this time incorporating the previous covariates as well as the study identifier (Equation 1).

(**Equation 1**) *Phenotype (1/0) ~ CNVs disrupting the gene (Absolut count)* + Sex (F/M) + Length (KB) + [Dataset (MGS/ClozUK), only when combined dataset was tested]

The nominal significance threshold was established at $P < 0.05$. To correct for multiple testing, Bonferroni correction was applied, with the nominal P value divided by the number of possible genes affected. However, not all genes were considered in this correction, as CNVs were restricted to segmental duplications and the potential number of genes affected was thus limited to segmental duplications and their surrounding genes (n = 2644 genes, $P < 1.89 \times 10^{-05}$). This gene-centric approach allowed for detection of signal from both single gene enrichments and larger recurrent events via contiguous gene enrichments.

Notably, logistic regressions were performed with the *logistf* function in R, since it is specifically designed for logistic regression with small to moderate sized datasets, particularly those with rare events.

2.5.5 CNV burden analyses

The burden of CNVs per individual was assessed through three different methods: (i) calculation of the number of CNVs per individual, (ii) quantification of the kilobase burden of CNVs, (iii) determination of the number of genes impacted by CNVs. Only CNVs that overlapped with coding exons were considered for gene counting. Individual burden was annotated with PLINK, and logistic regressions were conducted in R utilizing the *glm* function of the *stats* package (pooling variants increases number of events, which enables the use of standard methods of logistic regression). The number of CNVs was only adjusted for sex (Equation 2), while sex and the number of CNVs were considered as covariates for the total length of base pairs affected by CNVs (Equation 3). Finally, sex, the number of CNVs, and CNV length were considered as covariates for the total number of genes impacted (Equation 4).

(Equation 2) *Phenotype (1/0) ~ N CNVs (Absolut count)* + Sex (F/M) + [Dataset (MGS/ClozUK), only when combined dataset was tested]

(Equation 3) *Phenotype (1/0) ~ Length (KB)* + Sex (F/M) + N CNVs (Absolut count) + [Dataset (MGS/ClozUK), only when combined dataset was tested]

(Equation 4) *Phenotype (1/0) ~ N genes affected (Absolut count)* + Sex (F/M) + N CNVs (Absolut count) + Length (KB) + [Dataset (MGS/ClozUK), only when combined dataset was tested]

This approach allowed for the independent analysis of each type of burden, since it is expected that a higher number of CNVs results in longer segments of DNA affected by CNVs, and a higher number of genes disrupted too.

2.5.5.1 Genomic element enrichment

Genomic element sets that have been previously associated with brain development and/or schizophrenia and are enriched in segmental duplications were compiled from previous studies (Table 5). To test for an association between schizophrenia CNVs and disrupted genomic element sets, logistic regression was performed using the *logistf* package in R (again, because the number of elements disrupted was low, Table 2, Table 3). Sex, number of CNVs, and CNV length were included as covariates in the regression analysis (Equation 5). Additionally, for the gene set, the logistic regression was further adjusted for the global number of genes affected by CNVs to account for any non-specific association signals that may be reflective of CNV burden stratification between cases and controls (Equation 6).

(Equation 5) *Phenotype (1/0) ~ N elements affected (Absolut count)* + Sex (F/M) + N CNVs (Absolut count) + Length (KB) + [Dataset (MGS/ClozUK), only when combined dataset was tested]

(Equation 6) *Phenotype (1/0) ~ N genes from gene-set affected (Absolut count)* + Sex (F/M) + N global genes affected (Absolut count) + N CNVs (Absolut count) + Length (KB) + [Dataset (MGS/ClozUK), only when combined dataset was tested]

2.5.6 Validation

The *DRD5* duplication was validated through a meticulous inspection of the boundaries in the UCSC Genome Browser. Custom tracks were created from PennCNV output files for duplications and deletions to facilitate this process (custom R scripts). Additionally, the raw LRR and BAF traces were plotted using the *ggplot* package in R to ensure the accuracy of the CNV calls.

Table 5. Genomic annotations used in non-coding constraint study.

SET	DESCRIPTION	N	DOWNOADED FROM	REFERENCES
Segmental duplications	Summary of large genomic Duplications (>1Kb >90% similar)	51,599 non-independent regions	hg19/genomicSuperDups UCSC Genome Browser Track	(Bailey et al., 2001; Bailey and Eichler, 2006)
Global	Basic Gene Annotation (reference chromosomes only)	62,703 genes	GENCODE v42lift37	https:// www.gencodegenes.org/
Great-ape specific brain genes	Great-ape specific genes preferentially expressed in progenitors of fetal neocortex	50 genes	Supplementary material from reference	(Florio et al., 2018)
HAQERs	Human ancestor quickly evolved regions overlapping open chromatin windows in the developing brain	79 regions	https:// www.vertgenlab.org/ haqer2022.html	(Mangan et al., 2022)

Chapter 3: Contribution of constrained non-coding regions in copy number variants to schizophrenia genetic risk

3.1 Summary

Common and rare genetic variants reported in psychiatric disorders, especially schizophrenia, are concentrated in biologically essential genes with high constraint against loss-of-function mutations. Copy number variants (CNVs) in exclusively non-coding regions were assumed to be unimportant. A new constraint score that includes non-coding regions allows for further investigation of this aspect. CNV calls from three different studies were meta-analyzed, comprising 14,155 schizophrenia cases and 21,660 controls. Constraint Z score in 1kb genomic windows was used to annotate CNV constraint, and to then evaluate the contribution of constraint from coding and non-coding disrupted regions. CNVs associated with schizophrenia mapped to regions with increased non-coding constraint compared to CNVs in controls. The non-coding constraint signal in genic CNVs covaried with coding constraint, whereas non-genic CNVs showed higher non-coding constraint driven by candidate cis-regulatory elements (promoters and enhancers) in schizophrenia. These findings suggest that CNVs associated with schizophrenia are likely to occur in regions that are functionally important for gene regulation, even if they do not directly affect coding regions. This study highlights the importance of considering non-coding regions in disease genetics to improve our understanding of the functional impact and pathogenicity of CNVs in schizophrenia, which may lead to new strategies for risk prediction and targeted therapy.

3.2 Introduction

3.2.1 Coding and non-coding constraint

Genomic mutational constraint refers to the selection pressure that acts to prevent or reduce the accumulation of mutations in specific regions of the genome. These regions are usually essential for the proper functioning of biological processes or for maintaining the structural integrity of the genome. Therefore, mutations that occur in constrained regions are more likely to be deleterious and have a negative impact on fitness, leading to their removal from the population over time (Karczewski et al., 2020; Pollard et al., 2010).

There are several ways to quantify constraint estimates. Population genetic-based methods quantify the observed variation in a genomic region and compare it to the expected mutation rate, to determine whether the region is subject to purifying selection in the human lineage (Chen and The gnomAD consortium, 2022; di Iulio et al., 2018; Gussow et al., 2017; Halldorsson et al., 2022; Karczewski et al., 2020; Lek et al., 2016; Vitsios et al., 2021). A study from the Exome Aggregation Consortium (ExAC) of protein-coding variation from over 60,000 individuals developed a coding-exome-wide metric that estimated the probability of a gene being loss-of-function (LoF) intolerant (pLI). This metric originated the concept of constrained genes ($pLI \geq 0.9$, $n = 3{,}230$), which are characterized as being involved in core biological processes and widely expressed across tissues (Lek et al., 2016). This study preceded a larger analysis of over 125,000 whole-exome and 15,000 whole-genome sequences, as part of the Genome Aggregation Database (gnomAD). In contrast to the dichotomous use of pLI, the bigger sample size allowed for the generation of a continuous metric that reflected the degree

of intolerance to LoF variation in each gene, termed loss-of-function observed/expected upper bound fraction (LOEUF) (Karczewski et al., 2020).

Metrics that quantify gene constraint have greatly facilitated the interpretation of pathogenicity of genetic variants associated with disease. Rare variants (including SNVs, CNVs and *de novo* variants), which are more frequently deleterious, are enriched in constrained genes (pLI > 0.9) in traits with lower reproductive fitness such as schizophrenia, ASD and developmental disorders (Havdahl et al., n.d.; Karczewski et al., 2020; Singh et al., 2017). In schizophrenia, the enrichment in constrained genes persists in common variant associations and accounts for 30% of the SNP heritability (Pardiñas et al., 2018). Equivalent findings were observed when using the continuous LOEUF variable. Therefore, it is believed that LoF intolerant genes play a significant role in the genetic architecture of cognitive and psychiatric traits, especially in schizophrenia (Karczewski et al., 2020).

Other metrics of human population genetic-based constraint extend to the whole genome, allowing the measure of constraint in non-coding regions. These methods quantify the depletion of variation in artificial genomic windows of arbitrary sizes. A number of different metrics have been developed in recent years. The Orion metric was generated with data from over 1,500 whole-genome sequences, utilized windows ranging from 10 to 1,000 bp and integrated effects of trinucleotide context in the analysis (Gussow et al., 2017). Using a different strategy that relied on 7-nt motifs rather than genomic windows, the context-dependent tolerance score (CDTS) employed whole-genome sequence data from more than 11,000 individuals (di Iulio et al., 2018). Subsequent efforts raised the number of participants considerably. The genome-wide

residual variation intolerance score (gwRVIS) was constructed upon whole-genome sequencing data from over 60,000 individuals and was based on a non-overlapping 3-nt window method (Vitsios et al., 2021). Additionally, the UK Biobank released a Depletion Rank (DR) score, computed from the analysis of over 150,000 whole-genome sequences, that integrated previous methodological approaches with C>G *de novo* enriched region information from a family trio study (Halldorsson et al., 2022).

All these metrics (Orion, CDTS, gwRVIS, DR) showed association with GWAS and ClinVar catalog variants, highlighting their ability to prioritize non-coding variants. However, they failed to account for the effect of regional genomic features such as centromeres and telomeres on the mutation rate beyond the influence of trinucleotide context (di Iulio et al., 2018; Gussow et al., 2017; Halldorsson et al., 2022; Vitsios et al., 2021). To solve this issue, the Genome Aggregation Database (gnomAD), which accumulated whole-genome sequence data from over 70,000 individuals, adjusted the expected mutation rate of each trinucleotide context depending on the genomic feature, derived from the frequency of *de novo* mutations that these genomic features exhibited in previous family-based whole-genome sequencing studies. The gnomAD analysis computed a constraint Z score from non-overlapping 1,000 bp windows, which out-performed all previous constraint metrics in predicting non-coding variants associated with disease (Chen and The gnomAD consortium, 2022).

Putative functional non-coding regions (introns, cCREs, and non-coding genes) exhibit higher mutational constraint compared to genomic regions with no annotated function, as shown in the CDTS, DR and Z score studies (Chen and The gnomAD consortium, 2022; di Iulio et al., 2018; Halldorsson et al., 2022). Consequently, it is expected that

dosage alteration of constrained non-coding functional regions by CNVs contributes to the genetic risk of cognitive and psychiatric traits, in parallel with LoF-intolerant genes. Accordingly, non-coding Z score constraint presented a substantial increase in CNVs associated with developmental disorders compared to CNVs in controls. Notably, the enrichment survived correction for gene content and gene constraint, indicating that the dosage of constrained non-coding regions may underpin the pathogenicity of CNVs associated with developmental disorders (Chen and The gnomAD consortium, 2022).

Overall, constraint from genetic variation is an important concept in understanding the function and pathogenicity of variants in diverse genomic regions, as it aids in identifying regions that are under strong selection pressure to maintain their integrity. This is of utmost importance in psychiatric disorders, in which variants affecting constrained regions are enriched (Havdahl et al., n.d.; Karczewski et al., 2020; Pardiñas et al., 2018; Singh et al., 2017). The enrichment of variants of all frequencies in constrained genes in schizophrenia underlines their involvement in the development of the disorder. Therefore, the development of tools that classify CNVs according to the constraint of genes and other functional elements they disrupt would be a valuable addition to the study of schizophrenia and other disorders. In this chapter, the contribution of constraint from non-coding regions is explored in CNVs in schizophrenia.

3.2.2 Study design

In this chapter, the contribution of non-coding constrained regions to schizophrenia genetic liability is explored for rare CNVs using data from 14,155 schizophrenia cases and 21,660 controls. The same CNV calls that had been used in a previous study showed

that LoF-intolerant genes are enriched in schizophrenia CNVs (ClozUK1 with 6,307 cases and 10,675 controls and ClozUK2 with 5,648 cases and 8,414 controls) (Rees et al., 2014c; Singh et al., 2017). Here, the ClozUK1 and CLozUK2 datasets were combined with the MGS dataset (2,200 cases and 2,571 controls), for which CNV calls were made as described in chapter 2 of this thesis.

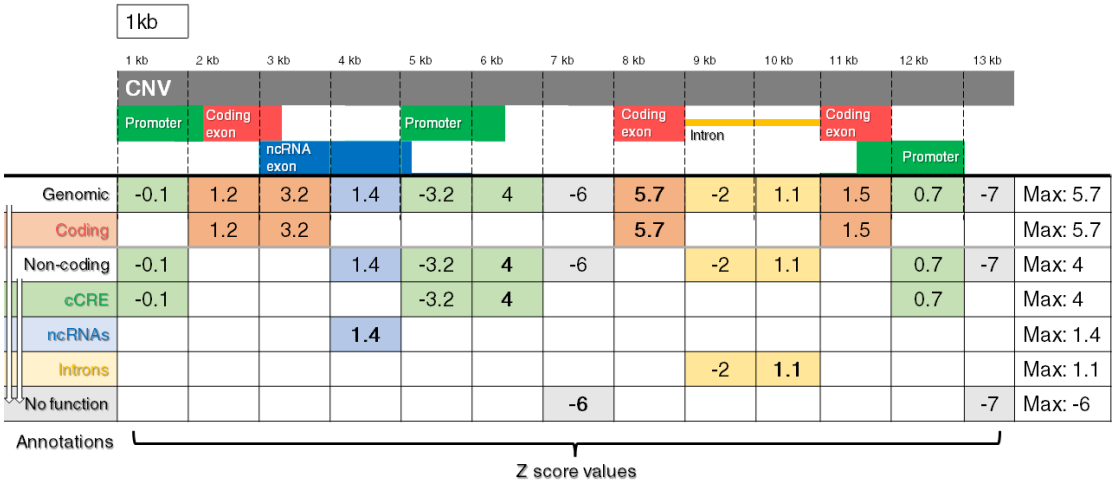
The constraint metrics used here differ from a previous study that evaluated the enrichment of disrupted constrained genes (established at a threshold of pLI > 0.9 for LoF-intolerance) in schizophrenia CNVs. In this chapter, the focus is on non-coding constraint (continuous variable based on 1 kb windows across the genome) (Chen and The gnomAD consortium, 2022; Lek et al., 2016). Previously, different individuals were annotated with the number of LoF-intolerant genes (pLI > 0.9) that were encompassed by CNVs. In that study, the probability of an individual having schizophrenia was tested against the number of LoF-intolerant genes disrupted (Singh et al., 2017). Because there is no non-arbitrary cut-off for variation intolerance in the Z score constraint windows, a different strategy was undertaken in this chapter. Here, CNVs were annotated with the maximum Z score of the disrupted 1kb windows, and the probability of a CNV occurring in an individual that has schizophrenia was tested against this score (Figure 9).

The 1kb defined windows could potentially overlap any type of genomic annotation, including coding genes (Chen and The gnomAD consortium, 2022). To test non-coding constraint exclusively, any Z score value from a 1kb window that spanned 1 bp or more of a coding exon was removed, and CNVs were assigned a constraint with maximum Z scores and tested for association with schizophrenia. This strategy is repeated

throughout the chapter to examine the contribution of different functional non-coding regions (cCREs, ncRNAs and introns) (Figure 9).

The research reported in this chapter is exploratory. The goal of an exploratory analysis is to identify patterns, trends, outliers, and relationships within the data, and to generate hypotheses for further investigation. Multiple tests are performed to formulate more targeted research questions or hypotheses for subsequent analyses. While this increases the likelihood of type I error in statistics (significant results purely by chance), corrections for multiple testing are typically not done in exploratory analyses, since they serve as a foundation for hypothesis generation followed by more rigorous statistical analyses (Grove and Andreasen, 1982).

The development of a framework for interpreting non-coding region variants will facilitate the discovery of new types of disease-causing variants, expanding our knowledge of additional contributors to disease genetic risk. In this chapter, the relevance of non-coding regions is studied through the Z score constraint metric, which provides insights into the degree of selective pressure acting upon them, and thus, their biological relevance. By evaluating the effects of genes and functional non-coding annotations (intron, cCREs, and ncRNAs) on the constraint of CNVs in schizophrenia patients, it is possible to discern which specific non-coding regions play a more critical role in the etiology of the disorder.



Test: CNV (Phenotype 1/0) ~ Annotation constraint (Maximum Z score) + Covariates

Figure 9. Analysis strategy: example of the assignment of constraint to CNVs conditioning on the genomic annotation.

Non-coding annotations include cCREs, ncRNAs, introns, and regions with no annotated function. Genomic annotation includes everything else. In cases where a window overlaps a coding exon and a cCRE, the window was categorized as coding (see hierarchy established in Table 6).

3.3 Results

3.3.1 CNVs in individuals with schizophrenia exhibit higher non-coding constraint than CNVs in controls

CNVs implicated in developmental delay have been shown to harbour increased Z score constraint compared to controls (Chen and The gnomAD consortium, 2022). To test whether this characteristic is shared with schizophrenia, CNVs from the three different studies (ClozUK1, Clozuk2 and MGS) were annotated with the maximum Z score constraint of disrupted 1kb windows. The probability of a CNV belonging to an individual that has schizophrenia was then tested against the maximum Z score of disrupted 1kb windows with a logistic regression (Figure 9). Effects of CNV length and other covariates were accounted for (See 3.5 Methods, page 113). Meta-analysis of the three different studies showed that schizophrenia CNVs displayed association with increased genomic Z score constraint (OR=1.08, $P < 2.53 \times 10^{-04}$; Figure 10A; Supplementary Table 10). The signal was consistent in all studies: MGS (OR=1.16, $P < 4.32 \times 10^{-05}$), ClozUK1 (OR=1.07, $P < 5.55 \times 10^{-02}$) and ClozUK2 (OR=1.04, $P < 1.79 \times 10^{-01}$; Figure 10A; Supplementary Table 10). Deletions contributed substantially to the association (meta-analysis OR=1.14, $P < 2.08 \times 10^{-04}$) in comparison to duplications (meta-analysis OR=1.05, $P < 9.70 \times 10^{-02}$; Figure 10A; Supplementary Table 10).

To test the contribution of the non-coding regions exclusively, Z score values from 1kb windows that overlapped at least 1bp of a coding exon were removed. The number of CNVs removed due to this filtering is reflected in the numbers of CNVs contributing to the regression in Figure 10B. CNVs were annotated again with the highest Z score from non-coding 1kb windows (hereafter referred to as non-coding constraint; 3.2

Introduction, Figure 9; 3.5 Methods; Table 6). Meta-analysis showed that non-coding constraint in the CNV is associated with schizophrenia (OR=1.05, $P < 4.53 \times 10^{-02}$; Supplementary Table 10). This observation was driven by deletions (OR=1.13, $P < 1.39 \times 10^{-03}$; Figure 10B; Supplementary Table 10) while duplications displayed no association (OR=0.99, $P < 7.29 \times 10^{-01}$; Supplementary Table 10).

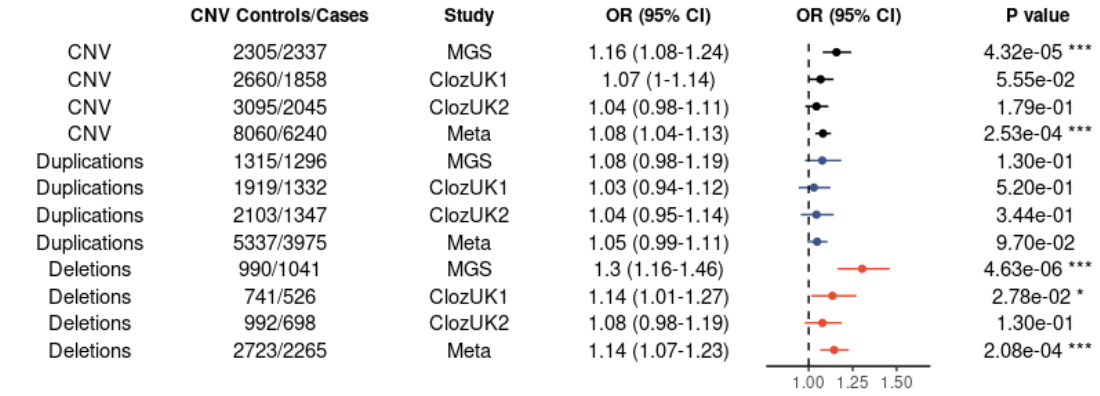
To evaluate different functional genomic elements separately, the Z score constraint of CNVs was annotated based on, firstly, 1kb windows overlapping coding exons. Any Z score from disrupted 1kb windows that overlapped a coding exon but also a non-coding exon, intron, or cCRE was considered in the coding exon group and excluded in the rest of the categories (3.2 Introduction, Figure 9; 3.5 Methods; Table 6). The maximum of these Z scores in the CNV is referred to as coding constraint in this chapter. Following the same pattern as the genomic constraint, coding constraint was increased in schizophrenia CNVs compared with controls, especially in deletions (OR=1.20, $P < 2.68 \times 10^{-05}$; Figure 10B; Supplementary Table 10). This discovery supports previous studies that reported an enrichment of constrained genes in deleterious rare variants in schizophrenia (Singh et al., 2017).

Additionally, the contribution of different functional non-coding regions was assessed (3.2 Introduction, Figure 9; 3.5 Methods; Table 6). The cCRE constraint, represented by the maximum Z score constraint from windows in the CNV that overlapped cCREs in non-coding regions (which could span non-coding exons and introns), showed association with schizophrenia (deletions OR=1.15, $P < 5.46 \times 10^{-05}$; Figure 10B; Supplementary Table 10). Separately, non-coding exon constraint, (from non-coding exons outside cCREs), was higher in schizophrenia (deletions OR=1.13, $P < 5.56 \times$

10^{-03}; Figure 10B; Supplementary Table 10). It is important to note that non-coding exon Z score windows could extend through introns of coding genes. When exclusively intronic Z score windows were used to annotate the maximum constraint in the CNV, the significant association held (deletions OR=1.13, $P < 2.70 \times 10^{-03}$; Figure 10B; Supplementary Table 10).

The association was further tested after removing Z scores coming from any genic region (coding exons, non-coding exons, and introns) and any cCRE (promoters, enhancers, CTCF-bound elements), in other words, excluding all coding and non-coding disrupted 1 kb windows with an annotated function (3.2 Introduction, Figure 9; 3.5 Methods; Table 6). In this case, while the number of CNVs included in the analysis was not substantially reduced, the association lost statistical significance (deletions OR=1.12, $P < 7.12 \times 10^{-02}$; Figure 10B; Supplementary Table 10), indicating that the signal is predominantly coming from functional non-coding regions (non-coding RNA genes, introns and/or cCRE regions).

A Genomic constraint



B Constraint from different annotations

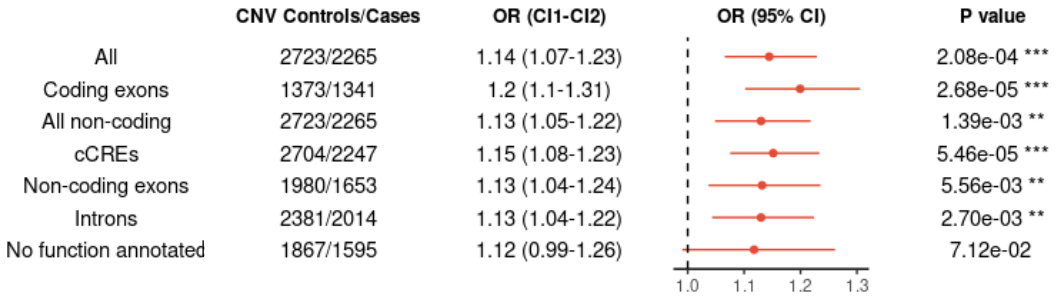


Figure 10. Primary findings. Z score constraint is increased in CNVs in schizophrenia cases compared to CNVs in controls.

A) Forest plot for the association of the maximum Z score of 1 kb windows disrupted by the CNV and schizophrenia case/control status across studies (MGS, ClozUK1, ClozUK2 and meta-analysis) and type of CNV (CNV coloured in black, duplications in blue and deletions in red). B) Forest plot for the meta-analysis of the association with schizophrenia case/control status of the Z score constraint in deletions only (All = maximum Z score from all disrupted windows; coding exons = maximum Z score from coding exons; all non-coding = maximum Z score from windows overlapping any non-coding region; cCREs = maximum Z score from windows overlapping cCREs in exclusively non-coding regions; non-coding exons = maximum Z score from windows overlapping non-coding exons; introns = maximum Z score from windows overlapping introns and lastly; no function annotated, maximum Z score from windows that do not overlap any functional annotation).

3.3.2 Sensitivity analyses

3.3.2.1 CNVs previously implicated in schizophrenia

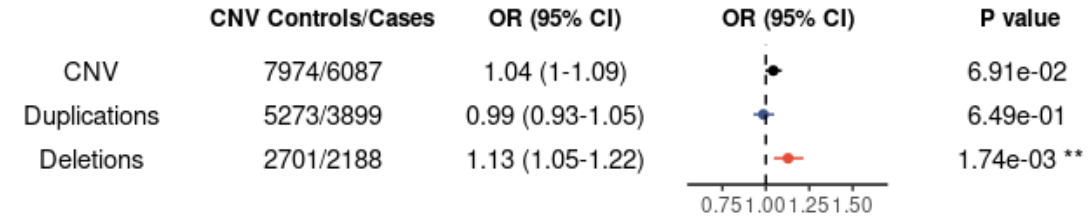
Much of the CNV burden, measured by the number of genes affected, is driven by CNV loci previously associated with schizophrenia (Marshall et al., 2017). To evaluate whether the non-coding signal is concentrated in these loci also, the analysis was repeated following removal of CNVs that overlap regions listed in Table 1 (Methods). Comparable effect sizes and significance levels were observed to the analysis that included previously implicated CNVs (deletions including reported loci meta-analysis OR=1.13, $P < 1.39 \times 10^{-03}$; Figure 10A; Supplementary Table 10; versus deletions excluding reported loci meta-analysis OR=1.13, $P < 1.74 \times 10^{-03}$); Figure 11A; Supplementary Table 11). This suggests that the non-coding constraint signal is coming predominantly from loci that have not been yet reported.

3.3.2.2 LoF intolerant genes (pLI < 0.9)

Rare deletions affecting LoF-intolerant genes are enriched in people with schizophrenia, while other deletions seemingly display the same frequency in people with and without schizophrenia (Singh et al., 2017). Non-coding regions surrounding LoF-intolerant genes should exhibit higher constraint against variation due to linkage disequilibrium. To check that the increased non-coding constraint in deletions in schizophrenia was not a consequence of the presence of LoF intolerant genes in the same variant, the number of LoF-intolerant genes disrupted was added as a covariate in the analysis (3.5 Methods). Non-coding Z score constraint in schizophrenia CNVs remained significantly increased (deletions meta-analysis OR = 1.10, $P = 8.80 \times 10^{-3}$); Figure 11B; Supplementary Table 12). Different functional annotations were tested

separately (cCRE constraint, non-coding exon constraint, intronic constraint), and major findings remained significant (Supplementary Table 12). Remarkably, coding Z score constraint retained a significant association with schizophrenia even after correcting for the number of LoF-intolerant genes disrupted (pLI < 0.9), which suggests pLI fails to cover the whole spectrum of variation in coding genes that has a substantial phenotypic effect (Supplementary Table 12).

A Previously implicated regions removed



B Correcting for LoF intolerance

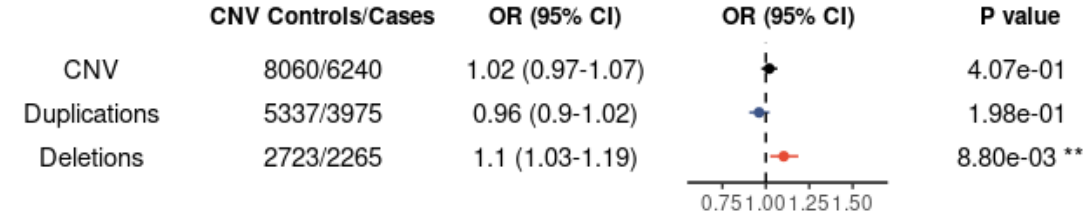


Figure 11. Sensitivity analyses of non-coding constraint.

A) Correction for previously implicated CNV loci in schizophrenia. B) Correction for the number of LoF-intolerant genes (pLI < 0.9) in the CNV. Forest plots show the meta-analysis of the association of the maximum non-coding Z score in the CNV and schizophrenia case/control status across types of CNV (all CNVs coloured in black, duplications in blue and deletions in red).

3.3.3 Constraint in genic and non-genic CNVs

Previous studies have shown that there is no increase in burden, measured by the length of the CNV, from CNVs that do not overlap a protein-coding gene (referred to as non-genic CNVs), in schizophrenia (Marshall et al., 2017). However, it could be possible that, albeit not longer, non-genic CNVs in schizophrenia disrupt regions with higher non-coding constraint compared to controls. To evaluate this, CNVs were separated into genic and non-genic, depending on whether they spanned at least one base pair of a protein-coding exon. Meta-analysis showed that non-coding Z score constraint enrichment in schizophrenia was only seen in genic deletions (meta-analysis OR=1.22, $P < 9.72 \times 10^{-04}$; Figure 12A, Supplementary Table 13). Meanwhile, non-coding constraint was not significantly increased in non-genic CNVs in schizophrenia compared with controls (deletions meta-analysis OR=0.95, $P < 3.39 \times 10^{-01}$; Figure 12A, Supplementary Table 14). It was noted that non-genic CNVs were substantially smaller than genic CNVs (Supplementary Figure 1). Since CNV calls had been filtered by size and number of SNP probes to reduce noisy signals in the primary analysis (> 50,000 bp and > 50 probes), non-genic CNVs whose Z score constraint contributed to pathogenicity could have been filtered out. Consequently, the analysis was repeated including smaller CNVs (> 10,000 bp but less than 50,000 bp). To ensure that quality CNV calls were prioritized, only CNVs that had more than 10 SNP probes and a probe density of more than 1 probe for every 5,000 were considered. All previous tests maintained the same trends except non-genic CNVs (Figure 12A-B; Supplementary Table 10; Supplementary Table 11; Supplementary Table 12; Supplementary Table 13; Supplementary Table 14; Supplementary Table 15; Supplementary Table 16; Supplementary Table 17; Supplementary Table 18). In non-genic CNVs, higher non-

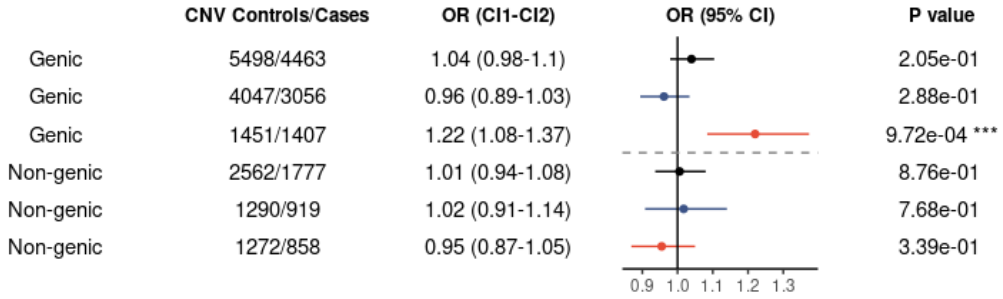
coding constraint in duplications became protective (meta-analysis OR=0.95, $P < 2.37 \times 10^{-02}$; Figure 12B) while non-coding constraint in deletions showed a trend towards association with schizophrenia (meta-analysis OR=1.04, $P < 5.53 \times 10^{-02}$; Figure 12B). When cCRE constraint was inspected, the association of deletions with schizophrenia reached significance (meta-analysis OR= 1.04, $P < 1.76 \times 10^{-02}$; Supplementary Table 14), a finding that was not observed for any other functional annotation. This finding suggests that smaller CNVs that do not affect coding genes directly but disrupt cCREs (regulatory regions) under high constraint could be involved in the pathogenesis of schizophrenia.

3.3.3.1 Coding Z score constraint effect in genic CNVs

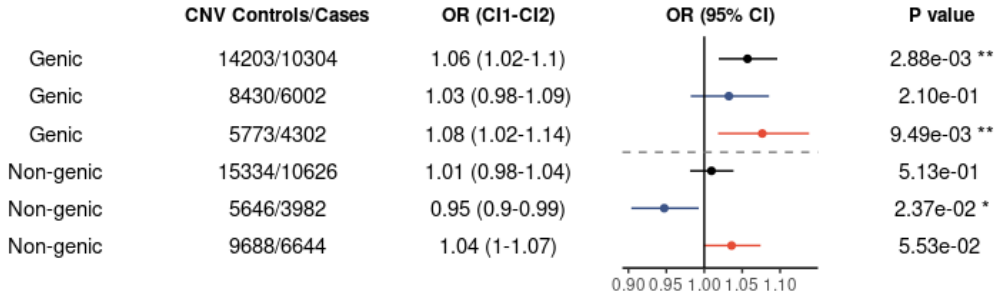
The association of coding Z score constraint with schizophrenia was independent of the number of LoF intolerant genes disrupted by the deletion (Supplementary Table 12). Because this indicated that the pLI metric may not be representing the whole spectrum of variation in coding genes with disease-relevant effects, the coding Z score needed to be tested in a sensitivity analysis. Therefore, the contribution of non-coding constraint Z score was tested again correcting for the maximum coding exon Z score in genic CNVs. This resulted in either the loss of statistical significance in deletions (meta-analysis OR = 1.09, $P = 3.17 \times 10^{-1}$; Figure 12C; Supplementary Table 15) or, surprisingly, the change of direction of effect sizes in CNVs and duplications (Figure 12C; Supplementary Table 15). This sensitivity analysis indicates that the association of non-coding constraint with genic deletions in schizophrenia is correlated with coding Z score constraint. The same loss of enrichment was observed when testing functional non-coding region Z scores separately (Supplementary Table 15). Additionally, there

could be a possible protective effect coming from the non-coding regions in genic duplications after accounting for the constraint of the genic part (Figure 12C). The same trend was observed in non-genic duplications (Figure 12B).

A Genic vs non-genic CNVs: Primary filters



B Genic vs non-genic CNVs: Secondary filters including CNVs smaller than 50 Kb



C Non-coding constraint in genic CNVs: correcting for coding constraint

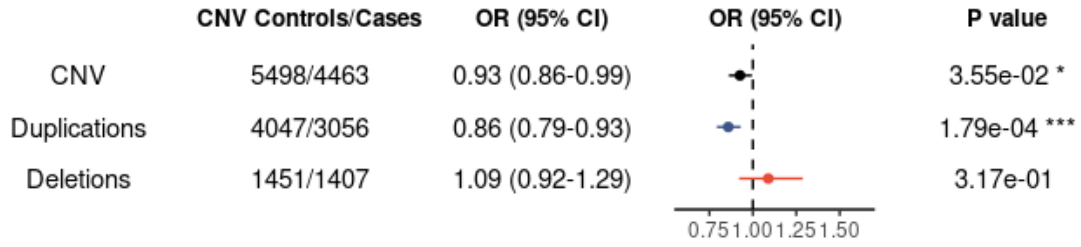


Figure 12. Comparison of the contribution of non-coding constraint in genic and non-genic CNVs.

A-B) Forest plot shows the meta-analysis of the association of the maximum non-coding Z score in the CNV and schizophrenia case/control status across different types of CNV. A) CNVs filtered by > 50 kb and > 50 SNP probes. B) Smaller CNVs filtered by > 10kb, > 10 SNP probes and > 1 probe every 5 kb. C) Effect of the addition of the coding-exon Z score correction in the association of the maximum non-coding Z score in genic CNVs with schizophrenia. Both types of CNV coloured in black, duplications in blue and deletions in red.

3.3.3 Other Z score measures

In addition to the maximum Z score, the mean, median and sum Z score in the CNV were used to evaluate whether other ways to annotate Z score constraint in CNVs were more effective predictors. The maximum Z score method allows for the identification of the putative critical region, whereas metrics like the mean, median and sum reflect the constraint of the CNV as a whole. In this case, all (coding and non-coding) Z scores were tested. Both mean (deletions meta-analysis OR = 1.17, P = 5.74×10^{-5}; Supplementary Table 17, and median (deletions meta-analysis OR = 1.17, P = 8.69×10^{-5}; Supplementary Table 18) exhibited higher effect sizes and significance than the maximum Z score showed in the primary analysis (Supplementary Table 10). In contrast, the sum of the Z score failed to report an effect despite reaching significance (Supplementary Table 16, ORs are always 1). This could be explained by the presence of a few windows in the CNVs with extremely negative Z scores that introduce noise.

3.4 Discussion

The present study investigated the role of Z score constraint of copy number variants (CNVs) in schizophrenia. The results indicate that CNVs in schizophrenia cases, and especially deletions, are significantly enriched for regions with higher levels of constraint compared to CNVs in controls, suggesting that the disruption of constrained genomic regions, including non-coding regions, plays a role in the development of schizophrenia. The sum of all Z scores in the CNV failed to replicate the primary findings obtained with the maximum Z score, which suggests that, while CNVs may span thousands of base pairs with extremely low constraint, the pathogenicity of the variant resides in the disruption of a highly constrained critical region.

These findings are consistent with previous research indicating that rare CNVs that disrupt LoF-intolerant genes confer significant genetic risk for schizophrenia (Singh et al., 2017). This is consistent with evidence from other psychiatric and neurodevelopmental disorders that the disruption of essential biological pathways may contribute to their pathogenesis (Havdahl et al., n.d.; Karczewski et al., 2020; Pardiñas et al., 2018; Singh et al., 2017; Trubetskoy et al., 2022). For example, rare, non-recurrent deletions affecting exons of *NRXN1* (with a pLI value of 0.99964) have been reported in schizophrenia as well as other psychiatric disorders (Ishizuka et al., 2020; Kirov et al., 2009; Lowther et al., 2017; Marshall et al., 2017; Schaaf et al., 2012).

The role of constrained non-coding regions in schizophrenia had not been examined previously. This study shows, for the first time, that non-coding regions with higher constraint are enriched in deletions in schizophrenia, similar to LoF-intolerant genes. Importantly, the signal is driven by those non-coding regions with an annotated

function, such as cCREs (promoters, enhancers, or CTCF-bound elements), non-coding RNA genes and introns.

The enrichment of non-coding constraint survived the exclusion of previously implicated CNV loci in schizophrenia, which highlights the potential use of Z score constraint for the discovery of novel loci. The association of non-coding constraint with deletions in schizophrenia remained significant after correcting for the number of LoF-intolerant genes disrupted by the CNV. Surprisingly, Z score constraint from coding regions was higher in schizophrenia CNVs independently of the number of LoF-intolerant genes disrupted by the CNV as well, which indicates that the Z score metric captures levels of constraint that underlie disease susceptibility but fail to be reflected by loss-of-function intolerant definitions (Chen and The gnomAD consortium, 2022; Lek et al., 2016).

A distinction between the contribution of genic CNVs and non-genic CNVs was made in a previous study and pointed towards a null contribution of non-genic CNVs (Marshall et al., 2017). In this study, genic deletions showed increased Z score constraint in schizophrenia, consistent with previous evidence (Singh et al., 2017). In terms of exclusively non-coding constraint in genic CNVs association with schizophrenia, this correlated with the coding Z score constraint of windows disrupted, but not with the number of LoF-intolerant genes affected. This finding has two implications. First, in genic CNVs, critical regions fall primarily in protein-coding genes. Secondly, Z score constraint metric may be more informative in CNVs than the number of affected genes with high probability of being loss-of-function intolerant (pLI).

Additionally, this study implicates non-genic CNVs with high constraint in schizophrenia genetic liability, in contrast to previous studies that suggested that non-genic CNVs were unimportant (Marshall et al., 2017). Another important finding of this study is that the effect of non-genic CNV non-coding constraint is concentrated in cCREs disrupted by smaller deletions (less than 50kb). Perhaps this is because longer CNVs are likely to span one or more coding exons that accumulate causal association, and therefore non-coding critical regions can only be found in smaller CNVs. Lastly, the protective signal observed from non-coding regions in both genic CNVs (after adjusting for gene constraint) and non-genic CNVs is intriguing. Overall, these results suggests that the loss of constrained non-coding DNA, for example, an enhancer, may disrupt pathways in the same way loss-of-function variants in constrained protein-coding genes do, whereas the duplication of such regulatory region may not result in an increased dosage of the protein-coding gene linked to it. Instead, the duplication of an enhancer could be contributing to redundancy (e.g., assuring the enhancer is somewhat in the right place to carry out its function). Nevertheless, a duplication of a TAD-boundary could have dramatic consequences in gene regulation. To add mechanistic clarity to this issue, different cis-acting regulatory regions should be treated separately.

This study has some limitations. First, our analysis depends on the resolution of the 1,000 bp windows (Chen and The gnomAD consortium, 2022). Any genomic window that overlapped 1 bp of a coding exon was considered as coding. This decision was made based on the hierarchy established upon the average Z score constraint of different functional elements (Table 6); however, it may be excessively strict and inflate the contribution of some categories, for example, coding exons. Second, the effect of linkage disequilibrium was not considered. It is possible that some non-coding regions,

even in non-genic CNVs, show higher constraint as a consequence of being physically close to a LoF-intolerant gene, and not due to an essential function attributed to its DNA sequence (Pardiñas et al., 2018; Tapper et al., 2005).

This study did not focus on specific loci. Because the functional consequences of the disrupted non-coding genomic elements were not investigated, it is unclear how constrained non-coding regions directly or indirectly contribute to the pathogenesis of schizophrenia. However, one can speculate that the disruption of enhancers and/or TAD-boundaries by smaller CNVs could cause the aberrant expression of genes with an essential role in, for example, synaptic processes. Future research is needed to better understand the functional consequences of CNVs in schizophrenia, and to identify additional genetic risk factors for the disorder. Recent efforts from different research groups and consortia, such as GTEx, PsychENCODE or the Allen Brain Atlas offer exciting possibilities for linking non-coding regulatory regions to coding genes, including expression data at the single cell resolution and genotyping data from psychiatric patients and controls, which would allow for the identification of eQTLs active in specific cell-types that are important in brain development and disease (Akbarian et al., 2015; Bryois et al., 2022; Lonsdale et al., 2013; Miller et al., 2014; THE GTEX CONSORTIUM, 2020). Also, the level of conservation of a DNA sequence across species, which can inform about its functionality, was not integrated into this analysis (Pollard et al., 2010). Such conservation scores were excluded because one of the main assumptions of this thesis is that recently evolved genomic regions comprise a substantial proportion of the genetic susceptibility to schizophrenia in humans.

In conclusion, this study provides novel evidence that the non-coding regions disrupted by CNVs in schizophrenia are more likely to be highly constrained, suggesting that the disruption of not only biologically essential coding genes, but also non-coding functional elements may contribute to the development of the disorder. In future, it may be more informative to categorize variants based on their downstream effects on gene expression in model organisms and patient derived cells and organoids. These findings highlight the potential for the use of gene constraint metrics together with non-coding constraint analysis for identifying novel loci that confer genetic liability for schizophrenia. Furthermore, this approach may ultimately lead to the development of more effective diagnostic and treatment strategies for the disorder.

3.5 Methods

3.5.1 CNV calls

Pre-made CNV calls that had already passed quality control filtering from three independent studies were included in this study. The studies were i) Molecular Genetics of Schizophrenia (MGS); ii) ClozUK1; and iii) ClozUK2. The MGS study used the Illumina Affymetrix 6.0 probe array, with 2,200 cases and 2,571 controls. ClozUK1 used the Human Omni Express 12v1 probe array, with 6,307 cases and 10,675 controls. ClozUK2 also used the Human Omni Express 12v1 array, with 5,648 cases and 8,414 controls. In total, 14,155 cases and 21,660 controls were used in this study. Two additional different CNV quality control thresholds were used to filter CNVs: CNVs filtered by > 50 kb and > 50 SNP probes, and > 1 probe every 5 kb; and CNVs filtered by > 10kb, > 10 SNP probes, and > 1 probe every 5 kb.

3.5.2 Annotation of CNV Z score constraint based on genomic features

For each CNV, a maximum Z score was annotated. First, all Z score windows were considered to assess genomic constraint. Because different genomic elements overlap with each other, a hierarchy was later established based on the average Z score of each feature, with coding regions having the highest priority, followed by cCREs, non-coding exons, introns, and ultimately, regions with no annotated function (Introduction Figure 9; Table 6). Gene and cCRE annotation from the Basic Gene Annotation Set (GENCODE Version 42lift37) and the Candidate cis-Regulatory Elements by ENCODE (Registry of cCREs V3) was employed. Scanning of overlapping features was performed with the scan_region.pl tool in the PennCNV v1.0.5 software.

3.5.3 Logistic regression and meta-analysis

For each study, a logistic regression analysis was performed, using the CNV case/control status (of the person harbouring the CNV) as the dependent variable, and the maximum Z score constraint (from the functional category of interest [Table 6]) as the independent variable. The logistic regression model was adjusted for sex and length of the CNV as covariates (Equation 7). Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated for each study. Effect sizes and standard errors were combined across the three studies by a meta-analysis using inverse-variance weighting fixed-effect model. All statistical analyses were performed with the *stats* and *rma* packages in R.

(Equation 7) *CNV case/control status (0/1) ~ Maximum constraint (Z score)* + length CNV (bp) + sex (male/female)

3.5.3.1 Sensitivity analyses

To assess whether known loci explained the majority of the signal, CNVs were filtered out if they spanned any of the previously implicated loci in Table 1 and logistic regression in the form of Equation 7 was performed. To test for correlation with the number of LoF-intolerant genes affected, this variable was added as a covariate to the regression (Equation 8). The list of LoF intolerant genes (pLI > 0.9) was available from a previous study (Lek et al., 2016). For genic CNVs, an additional sensitivity analysis was performed, testing for the association with the maximum coding constraint Z score (Equation 9). All statistical analyses were performed with the *stats* and *rma* packages in R.

(Equation 8) *CNV case/control status (0/1) ~ Maximum constraint (Z score)* + length CNV (bp) + sex (male/female) + LoF-intolerant genes affected (absolute number)

(Equation 9) *genic CNV case/control status (0/1) ~ Maximum non-coding constraint (Z score)* + length CNV (bp) + sex (male/female) + Maximum coding constraint (Z score)

Table 6. Hierarchy for the different categories of constraint Z scores windows tested.

	CATEGORY	DEFINITION	MEAN Z SCORE
CODING	Coding exons	Coding exons overlapping any other functional annotation	1.51
NON-CODING	cCREs	cCREs not overlapping any coding exon, but overlapping non-coding exons and introns	0.66
	Non-coding exon	Non-coding exons not overlapping any coding exon or cCRE, but overlapping introns	0.08
	Introns	Introns not overlapping any other functional element	-0.1
	No function annotated	Intergenic and non-regulatory (no cCRES)	-0.68
ALL	Genomic	-	0.03

Chapter 4: *In silico* characterization of *PSG8-AS1* long non-coding RNA in the segmentally duplicated *PSG* locus and its involvement in human brain development and disease

4.1 Summary

The segmentally duplicated Pregnancy-specific Glycoprotein (*PSG*) gene family may be the most rapidly evolving locus in the human genome. Genes derived from segmental duplications are important drivers of brain evolution. This chapter describes the *in-silico* characterization of *PSG8-AS1*, a human oligodendrocyte-specific lncRNA, emphasizing the possibilities of utilizing bioinformatic resources for the selection of candidates for study of brain evolution relevant genes specific to humans. Expression of *PSG8-AS1* orthologs was evaluated in the brain of several primate species and was found only in humans. In addition, its putative functions in oligodendrocytes and involvement in oligodendroglioma were examined. *PSG8-AS1* was found to be highly correlated with the expression of myelin components, as well as predicted to bind many proteins implicated in the transport of plasma membrane constituents. Additionally, *PSG8-AS1* may be a promising candidate for a diagnostic biomarker in glioma, with prognostic value in oligodendrogliomas. Understanding the functions of lncRNAs such as *PSG8-AS1*, generated in segmental duplications, in the human brain will contribute to our knowledge of the molecular mechanisms underpinning human cognitive capacities and developing new therapeutic approaches for brain diseases.

4.2 Introduction

4.2.1 The pregnancy-specific glycoprotein (*PSG*) locus

The human pregnancy-specific glycoprotein (*PSG*) segmentally duplicated locus located at 19q13.2-13.31 encodes 10 proteins (PSG1-9, PSG11) predominantly produced in the placenta (Moore et al., 2022; Moore and Dveksler, 2014). During pregnancy, PSGs are secreted at high levels into the maternal bloodstream, where they may modulate the immunological and vascular systems (Rattila et al., 2019; Shanley et al., 2013; Towler et al., 1977, 1976; Zimmermann and Kammerer, 2021). PSGs can activate soluble and extracellular matrix bound TGF-β, potentially regulating processes in diverse tissues (Ballesteros et al., 2015). Deregulation and genetic variation in the *PSG* locus have been identified as a risk factor for pregnancy-related complications, such as preeclampsia (Arnold et al., 1999; Blankley et al., 2013; Chang et al., 2016; Gormley et al., 2017; Hertz and Schultz-Larsen, 1983; MacDonald et al., 1983; Rattila et al., 2019; Temur et al., 2020; Zhao et al., 2012). Preeclampsia is characterized by high blood pressure and is one of the leading causes of maternal and foetal morbidity and mortality (Phipps et al., 2019). Apart from the placenta, PSGs are expressed throughout the gastrointestinal system and in some tumours (Houston et al., 2016; Kawano et al., 2007; Mathews et al., 2020). In a recent study, approximately 20% of patients with lung, breast, uterine, and colon cancer were found to have an altered PSG expression profile, which was associated with a poor prognosis (Mathews et al., 2020). In conjunction with prior discoveries linking PSGs to immunological tolerance, these findings implicate PSGs in a putative immune tolerance mechanism of malignancy.

There is solid evidence that *PSG* genes are subject to rapid evolution, including CNVs, divergence in the composition and structure of protein domains, and non-conservative

changes in the amino acid sequence of open reading frames (Chang et al., 2016; Sudmant et al., 2013; van der Lee et al., 2017). Some orthologous gene pairs exist between mouse and rat, as well as between humans and apes (Moore and Dveksler, 2014). Despite the rapid evolution of *PSG* genes, the encoded proteins in all analysed species—including those of the mouse, horse, and human—share analogous functions (Kammerer et al., 2020; Shanley et al., 2013; Warren et al., 2018). Currently, selection for enhanced gene product dosage rather than for functional diversification provides a better explanation for the growth and rapid development of *PSG* families across species (Haig, 2008, 1993; Moore, 2012).

The *PSG* locus segmental duplication has been identified as the locus most prone to copy number variation in the human genome (Sudmant et al., 2013). The outstanding evolutionary pressure on the *PSG* locus has been interpreted as a consequence of maternal-foetal conflict. The maternal-foetal conflict theory, derived from parent-offspring conflict, is based on the different selective forces in the mother and the embryo governing the level of maternal investment in the pregnancy. Therefore, selection of maternal genes will limit resource transfer (such as glucose) to the embryo, whereas embryonic genes will favor it (Haig, 2008, 1993; Moore, 2012). Such conflicts may encourage rapid evolution within and between species and underpin the generation of new genes in the *PSG* locus including the uncharacterized lncRNA *PSG8-AS1*, whose function is explored here.

4.2.2 The non-coding content of the *PSG* locus

Two of the eleven *PSG* genes, specifically *PSG10* and *PSG7,* may not function predominantly as proteins. *PSG10* has been commonly excluded from the PSG protein

family (PSG1-9, PSG11) since it has been annotated as a pseudogene. *PSG10P* harbours a very premature stop codon in the majority of the population, with less than 1% of alleles of SNP rs1367178999 (T/A/C) leading to an intact open reading frame in humans (Sherry et al., 2001). Although not producing a protein, the transcript is still expressed and may retain functionality (N. Wang et al., 2019). The *PSG10P* transcript has been described as a sink for microRNA miR-19a-3p, which regulates the expression of *IL1RAP* mRNA by binding to its 3' untranslated region (UTR). The *PSG10P*/miR-19a-3p/IL1RAP axis may influence the invasion capacity of trophoblasts under hypoxia and be dysregulated in preeclampsia, which implicates the non-coding version of *PSG10P* in pregnancy in parallel with the PSG proteins (N. Wang et al., 2019).

Another common SNP responsible for a truncation of a PSG protein is rs113247044, which results in a stop codon in *PSG7* exon 2 in approximately 70% of people. The effect of rs113247044 on *PSG7* mRNA levels was investigated alongside the investigations contributing to this thesis and published (Moore et al., 2022). Individuals inferred to have the frequent, truncating genotype showed significantly lower levels of PSG7 mRNA, indicating that the rs113247044 stop codon allele may cause nonsense-mediated mRNA decay. It is unknown whether the *PSG7* transcript is functional as a ncRNA or if the exclusion of exon 2 or initiation of translation from a downstream AUG codon results in a protein product (Moore et al., 2022).

The lncRNA *PSG8-AS1* overlaps the opposite strand of a non-canonical *PSG8* intron isoform. Its sequence is not derived from the *PSG* coding exons, suggesting it did not evolve from a truncation of a PSG protein. In this chapter, publicly available transcriptomic and genomic datasets are used to elucidate *PSG8-AS1* possible

relationship with the coding *PSG* genes and functions in human biology and disease. Evidence provided in this chapter indicates that *PSG8-AS1* genomic sequences have emerged in the primate linage. However, it may only be expressed in humans, where it may be a brain-specific lncRNA involved in myelin sheath formation by oligodendrocytes.

4.2.3 The human brain is uniquely complex

The human brain is unique in its complex functions and abilities, and this is reflected in its epigenetic and transcriptomic diversity compared to other tissues and species (An et al., 2023; Berg et al., 2021; Florio et al., 2015; Jeong et al., 2021; Laukoter et al., 2020; Liu et al., 2022). Epigenetic diversity refers to differences in DNA methylation or histone modifications between different tissues, while transcriptomic diversity refers to differences in gene expression patterns between tissues.

One of the main differences between the brain and other tissues is its unique set of epigenetic signatures. Histones are proteins that package DNA into structures called chromatin and their modification patterns are highly tissue-specific. Depending on the degree of compaction of nucleosomal DNA, the rate of transcription will vary (Hyun et al., 2017; Lewis et al., 2021; Sadakierska-Chudy and Filip, 2015). DNA CG methylation is a stable epigenetic alteration of genomic DNA that represses transcription and plays a crucial part in brain development. DNA methylation in a non-CG context, namely CH methylation (where H = A, C, and T), is more frequent in the brain compared to other tissues and is linked to postnatal neuronal development and cell-type-specific transcriptional activity (Jeong et al., 2021; Stroud et al., 2017). During the evolution of the human brain, significant modifications in DNA methylation

have taken place in cell-type- and cytosine-context-specific ways. These DNA methylation changes have a significant impact on the regulatory architecture of the human brain and disease vulnerability, including in schizophrenia (Jeong et al., 2021).

Ultimately, distinct epigenetic signatures such as histone modifications and DNA methylation in the brain result in a unique set of gene expression patterns. A recent study used the Genotype-Tissue Expression (GTEx) transcriptomic resource to build a brain-specific gene database based on the specificity index τ parameter (Liu et al., 2022; Lonsdale et al., 2013; Yanai et al., 2005). Housekeeping genes are characterized by specificity index τ values close to 0, whereas tissue specific genes tend towards the maximum value of 1. After removing many genes with very low expression, the specificity index τ was calculated and brain-specific genes were defined as those upregulated in brain tissues and with a specificity index $\tau>0.9$. A total of 475 coding genes and 134 non-coding RNAs were identified as brain-specific, which indicates that more than 600 genes are expressed exclusively in the brain. This includes genes involved in synaptic plasticity, learning and memory, and brain development (Liu et al., 2022). Of note, the previously uncharacterized lncRNA *PSG8-AS1* is contained within this brain-specific gene list as *AC004603.4,* an alias used in previous genome assemblies. Possibly, many other cell-specific non-coding transcripts have been excluded due to their poor representation in bulk-RNAs sequencing datasets. Understanding how these cell specific gene networks are uniquely expressed in the brain can provide new insights into the evolutionary processes that have shaped human cognitive function and brain disease.

4.2.4 Specific cell types in the brain

The brain is composed of many different types of cells, each with a specific function and gene markers. Neurons are the primary cells of the nervous system and are responsible for transmitting signals in the brain. They have a unique shape characterized by a cell body, dendrites, and an axon (Elston et al., 2001). Providing support and insulation to neurons, glial cells are involved in a variety of metabolic and immunological functions and include astrocytes, oligodendrocytes, and microglia (Allen and Lyons, 2018). Ependymal cells line the ventricles of the brain and play a role in the production and circulation of cerebrospinal fluid (Jiménez et al., 2014). These cell types are produced by neural stem cells (NSCs), which play a key role in brain development, and repair and recovery after injury (Zhao and Moore, 2018). Lastly, radial glial cells serve as scaffolds for the migration and differentiation of neurons during brain development (De Juan Romero and Borrell, 2015).

In this chapter, *PSG8-AS1* is characterized as an oligodendrocyte (OL) specific lncRNA. In the central nervous system (CNS), oligodendrocytes are the cells engaged in myelinating axons (de Faria et al., 2021; Grotheer et al., 2022; Marton et al., 2019; Ouyang et al., 2019). Myelination is the process by which axons are covered with oligodendrocyte membrane material that has a unique protein and fatty acid composition. This covering, termed myelin, helps the nerve fibres to conduct electrical signals more efficiently, resulting in faster and more reliable communication between neurons (Grotheer et al., 2022; Knowles et al., 2022). Not only myelin gene expression, but also other factors such as cell-cell adhesion between the axon and oligodendrocyte membrane are essential for the biogenesis of functional myelin (Asou et al., 1995; Kachar et al., 1986; Laursen et al., 2009; Marton et al., 2019). Thus, the

oligodendrocyte differentiation program in humans is a precisely orchestrated process that involves the coordination of thousands of genes (Knowles et al., 2022). It is critical to broaden our knowledge in the oligodendrocyte-specific transcriptome machinery since functional myelination is essential for cognition (Maas et al., 2020; Steadman et al., 2020), and oligodendrocyte dysregulation influences a variety of demyelinating conditions such as multiple sclerosis (Dobson and Giovannoni, 2019); acute disseminated encephalomyelitis (Paolilo et al., 2020); and psychiatric disorders such as schizophrenia and ASD (Fessel, 2022; H. Lee et al., 2019).

A myriad of oligodendrocyte-specific lncRNAs have been studied in mice (Dong et al., 2015; He et al., 2017; Li et al., 2018; Mercer et al., 2010; Stolt et al., 2005; Tochitani and Hayashizaki, 2008; Wei et al., 2021), but replication of these findings in humans is unlikely due to high interspecies divergence of non-coding transcripts and divergent oligodendrogenesis regulation (Bradl and Lassmann, 2010; Cabili et al., 2011; Paralkar et al., 2014; Zhao and Moore, 2018). The lack of human oligodendrocyte models impedes the transition of lncRNA research to a human brain-specific focus. Only one lncRNA, *OLMALINC*, has been identified as being specifically associated with human oligodendrocytes (Mills et al., 2015), but the MO3.13 oligodendrocyte cell line used in that study has been shown to be ineffective at modelling oligodendrocyte behaviour (De Kleijn et al., 2019). However, the scientific community's increasing access to bioinformatics resources provides an opportunity to identify and investigate new candidates.

Each of these cell types is essential for the activity of the brain and the integration of neural signals; therefore, disruptions in the function or number of any of these cell types could contribute to a variety of neurological and psychiatric disorders.

4.2.5 Brain disease

Brain disease refers to a broad and complex range of medical conditions that affect the structure or function of the brain, leading to various levels of impairment. There are many types of brain pathologies, including neuropsychiatric disorders, neurological diseases, and brain tumours. Most of the genetic signal in neurologic and neuropsychiatric disorder GWAS comes from non-coding regions (Hattori et al., 2003; Umeton et al., 2022; Xiao et al., 2017), emphasizing the relevance of non-coding RNA in the brain. In addition, many lncRNAs are dysregulated in human brain disease (Table 7) (Cai et al., 2020; Faghihi et al., 2008; Katsel et al., 2019; Katsushima et al., 2016; Lipovich et al., 2012; Liu et al., 2016; Talkowski et al., 2012; Voce et al., 2019; Yue et al., 2019). Neuropsychiatric disorders have been reviewed in the general introduction in this thesis. Neurologic diseases occur because of structural, biochemical, or electrical abnormalities in the brain and spinal cord. They can be caused by a variety of factors, including brain injury, infection, and genetic mutations. Some common examples of neurologic diseases comprise neurodegenerative disorders such as Alzheimer's and Parkinson's disease, demyelinating diseases like multiple sclerosis and acute disseminated encephalomyelitis, and other conditions including epilepsy and strokes (Faghihi et al., 2008; Galovic et al., 2021; Höftberger and Lassmann, 2018; Lamptey et al., 2022).

Table 7. LncRNAs dysregulated in brain disease

LncRNAs are listed by alphabetical order of the references.

LNCRNA NAME & CHANGE	**DISEASE**	**REFERENCE**
***LINC00998* downregulation**	Glioblastoma	(Cai et al., 2020)
***BACE1-AS* upregulation**	Alzheimer's disease	(Faghihi et al., 2008)
***NEAT1* downregulation**	Schizophrenia	(Katsel et al., 2019)
***TUG1* upregulation**	Glioblastoma	(Katsushima et al., 2016)
***BDNFOS* downregulation**	Epilepsy	(Lipovich et al., 2012)
***LOC646329* upregulation**	Glioblastoma	(Liu et al., 2016)
***LINC00299* genomic rearrangement**	Neurodevelopmental delay	(Talkowski et al., 2012)
***MALAT1* upregulation**	Glioblastoma upon temozolomide treatment	(Voce et al., 2019)
***TUG1* upregulation**	Multiple sclerosis	(Yue et al., 2019)

Brain tumours also fall in the category of neurologic illness. The most common form of malignant brain tumour arises from glial cells or their precursors. Gliomas continue to be among the most challenging cancers to treat, with a 5-year overall survival of 35%. Treatment commonly involves a combination of surgery, radiation therapy, and chemotherapy, but the prognosis and treatment options vary depending on the tumour subtype and stage (Lapointe et al., 2018).

The understanding of glioma pathogenesis has greatly improved in the last 15 years, leading to the discovery of several clinically different genetic alterations that, together with histology, have contributed to improved diagnosis, grading and treatment decision making (Lapointe et al., 2018; Louis et al., 2016). Briefly, diffuse gliomas can be classified by the mutation in the isocitrate dehydrogenases 1 and 2 (*IDH1* and *IDH2*). Wildtype tumours (IDH-wt) mostly develop into grade IV glioblastomas, the most aggressive subtype of glioma with a median overall survival of one year (Eckel-Passow et al., 2015; Wick et al., 2009). In the cases where the *IDH* gene is mutated (IDH-mt), tumours are further categorized by the codeletion of chromosomes 1p and 19q from a non-balanced centromeric translocation t(1;19)(q10;p10), termed 1p/19q codeletion. The detection of 1p/19q codeletion distinguishes oligodendroglioma from astrocytoma and confers a favourable prognosis (median overall survival of more than twelve years) (Ducray et al., 2008; Eckel-Passow et al., 2015; van den Bent, 2007). The remaining IDH-mt gliomas are associated with an intermediate prognosis, which varies between two years and twelve years median overall survival depending on the grade of the tumour (Lapointe et al., 2018; Louis et al., 2016). Of note, it has been proposed that oligodendroglioma-specific tumour suppressor genes reside in 1p/19q, such as *FUBP1* at 1p and *CIC* at19q, among others (Bettegowda et al., 2011; Gladitz et al., 2018).

Because of the differences in gene distribution on chromosomes between humans and model species, this genomic signature of oligodendrogliomas is most likely unique to humans (Lv et al., 2021).

The study described in this chapter is the first attempt to characterize a human brain-specific lncRNA derived from a segmental duplication. The under-representation of lncRNAs in bulk-RNA-seq studies and the lack of human oligodendrocyte cell models has been overcome by the utilization of publicly available brain single cell and spatial RNA-seq resources. Another major challenge encountered has been the camouflaged nature of the *PSG* locus, which has hindered the association of variation of *PSG8-AS1* with phenotypes. However, the specific expression of *PSG8-AS1* in oligodendrocytes and its location at 19q, whose deletion together with 1p and the *IDH* mutation characterizes oligodendrogliomas, were the basis for the study of the possible involvement of *PSG8-AS1* in gliomas.

4.3 Results

4.3.1 *PSG8-AS1* is an oligodendrocyte-specific lncRNA

The relationship of *PSG8-AS1* with the coding *PSG* genes was initially examined by comparing their respective tissue expression patterns. According to GTEx, *PSG8-AS1* is considerably enriched in all compartments of the brain, in contrast to coding *PSGs*, which are expressed in the placenta and other tissues at lower levels (placental datasets GSE109082 and GSE148241, Table 9, Figure 14A-B) (Lonsdale et al., 2013). Antisense lncRNAs are known to regulate the transcription of parental mRNA (Boque-Sastre et al., 2015). However, due of the non-overlapping tissue expression patterns and the extremely low transcription of the *PSG8* intron which *PSG8-AS1* overlaps across all tissues, downregulation of *PSG8* expression by *PSG8-AS1* seems unlikely (Figure 13) (Figure 14A-B).

There is evidence that lncRNAs are cell type-specific (Cabili et al., 2011). Cell specificity was analysed using two publicly available brain single cell RNA sequencing (scRNA-seq) datasets containing cell identity information (GSE67835 and GSE118257, Table 9). *PSG8-AS1* was preferentially expressed in oligodendrocytes across all cell subtypes, despite being less abundant than marker genes in both datasets (Figure 15A-B). This observation can be explained by the fact that lncRNAs will be more challenging to detect in scRNA-seq experiments due to their low expression levels compared to coding RNAs (Cabili et al., 2011). When *PSG8-AS1* expression is present in other cell types, such as endothelial cells, oligodendrocyte markers are also expressed (Figure 15A-B). The Brain Cell Type Specific Gene Expression R/Shiny tool (http://celltypes.org/brain/) (McKenzie et al., 2018) further validates oligodendrocyte specificity (Supplementary Figure 2).

Transcription factor binding to ENCODE cCREs in proximity to the genomic sequence of *PSG8-AS1* was scrutinised to further examine its specificity for oligodendrocytes (Supplementary Table 19). Up to 60 general gene regulatory proteins act on *PSG8-AS1*-associated enhancers, 20 of them having a role in CNS function. The RE1-Silencing Transcription factor (REST) complex, which represses transcription of genes underpinning neural cell fate decisions in non-neural tissues, also targets *PSG8-AS1* regulatory regions (Figure 16). Interestingly, commitment to the oligodendrocyte lineage relies on the coordination of REST and CoREST regulatory modules (Abrajano et al., 2009). Of particular interest, two transcription factors essential for terminal oligodendrocyte differentiation and myelination, ZNF24 and TCF7L2 (Elbaz et al., 2018, p. 24; Zhang et al., 2021), interact with enhancer EH38E1955813, 10 kb downstream of *PSG8-AS1* promoter (Figure 16; Supplementary Table 19). In addition to proximity, the accumulation of expression quantitative trait loci (eQTL) active in brain for *PSG8-AS1* expression in the same genomic location indicates that enhancer EH38E1955813 is likely to regulate *PSG8-AS1* transcription (Figure 16; Supplementary Table 20). The tumour suppressor protein p53 binds to enhancers EH38E1955805 and EH38E1955806 1 kb upstream of the promoter (Figure 16; Supplementary Table 19). Since p53 promotes cell differentiation and coordinates the tumour suppressor transcriptome across all cancers (Hafner et al., 2019), this interaction suggests not only a potential involvement of *PSG8-AS1* in the final maturation of oligodendrocytes, but also in tumours arising from glial cells.



Figure 13. Expression of PSG8 is low across tissues.

The exon that involves *PSG8-AS1* overlapping intron is never detected in GTEx data (Lonsdale et al., 2013).

A



B

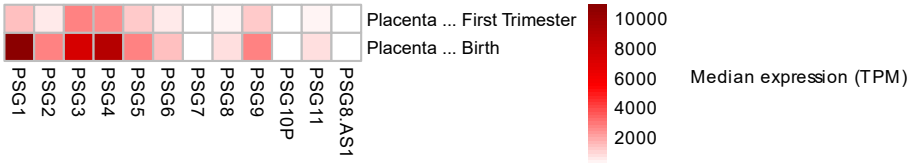
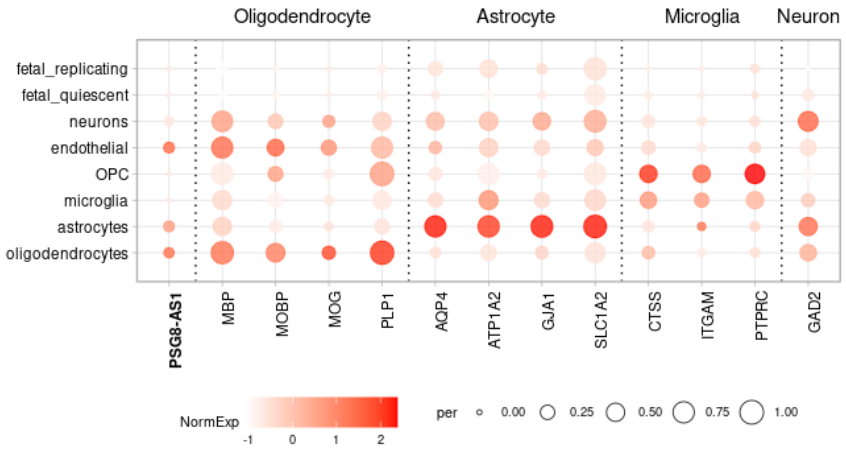


Figure 14. Different tissue expression of the *PSG* genes and *PSG8-AS1*.

A) The median expression in Transcripts per Million (TPM) for each gene and tissue from GTEx is shown as a heatmap. B) The median expression in TPM for each gene in two different placenta datasets is shown as a heatmap (placental datasets GSE109082 and GSE148241, Table 9).

A



B

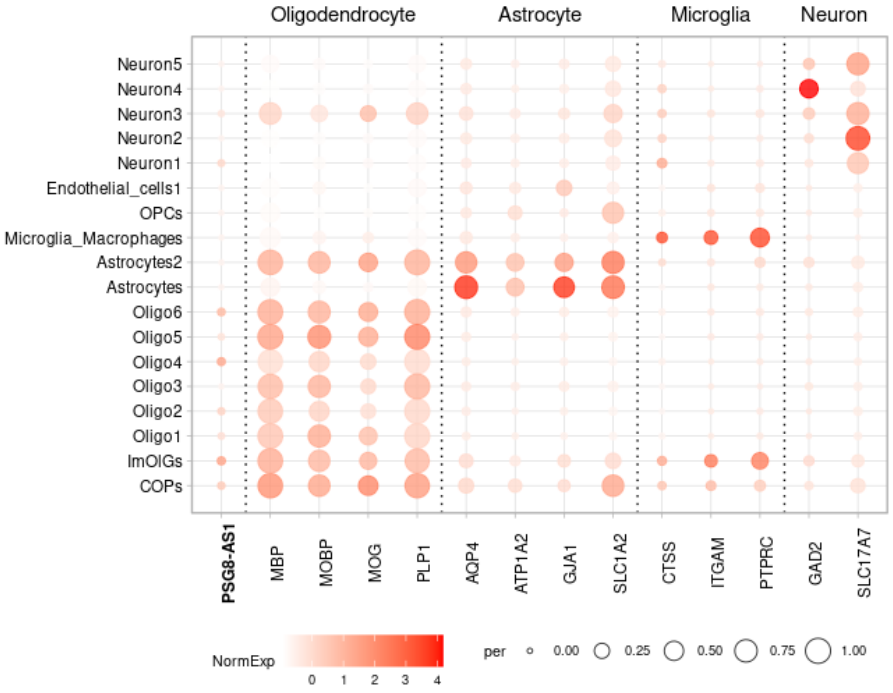


Figure 15. Comparison of single cell-specific expression of *PSG8-AS1* with brain cell marker genes.

Normalized expression is shown as colour intensity, whereas the percentage of cells where gene expression is identified is represented by the size of the circle in the dotplot. A) Data from GSE67835 (scRNA-seq from human brain samples). B) Data from GSE118257 (scRNA-seq from human brain samples). For more details about datasets, please see Table 9.

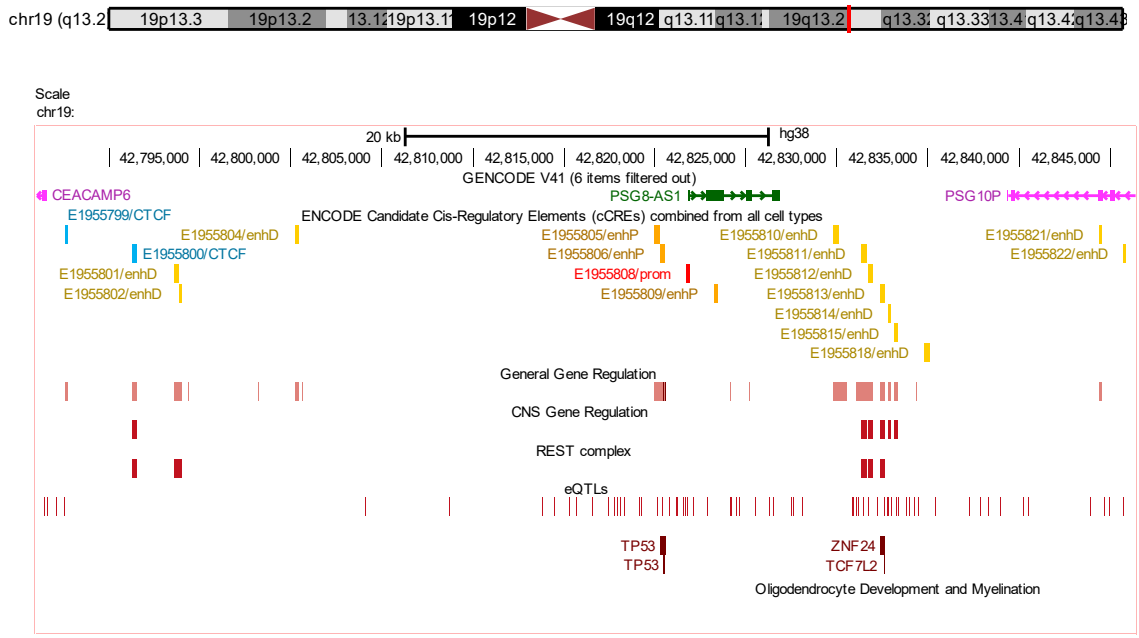


Figure 16. Localization of cCREs, different sets of transcription factor binding proteins, and eQTLs that potentially regulate *PSG8-AS1*.

Modified from UCSC browser (genome version hg38). GENCODE V41 track shows lncRNA genes in green and pseudogenes in pink. ENCODE cCREs track shows CTCF-bound only in blue, distal enhancers in light yellow, proximal enhancers in orange, and promoters in red. Transcription factor binding sites (General, CNS, REST complex, TP53, ZNF24 and TCF7L2) and eQTLs are shown in different shades of red.

4.3.2 *PSG8-AS1* emerged in the Catarrhini lineage

To retrieve potential orthologs of the human *PSG8-AS1* gene, the *blastn* online tool (https://blast.ncbi.nlm.nih.gov/) was used to interrogate the most up to date collections of genomes and validated or predicted RNAs from all species (termed the *nr* collection, which includes non-redundant sequences of genomic DNA, plasmid DNA, contigs, etc. in GenBank, EMBL, DDBJ, PDB and RefSeq databases). The sequence of human *PSG8-AS1* RNA (NR_036584.1) was introduced as input and the alignments were filtered by more than 80% span and similarity. Annotated lncRNAs from the gorilla (Homininae), orangutan (Hominidae), gibbon (Hominoidea), Tibetan macaque, drill and baboon (Catarrhini) were found in the putative orthologs list. The rhesus macaque was also included in the results list; however, similarity was found with genomic sequence instead of an annotated, cloned RNA entity. The genomic coordinates in the rhesus macaque were used to specify a preliminary annotation for the orthologous *PSG8-AS1* lncRNA in this species (Table 8).

Transcript annotation in primate genomes is consistent with the conservation at the genome level (Figure 17). Best alignments (represented by "nets") in the UCSC browser Mammals Multiz Alignment & Conservation (27 primates) track were inspected. Segments spanning exons 2 and 3 are either not continuous or poorly assembled in the chimpanzee and bonobo genomes. One explanation for why this segment remains intact in the human genome might be that the chimpanzee and bonobo individuals whose DNA was sequenced for the reference genome assembly happened to have a structural variant, or the assembly algorithm failed to resolve this segmentally duplicated locus. Since segmental duplications are challenging to assembly, *PSG8-AS1* might simply be camouflaged in the latest chimp and bonobo genomic assemblies.

Another explanation is that the common ancestor between humans, chimps and bonobos had a functional copy of *PSG8-AS1,* but it was only retained in the human lineage and lost in the chimp and bonobo. Although significant sequence is conserved (Figure 17; Figure 18), gaps in alignment nets with other species are more abundant due to the common structural rearrangements in segmental duplications and different chromosomal architecture.

The conservation at the RNA level was also inspected by the multiple alignment of RNA sequences of *PSG8-AS1* orthologs (Figure 18; Table 8). The exon composition is not the same across orthologous lncRNAs, with the macaques, drill and baboon losing exon 1 almost entirely, and gibbons, who are depleted of significant sequence of exon 4. The magnitude of functional change these alterations have in other species is unknown; however, it suggests that exon 2 and exon 3 might be more important for *PSG8-AS1* functionality.

To evaluate if *PSG8-AS1* candidate orthologs are expressed in primates other than the human, RNA-seq reads from brain and other tissues from several primate species were realigned to capture *PSG8-AS1* (GSE100796 and GSE30352, Table 9, Figure 19A-B). In the case of rhesus monkey, a modified annotation file was created to include the predicted *PSG8-AS1* ortholog. The levels of *PSG8-AS1* RNA were higher in the human brain compared to other human tissues and other primates (Figure 19A-B). Of note, only one or two individuals for each primate are included in the datasets, except for humans. Although more primate individuals should be assessed, this finding suggests that transcription of *PSG8-AS1* orthologs in other primates might be inconsequential, whereas the active expression is enriched in the human brain. The emergence of genes

important for brain evolution in primates with enriched expression in the human brain has been described previously and it has been speculated that specific changes in regulatory elements occurred in these genes during human evolution (Florio et al., 2018).

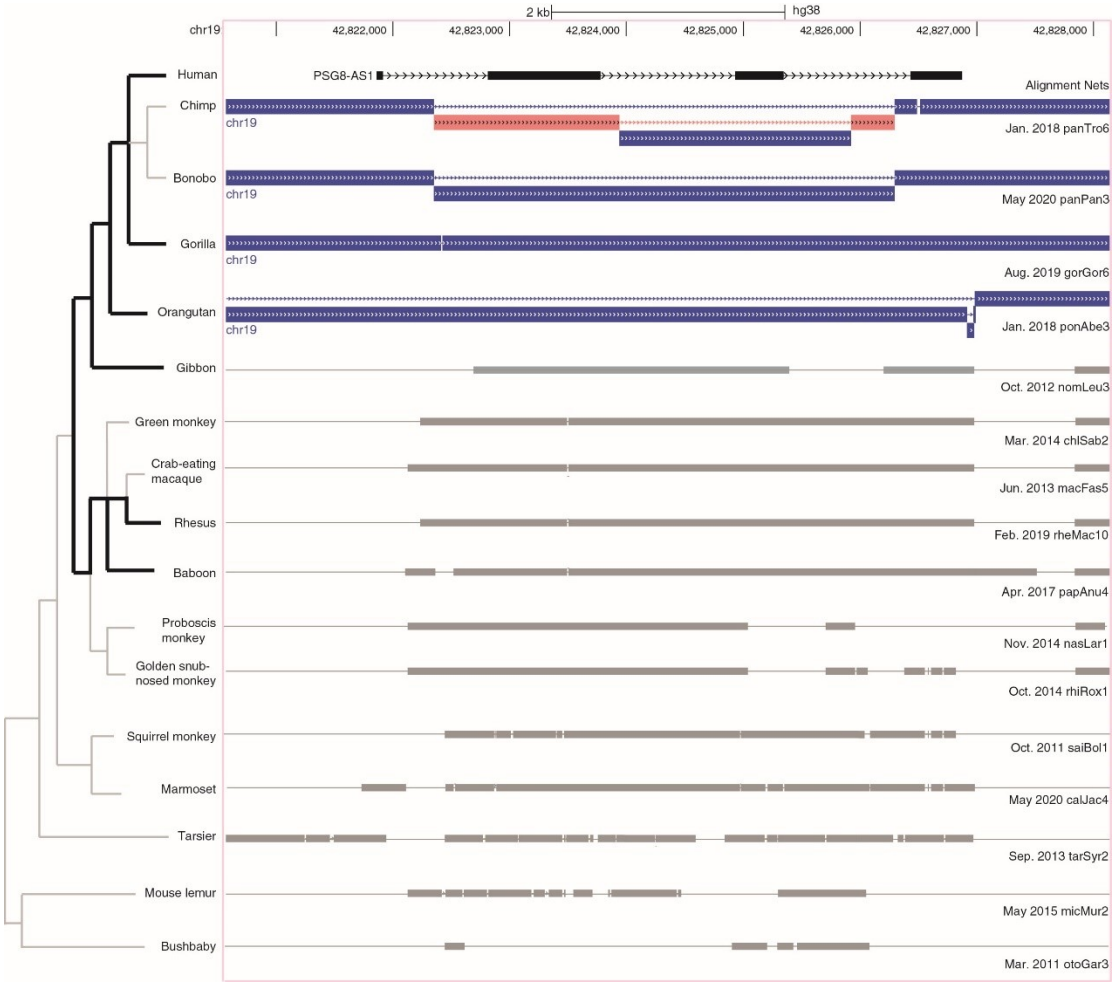


Figure 17. Best alignments to the human genome (represented by "nets") in the UCSC browser Mammals Multiz Alignment & Conservation (27 primates) to *PSG8-AS1*.

Human genome version is hg38. *PSG8-AS1* sequence in the human genome is shown at the top (boxes=exons, lines=introns, arrows=strand). For alignments, boxes represent ungapped alignments; lines represent gaps. In the chimp, bonobo, gorilla and orangutan, blue boxes indicate alignment to chromosome 19, whereas red boxes indicate an alignment to clone contigs that cannot be confidently placed on a specific chromosome. Arrows indicate the direction of the alignment. In the rest of the primates the colour is simplified to grey only. Again, boxes represent ungapped alignments and lines represent gaps. Primate phylogeny is show to the left, where thicker lines represent species that harbour a *PSG8-AS1* ortholog based on the "blastn" result.

Table 8. Putative *PSG8-AS1* orthologs obtained from blastn.

ACCESSION	SPECIES	COMMON NAME	QUERY COVER	PER. IDENTITY	LENGTH (BP)
XR_004068020.1	*Gorilla gorilla*	Gorilla	100%	96.16%	1866
XR_002912515.1	*Pongo abelii*	Orangutan	100%	93.94%	1880
XR_004026472.1	*Nomascus leucogenys*	Gibbon	80%	93.44%	1508
XR_004251333.1	*Hylobates moloch*	Silvery gibbon	80%	93.71%	1596
NA	*Macaca mulatta*	Rhesus macaque	97%	90.11%	NA
XR_007721678.1	*Macaca thibetana*	Thibetan macaque	97%	90.13%	1814
XR_001007954.1	*Mandrillus leucophaeus*	Drill	97%	89.93%	1900
XR_003116959.1	*Theropithecus gelada*	Gelada baboon	97%	89.58%	1813

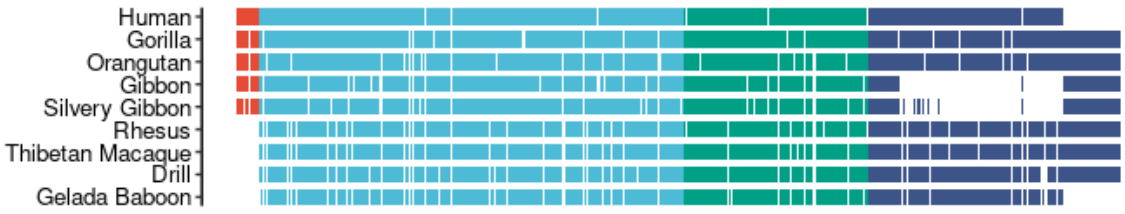
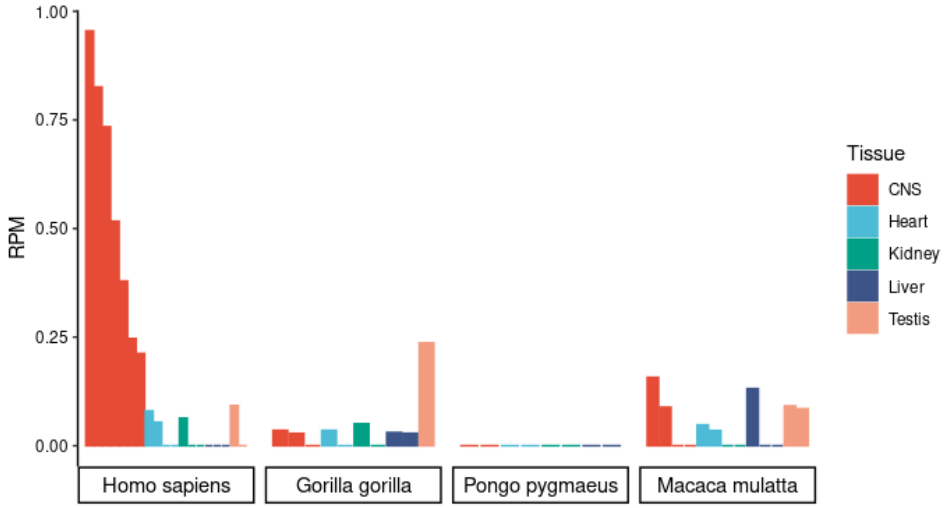


Figure 18. Multi-Alignment of the RNA sequences of the *PSG8-AS1* orthologs

Gaps or nucleotide substitutions are white; matches in exons (as annotated in the human) are different colours (Exon 1 = red, Exon 2 = blue, Exon 3 = green, Exon 4 = dark blue). The white in the human represents gaps only.

A



B

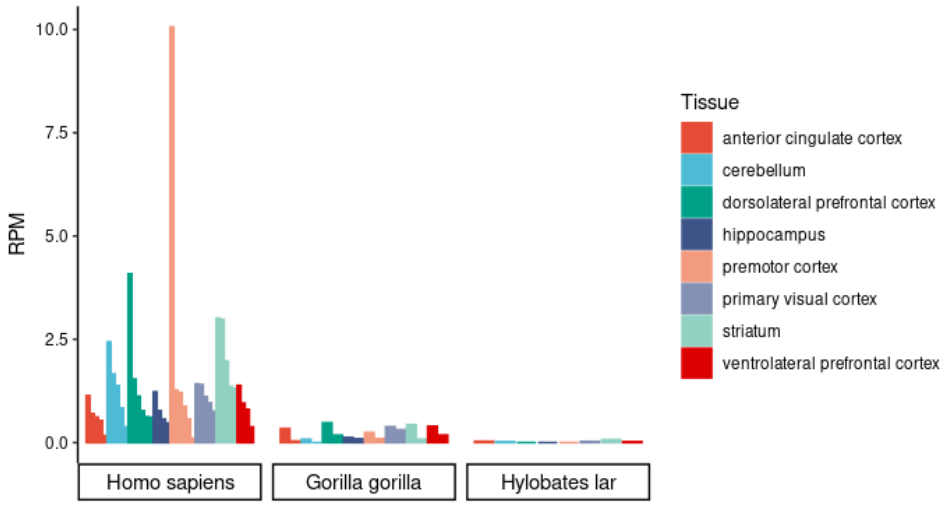


Figure 19. Barplots showing *PSG8-AS1* and ortholog expression levels in different primate species and tissues.

Expression units is Reads per Million (RPM). More than one bar in the same colour and species indicates different individuals. A) Expression data from different tissues, including the Central Nervous System (CNS), in human, gorilla, orangutan and macaque (GSE100796). Different colours represent distinct tissues. B) Expression data from different sections of the brain in human, gorilla, and gibbon (GSE30352). Different colours represent distinct brain regions.

4.3.3 *PSG8-AS1* may be involved in myelination

Since many novel genes important for human brain evolution are expressed in developmental stages, *PSG8-AS1* expression was inspected in the Allen Brain Atlas human brain development transcriptomic resource (https://portal.brain-map.org/), (Hawrylycz et al., 2012). This dataset includes 31 different time points pre- and post-birth from 26 brain regions. Levels of *PSG8-AS1* transcription were compared with specific brain cell markers (McKenzie et al., 2018) (Figure 20). Coincident with the start of oligodendrogenesis and the period of most rapid myelination in humans (25-27 postconception [pcw] weeks and 2 years after birth, respectively), *PSG8-AS1* expression peaks at the same time as oligodendrocyte markers *MBP* and *MOG* (de Faria et al., 2021; Ouyang et al., 2019). In contrast, astrocyte markers *GJA2* and *AQP4* start increasing before the initiation of gliogenesis at approximately 24 pcw, similar to neural markers *GAD2* and *SLC17A7.* Expression of the latter two markers declines around the start of the synaptic pruning process (2-3yrs) (Ouyang et al., 2019).

To get a clear picture of the genes that are co-expressed with *PSG8-AS1,* Weighted Gene Correlation Network Analysis (WGCNA) (Langfelder and Horvath, 2008) on the Allen Brain Atlas dataset was performed. This method suits the scientific question as it is used to find clusters or "modules" of highly correlated genes in a multidimensional transcriptomic dataset (Langfelder and Horvath, 2008). The genes most correlated with *PSG8-AS1* inside the module were *MBP*, *FOXO4* and *KIF1C* (Supplementary Table 21). A pathway enrichment analysis was conducted on the *PSG8-AS1* module using the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool (Dennis et al., 2003). The most significant pathways were related to myelination (Structural Constituent of Myelin Sheath, Myelin Sheath, Compact Myelin) and cellular

membrane composition (Integral Component of Membrane, Integral Component of Plasma Membrane, Plasma Membrane, Sphingolipid de novo Biosynthesis, Sphingolipid Metabolism) (Figure 21; Supplementary Table 22). Of note, sphingolipids are crucial for the synthesis and maintenance of the myelin sheath (Giussani et al., 2021). Although only marginally significant, Oligodendrocyte Differentiation was also enriched.

According to the LncATLAS resource (https://lncatlas.crg.eu/), *PSG8-AS1* is preferentially expressed in the nucleus in several cell lines (Mas-Ponte et al., 2017). Alignment data from the neuroblastoma SK-N-SH cell line from ENCODE, used in the LncATLAS study, shows signal only in the plus strand of the nuclear fraction (Figure 22). In the nucleus, lncRNAs recruit proteins for a variety of purposes (Yao et al., 2019). To investigate which RNA-binding proteins might interact with *PSG8-AS1*, catRapiD v2.1 predictor tool was used (Armaos et al., 2021). catRapiD v2.1 utilizes the nucleotide sequence and known interactions between protein domains and RNA to calculate an Interaction Propensity Score (IPS). Proteins with a IPS above 90 were selected for functional annotation analysis with the DAVID online tool (Figure 23A; Supplementary Table 23). Membrane protein transport and tight cell-cell junctions were among the most significant results (Figure 23B; Supplementary Table 24), which moderately resembles the set of pathways enriched in the WGCNA *PSG8-AS1* module. Interestingly, the third most correlated gene with *PSG8-AS1* in the WGCNA module was *KIF1C* after *MBP* and *FOXO4* (Supplementary Table 21), which encodes for a kinesin-like protein that participates in the transport of proteins through the Golgi system, predominantly in neurons (Lipka et al., 2016; Siddiqui et al., 2022). KIF1C was predicted to bind *PSG8-AS1* with a PRS of 80.08 (Supplementary Table 23), a value

that is still over the 3rd quantile in the IPS distribution (Figure 23A). Mutations in *KIF1C* have been associated with hereditary spastic paraplegia and cerebellar dysfunction in which focal cerebral white matter lesions occur (Dor et al., 2014; Marchionni et al., 2019; Oteyza et al., 2014).

Studies have shown that oligodendrocytes transport mRNAs and proteins to the myelin membrane through microtubule-based transport. These ribonucleotide protein complexes deliver mRNAs to the myelin compartment for local protein synthesis (Carson et al., 1997; Torvund-Jensen et al., 2018, 2014). In fact, in the mouse, *Mbp* mRNA is transported to the myelin sheath by a kinesin-dependent mechanism (Carson et al., 1997). In this way, *PSG8-AS1* may act in the nucleus as a link between mRNAs and ER/Golgi system proteins like KIF1C for their translation and transport to the membrane.

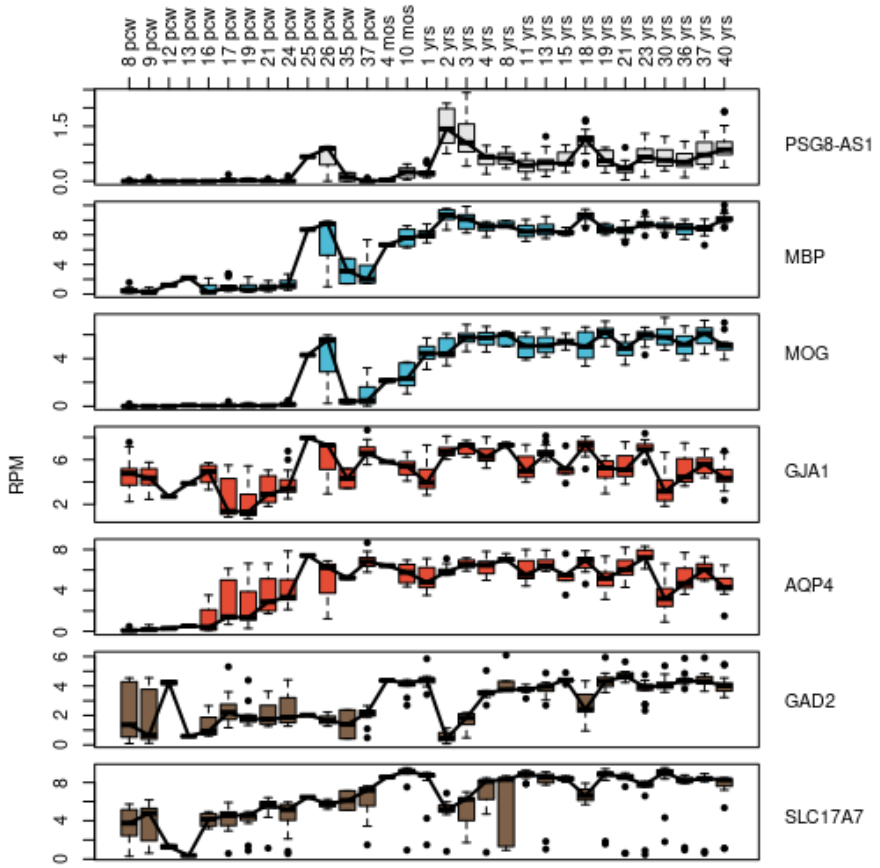


Figure 20. Expression of *PSG8-AS1* and several brain cell markers across brain developmental stages.

Expression units are Reads per Million (RPM). *PSG8-AS1* boxplots are shown in grey, oligodendrocyte markers in blue, astrocyte markers in red and neuron markers in brown.

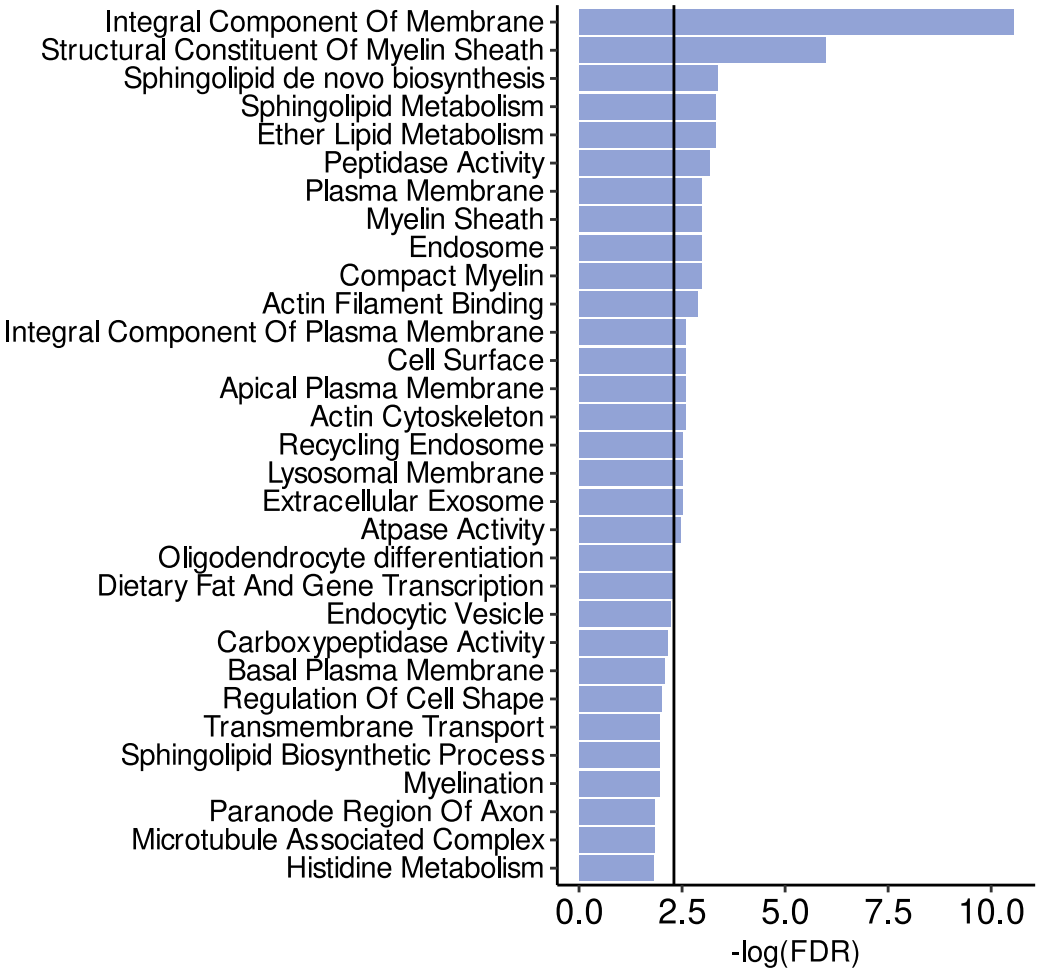


Figure 21. Barplot showing False Discovery Rate (FDR) of enriched pathways in the gene set enrichment analysis of the *PSG8-AS1* high-correlated gene expression module.

Pathway name format was simplified for visualization. Gene set enrichment analysis was performed with DAVID online tool.

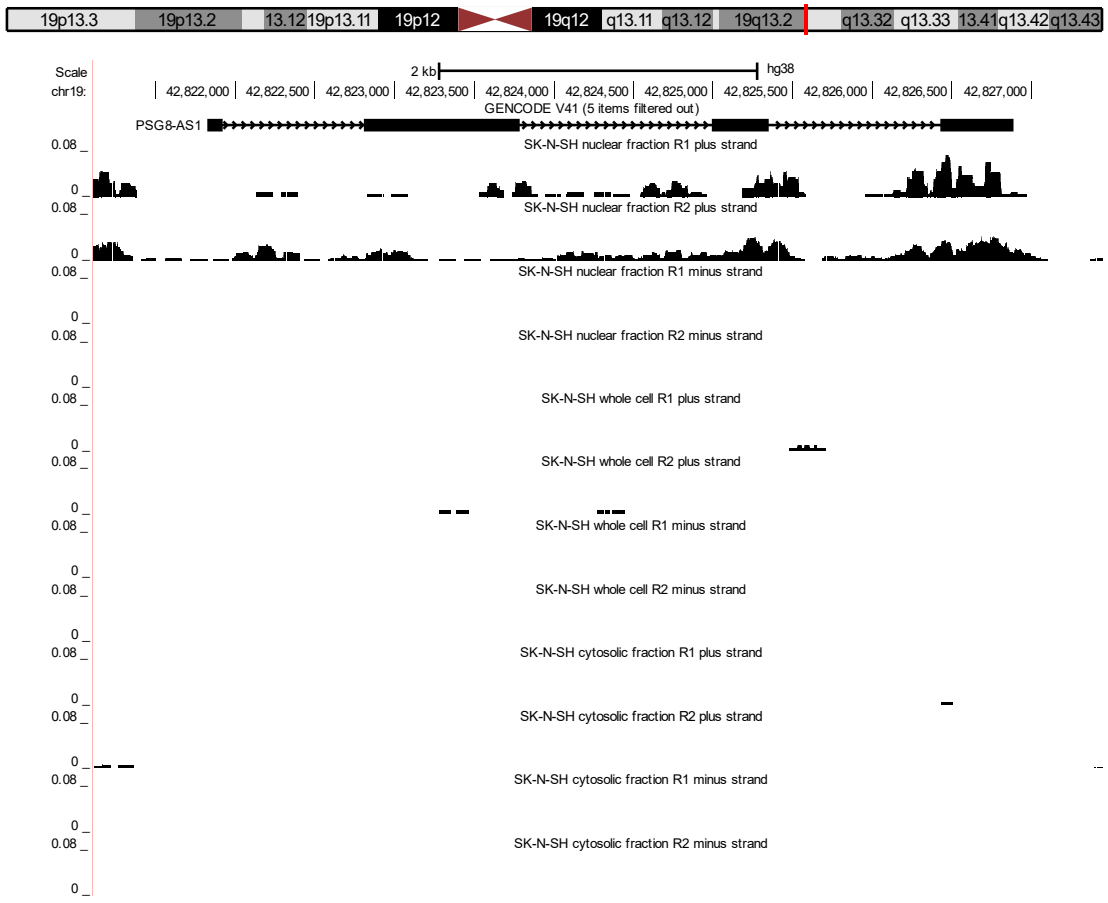
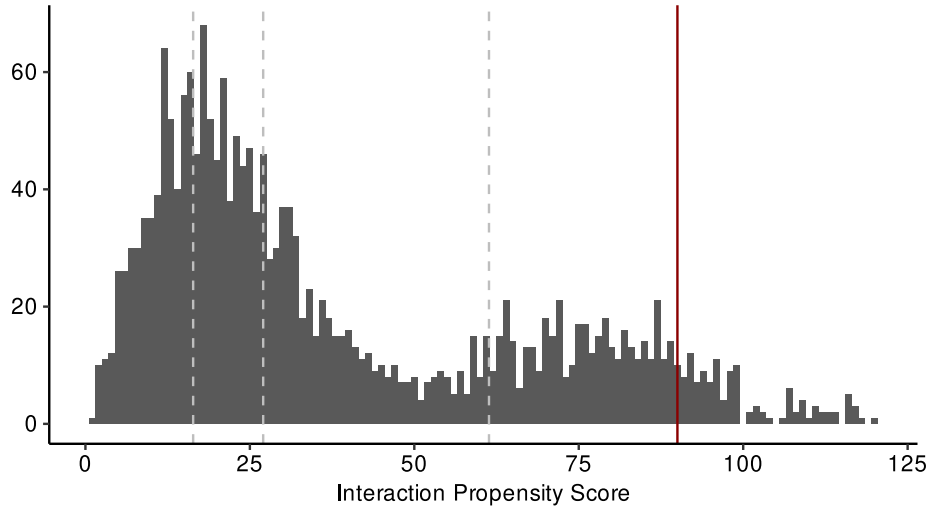


Figure 22. Alignment coverage density across *PSG8-AS1* in the SK-N-SH neuroblastoma cell line.

Genome version is hg38. *PSG8-AS1* is shown at the top. Boxes represent exons, while lines represent introns. The arrows indicate the *PSG8-AS1* is on the plus strand. SK-N-SH cell fractions shown are nuclear (top), whole (middle), and cytosolic (bottom). For each fraction, two isogenic replicates and strand specific reads are available. Although the peaks of read counts are enriched in exonic regions, some intronic noise signal can be observed, except for exon 1. It could be possible that, in cell line SK-N-SH, the transcription start site is located a few tens of base-pairs upstream of the reference.

A



B

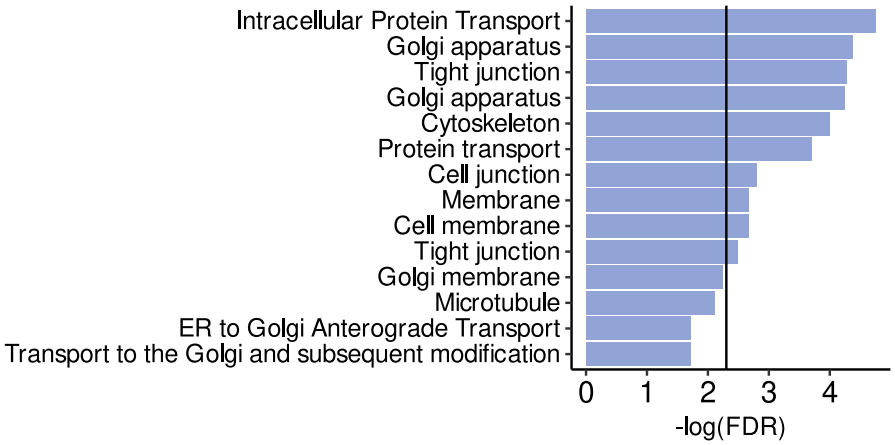


Figure 23. Prediction of protein interaction with *PSG8-AS1* results.

Prediction was performed with catRapiD v2.1. A) Distribution of IPS of the RNA-binding proteome to *PSG8-AS1*. 1st, 2nd, and 3rd quantile are highlighted in grey dotted lines. The filter threshold at 90 is indicated by a solid red line. B) Barplot showing False Discovery Rate (FDR) of enriched pathways in the gene set enrichment analysis performed with DAVID online tool of the highest scoring proteins predicted to interact with *PSG8-AS1*. Pathway name format was simplified for visualization.

4.3.4 *PSG8-AS1* as a potential biomarker in glioma

A biomarker refers to a measurable molecular feature that can be used to diagnose, classify, or predict the behaviour of a tumour (Kerr and Yang, 2021). Biomarkers in glioma can include genetic mutations, epigenetic alterations, changes in gene or protein expression, and changes in other features, such as histological parameters, or contrast enhancement on magnetic resonance imaging (MRI) (Lapointe et al., 2018; Louis et al., 2016). No substantial expression of *PSG8-AS1* in any other tumour or healthy tissue was observed in TCGA data (accessed through GEPIA 2, http://gepia2.cancer-pku.cn/). According to mutational TCGA data accessed through cBioPortal (https://www.cbioportal.org/), no germline or somatic mutations are found in *PSG8-AS1* in grade II and III oligodendroglioma and astrocytoma (low-grade gliomas or LGG) (Cerami et al., 2012; Gao et al., 2013; The Cancer Genome Atlas Research Network, 2015). Copy number alterations of *PSG8*-AS1 are found in around 1% of low-grade gliomas. However, copy number variants happen frequently in the *PSG* locus as it constitutes a segmental duplication, therefore, other copy number changes like double deletions or amplifications might be germline, non-tumour causing variants. The copy number of *PSG8-AS1* in IDHmt 1p/19q codeleted gliomas is always 1 in TCGA data, consistent with the location of this gene in 19q13.2 (Cerami et al., 2012; Gao et al., 2013; The Cancer Genome Atlas Research Network, 2015).

The RNA levels of *PSG8-AS1* in non-tumoral and tumoral tissue were assessed in two different datasets (GSE165595 and GSE59612, for details see Table 9). *PSG8-AS1* expression is lower in any subtype of glioma when compared with paired non-tumour brain tissue (Pairwise Wilcoxon Rank Sum Tests; “non-tumoral” vs. glioblastoma p=0.0037; “non-tumoral” vs. IDHmt p=0.0002; “non-tumoral” vs. IDH-mt 1p/19q

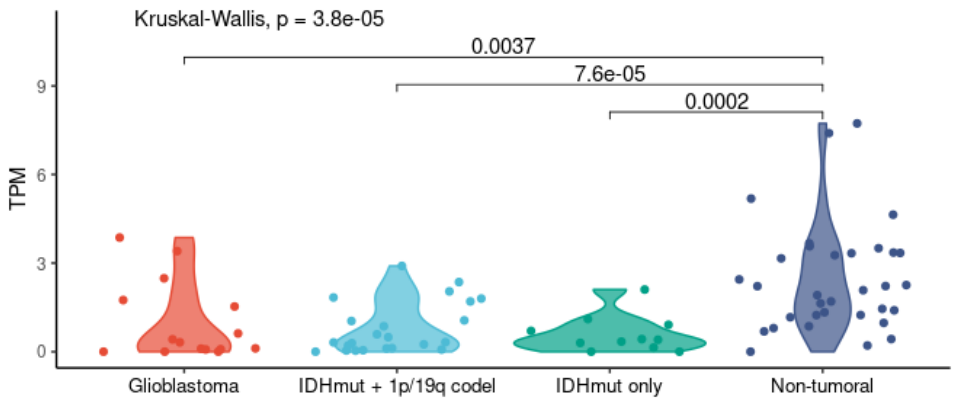
codeletion = 7.6e-05) (Figure 24A). Contrast enhancement on MRI, which is associated with the destruction of the blood-brain barrier and higher malignancy, is specific to high-grade gliomas (Gill et al., 2014; Pronin et al., 1997). When comparing non-neoplastic brain tissue with contrast-enhancing glioma and non-enhancing glioma, the expression of *PSG8-AS1* was only significantly lower in contrast-enhancing samples, meaning *PSG8-AS1* could be an indicator of glioma malignancy (Pairwise Wilcoxon Rank Sum Tests; non-neoplastic vs. contrast-enhancing glioma p=6.5e-08; non-neoplastic vs. non-enhancing glioma p=0.13; non-enhancing glioma vs contrast-enhancing glioma p=4.5e-09) (Figure 24B).

To investigate the potential clinical significance of *PSG8-AS1* downregulation, the association between *PSG8-AS1* state and survival was examined. Given mutations and gene copy number could not be used for survival analyses since they were the same in all samples, TCGA patients were stratified by molecular subtype and *PSG8-AS1* expression (high = top 25% and low = bottom 25% in each molecular subtype). Kaplan-Meier analysis and Log-rank tests were carried out to evaluate the effect on overall survival. In the IDH-mt 1p/19q-codeleted glioma, high expression levels of *PSG8-AS1* significantly correlated with longer patient survival time (Log-rank p = 0.019), whereas there was no effect on survival for the no-codeletion subtypes (Figure 25).

The results of these analyses suggest that *PSG8-AS1* levels might have diagnostic value in gliomas. Furthermore, *PSG8-AS1* expression could serve as a prognostic biomarker in IDHmt 1p/19q codeleted gliomas (oligodendrogliomas). The specific influence of *PSG8-AS1* expression on oligodendroglioma overall survival and no other gliomas support the previous evidence shown in this chapter that links *PSG8-AS1* with

oligodendrocyte function and implies that *PSG8-AS1* levels are reduced in oligodendroglial tumour-transformed cells with enhanced malignancy.

A



B

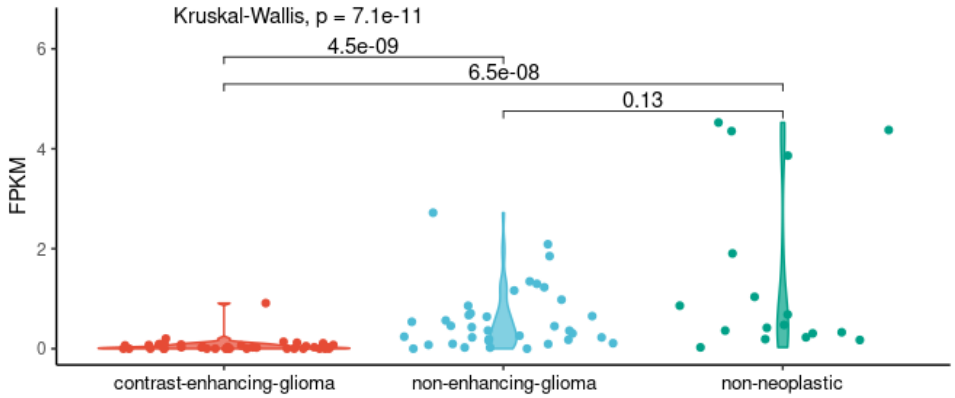


Figure 24. Expression of *PSG8-AS1* in tumours compared to paired non-tumoral tissue.

A) Violin plots showing Transcript Per Million (TPM) values from the GSE165595 dataset. B) Violin plots showing Fragments Per Kilobase of exon per Million mapped fragments (FPKM) values from the GSE59612 dataset. Individual values are shown in dots. Significant differences between the groups was assessed with a Kruskal–Wallis test (nonparametric one-way ANOVA).

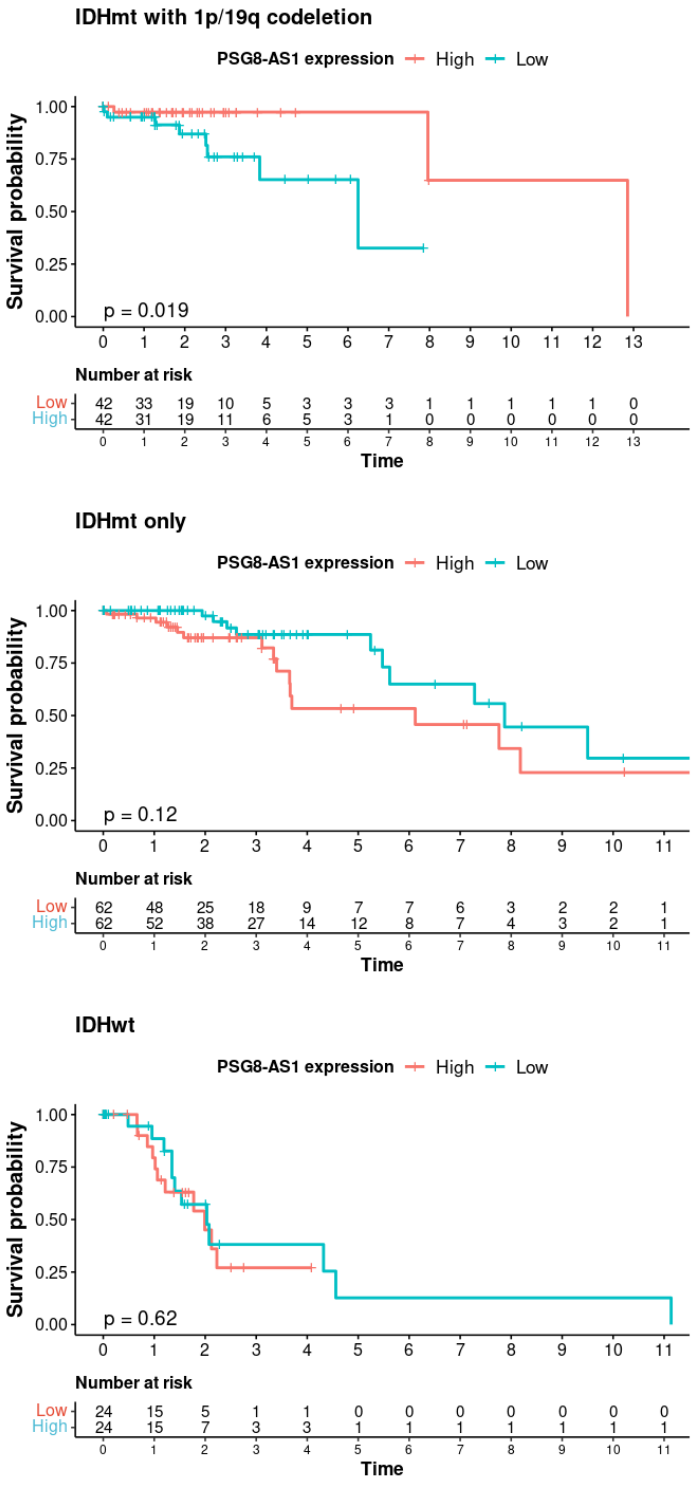


Figure 25. Kaplan-Meier survival curves in patients from the LGG TCGA study stratified by *PSG8-AS1* expression and molecular subtype.

Time is expressed in years.

4.4 Discussion

The aim of this study was to characterize *PSG8-AS1,* a lncRNA located in the rapidly evolving *PSG* locus, which harbours the placenta expressed Pregnancy-Specific Glycoprotein genes. The results demonstrate that *PSG8-AS1* exhibits a non-overlapping expression pattern with the coding *PSGs (PSG1-9, PSG11)* and is predominantly expressed in the brain. The initial characterization of *PSG8-AS1* in the brain included the analysis of its expression pattern in various cell types using single-cell RNA data. *PSG8-AS1* was found to be present at very low levels in most cells with predominant expression in oligodendrocytes. Several other cell types were also found to express *PSG8-AS1*, but only when oligodendrocyte markers were also present. This observation might be explained by an incorrect classification of, for example, a subgroup of cells as endothelial cells instead of oligodendrocytes. A major caveat of scRNA-seq is that the cell identity annotation is often made manually, which prevents reproducible annotations (Abdelaal et al., 2019; Ianevski et al., 2022). Transcription factors ZNF24 and TCF7L2, important regulators of oligodendrocyte differentiation and myelination in humans (Elbaz et al., 2018; Zhang et al., 2021), bind to putative enhancers of *PSG8-AS1,* which supports the idea that *PSG8-AS1* is an oligodendrocyte specific gene.

PSG8-AS1 was anticipated to be human-specific due to its localization to a segmental duplication, which are known to give rise to novel genes important to human brain evolution (Florio et al., 2018; Vollger et al., 2022). Conservation analyses here showed that *PSG8-AS1* emerged in the Catarrhini lineage, although its expression was primarily observed in humans. Examples of human brain-specific genes that are present in primates but only exhibit transcription in humans have been described previously

(Florio et al., 2018). Additionally, research on recent changes in human brain evolution has focused on neurons in the past, however, it is expected that neurons alone cannot explain the enhanced cognitive capabilities in humans (Berto et al., 2019; Donahue et al., 2018; Rilling and van den Heuvel, 2018; Sousa et al., 2017). According to a recent study, oligodendrocytes in the human brain exhibit more rapid evolution than neurons (Berto et al., 2019). Moreover, humans exhibit more complex structural and functional connectivity that relies on myelination compared to non-human primates, requiring a longer time for myelination to occur in humans (Donahue et al., 2018; Miller et al., 2012; Rilling and van den Heuvel, 2018). Oligodendrocytes constitute 75% of the non-neural cell populations in the human cortex (Pelvig et al., 2008). This, together with the growing evidence linking myelination and oligodendrocyte function to cognition and neuropsychiatric disorders, suggests that the characterization of human oligodendrocyte-specific genes such as *PSG8-AS1* might shed more light on the processes that underpinned the evolution of our unique cognitive capacities and that are dysregulated in human brain pathology (Fields et al., 2014; Maas et al., 2020; Mighdoll et al., 2015; Voineskos et al., 2013).

In this chapter, , *PSG8-AS1* was found to be co-expressed with oligodendrocyte markers during brain development. Genes highly correlated with *PSG8-AS1* were implicated in myelination and the composition of the plasma membrane (Figure 20, Figure 21). The pathways enriched in the putative *PSG8-AS1*-interacting proteome were associated with the plasma membrane, tight junctions, and protein transport across the Golgi systems (Figure 23B). One promising binding partner candidate, KIF1C, is encoded by the second most correlated gene with *PSG8-AS1* and was found in the putative *PSG8-AS1*-interacting proteome (Supplementary Table 23). Overall, the findings presented in

this chapter suggest that *PSG8-AS1* might participate in the transport of myelin sheath constituents to the cell membrane. The transport of the myelin membrane components and the capacity for coordinated membrane expansion are key elements in the mechanism of myelin assembly (Baron et al., 2015; Feldmann et al., 2011; Lam et al., 2022; White and Krämer-Albers, 2014). The mechanism by which *PSG8-AS1* might regulate this process remains unclear. As reported here, its predominant localization to the nucleus (Figure 22) indicates that *PSG8-AS1* might be involved in the regulation of gene expression or mRNA processing relating to proteins comprising myelin and membrane components (Mattick et al., 2023). Nevertheless, although not validated, the predicted interacting-proteome suggests that *PSG8-AS1* could have a role outside the nucleus. From an evolutionary perspective, the expression of *PSG8-AS1* specifically in human oligodendrocytes could be responsible, in part, for distinctive myelination mechanisms that may underpin unique human cognitive abilities, such as the timing and the pattern of myelination during development (Glasser et al., 2014).

One of the main limitations of this study is the inability to find genetic associations in the *PSG* locus. Repetitive genomic regions where alignment of reads is challenging or specificity of array probes is low, referred to as camouflaged genomic regions, are depleted of SNPs or CNVs associated with disease in GWAS studies. The absence of genetic association with disease is mainly due to the current bioinformatic methods inefficiently capturing camouflaged regions. However, disease-relevant genes in camouflaged regions exist, especially in humans, as seen in chapter 2 and previous evidence (Ebbert et al., 2019).

It was hypothesized that *PSG8-AS1* might be involved in glioma since the tumour suppressor transcription factor p53 binds a proximal enhancer. In addition, the codeletion of 19q, where *PSG8-AS1* is located, together with 1p, is used as a genetic marker for oligodendroglioma. Accordingly, *PSG8-AS1* expression showed diagnostic value for all subtypes glioma, and prognostic value specifically for oligodendroglioma.

Another limitation of this study is the absence of confirmatory studies in cell and animal models, or in human tissue samples. There are several reasons for this. First, the utilization of human oligodendrocyte cell lines HOG and MO3.13, a common practice in many previous studies of oligodendrocyte function, has been discouraged due their reduced myelination capacity under differentiating conditions (De Kleijn et al., 2019). On the other hand, glioma cell lines harbouring the IDH mutation and the 1p/19q codeletion are difficult to grow due to their slow doubling time. Only two other cell lines, BT054 and BT088, have the relevant genomic alterations. These cell lines were developed more than ten years ago but lack genomic and transcriptomic validation (Kelly et al., 2010). Secondly, the most fundamental aspects of human brain development and function are difficult to replicate in animal models (Zhao and Bhattacharyya, 2018). A negative result in a mouse model, for example, would not negate a role for *PSG8-AS1* in myelination in humans. In future research, utilization of oligodendrocyte progenitors derived from human induced pluripotent stem cells (hiPSCs) or human cerebral organoids together with the perturbation of *PSG8-AS1* expression should be considered to fully characterize this human oligodendrocyte-specific lncRNA (Lopez-Caraballo et al., 2020; Mahmoud et al., 2019; Pantoja et al., 2020; Xu et al., 2021).

In summary, this chapter provides the first *in-silico* characterization of *PSG8-AS1,* a human oligodendrocyte-specific lncRNA mapping to a segmental duplication and suggests that it may play a role in the regulation of several key signalling pathways involved in the transport of myelin and membrane components. The findings also suggest potential biomarker utility for *PSG8-AS1* in oligodendrogliomas. Future studies will be required to investigate and validate the precise mechanisms by which *PSG8-AS1* regulates these pathways and to explore its potential as a therapeutic target for oligodendrocyte-dysregulation-related disorders.

4.5 Methods

4.5.1 Gene expression datasets

Gene expression and exon usage data of the *PSGs* and non-coding *PSG8-AS1* in different adult tissues was accessed through the GTEx Portal (www.gtexportal.org). The GTEx consortium (V8) repository provides a collection of RNA-sequencing (RNA-seq) data from 54 tissues obtained from a total of 948 human donors. The expression data used were gene-level transcript per million (TPM) quantifications generated by the GTEx consortium (Lonsdale et al., 2013).

BrainSpan, an atlas of the developing human brain, which includes RNA-seq data from 31 different time points pre- and post-birth from 26 distinct brain regions obtained from a total of 42 human donors, was accessed through the Allen Brain Atlas portal (https://portal.brain-map.org/). Normalized expression data in RPKM (reads per kilobase of exon model per million mapped reads) generated by the BrainSpan resource was used for analyses (Miller et al., 2014).

The Cancer Genome Atlas (TCGA, PanCancer Atlas) Low Grade Gliomas (LGG) collection, comprising 514 LGG brain tumour samples subject to RNA-seq were accessed through *TCGAbiolinks* R package (Script 2) (The Cancer Genome Atlas Research Network, 2015).

(Script 2) TCGAbiolinks::GDCquery(project = CancerProject, data.category = "Transcriptome Profiling", data.type = "Gene Expression Quantification", sample.type = c("Primary Tumor"), workflow.type = "HTSeq - Counts")

Gene counts were normalized by the trimmed mean of M-values (TMM) method (Robinson and Oshlack, 2010). Briefly, the TMM method uses a reference sample (the sample with the median/mean library size) to calculate library size factors with the trimmed mean of the log-ratios of the counts between each sample and a reference sample. Then, the library size factors are used to adjust the counts for each gene, to account for differences in sequencing depth and composition between the samples. After TMM normalization of the gene counts, gene expression data was converted to Counts Per Million (CPM). Data processing was performed with *edgeR* R package.

Publicly available datasets that did not belong to a public consortium or atlas were accessed through the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) (Table 9).

Table 9. Gene expression datasets accessed through GEO.

ACC = GEO accession, EXP = Experiment type, snRNA-seq = single-nuclei RNA-seq, SPEC = Species, SEC = Section in the chapter. GEO = Expression format in GEO, NORM = Data transformation for analysis. TMM = trimmed mean of M-values normalization, RPM = reads per million, TPM = transcript per million, FPKM = Fragments Per Kilobase of transcript per Million mapped reads.

ACC	TISSUE	EXP	SPEC	SEC	N	GEO	NORM
GSE109082	Placenta	bulk RNA-seq	Human	4.3.1	39 individuals	Counts	TMM-RPM
GSE148241	Placenta	bulk RNA-seq	Human	4.3.1	32 individuals	Counts	TMM-RPM
GSE67835	Brain	scRNA-seq	Human	4.3.1	466 cells	Counts	Counts
GSE118257	Brain	snRNA-seq	Human	4.3.1	5 samples (6591 cells)	Counts	Counts
GSE100796	Brain and others	bulk RNA-seq	Primates	4.3.2	5 humans 6 chimps 2 gorillas 1 gibbon	Counts	RPM
GSE30352	Brain	bulk RNA-seq	Primates	4.3.2	9 humans 2 chimps 2 bonobos 2 gorillas 2 rhesus	Counts	RPM
GSE165595	Brain (glioma and non-tumour)	bulk RNA-seq	Human	4.3.4	41 patients (matched pair of tumour and normal samples	TPM	TPM
GSE59612	Brain (glioma and non-tumour)	bulk RNA-seq	Human	4.3.4	39 contrast-enhancing samples 36 non-enhancing glioma samples 17 non-neoplastic samples	FPKM	FPKM

4.5.2 Transcription factor binding and expression quantitative trait loci (eQTLs)

The University of California, Santa Cruz Genome Browser (UCSC Browser, https://genome.ucsc.edu/) was used to visualize the genomic location of regulatory regions and eQTLs potentially influencing *PSG8-AS1* expression in hg38. Candidate cis-regulatory elements (cCREs), which include promoters, enhancers and CTCF-binding elements were retrieved from the ENCODE v3 library (Moore et al., 2020). Transcription factor binding information from the same ENCODE library was integrated with the ORegAnno database (Lesurf et al., 2016). Transcription factors were classified by their summary in the GeneCards database (https://www.genecards.org/). If the central nervous system was mentioned, the protein was classified as CNS regulation (meaning it may not have a fundamental role in the CNS, however, it has some level of influence). If the REST complex was mentioned, the protein was classified as REST complex. eQTLs were obtained from GTEx and PhychENCODE resources (Akbarian et al., 2015; Lonsdale et al., 2013).

4.5.3 Conservation

The online tool *blastn* was used to retrieve potential *PSG8-AS1* orthologs. Sequences aligning to human *PSG8-AS1* RNA sequence (NR_036584.1) were filtered by at least 80% alignment span and similarity. Synthetic constructs were excluded. Status of the genome sequence of *PSG8-AS1* across primate species was explored in the UCSC browser Mammals Multiz Alignment & Conservation (27 primates) track. Briefly, they used a Large-Scale Genome Alignment Tool (lastz) to generate pairwise alignments with the human genome (Kent, 2002). Pairwise alignments are then organized into gapless “chains” and scored by a Nearest Neighbour Algorithm (Chiaromonte et al., 2002). Alignment “nets” show the highest-scoring chain at the top, filling the gaps with

lower-scoring chains (Kent, 2002). This track is convenient for exploring orthologous regions and genome rearrangements across species. The UCSC browser search result was modified to highlight rearrangements occurring in apes. Only the highest-scoring chain is shown in non-ape species for simplicity.

Multi-Alignment of the RNA sequences of *PSG8-AS1* orthologs was performed with the *msa* package in R, using the ClustalW algorithm (Thompson et al., 1994). Expression of *PSG8-AS1* orthologs was assessed by the realignment of raw sequencing reads from GSE100796 and GSE30352 (SRP111096 and SRP007412 Sequence Read Archive, respectively) to the genomes of the human (hg38), gorilla (gorGor6), orangutan (ponAbe3), gibbon (nomLeu3) and rhesus macaque (rheMac10). In the case of the rhesus macaque, the gene annotation file was modified to incorporate the putative *PSG8-AS1* ortholog. Gene read counts were converted to reads per million (RPM) in R.

4.5.4 Weighted gene co-expression network analysis

A weighted co-expression network was constructed using the *WGCNA* package in R (Langfelder and Horvath, 2008). Genes with extremely low expression were removed (genes were kept when the expression was higher than 0.01 RPKM across two thirds of the samples), resulting in 24,267 remaining genes. Soft-thresholding power of 7 was selected. The adjacency matrix was calculated using the Pearson correlation coefficient, and the topological overlap matrix (TOM) was computed to identify highly connected modules. The modules were identified using the dynamic tree-cutting algorithm with a minimum module size of 30 and a merge cut height value of 0.25. A total of 48 modules

were identified. The *PSG8-AS1* containing module included a total of 365 genes (Supplementary Table 21).

4.5.5 Prediction of binding partners

The online tool catRAPID omics V2.1 (http://www.tartaglialab.com/) was used to generate a predicted *PSG8-AS1*-binding proteome (Armaos et al., 2021). catRAPID omics V2.1 computes an interaction propensity score (IPS) using the transcript sequence as an input, based on secondary structures and physiochemical properties of precompiled RNA-binding motif libraries. They obtained the RNA-binding proteome from UniprotKB/Swiss-Prot 2020_05 database.

4.5.6 Functional annotation

The functional annotation online tool from the Database for Annotation, Visualization and Integrated Discovery (DAVID) (https://david.ncifcrf.gov/) (Dennis et al., 2003) was used to perform gene-set enrichment analyses. Briefly, EASE Score tests (which are Fisher Exact tests that eliminate 1-hit gene-lists), are performed to assess enrichment of genes of a particular pathway in a submitted list compared to a background list. P-values are then adjusted by Bonferroni, Benjamin, and False Discovery Rate corrections. For the functional annotation of the *PSG8-AS1* WGCNA module, the default background was utilized. For the functional annotation of the predicted *PSG8-AS1*-binding proteome, the background gene set was changed to the RNA-binding proteome from UniprotKB/Swiss-Prot 2020_05.

4.5.7 Glioma diagnostic and survival analyses

Overall expression of *PSG8-AS1* across tumours was examined in GEPIA 2 online tool (http://gepia2.cancer-pku.cn). To evaluate if *PSG8-AS1* was downregulated in glioma,

the expression in tumour and non-tumoral tissue from glioma patients and controls was compared in datasets GSE165595 and GSE59612. Several glioma subtypes were included to assess the diagnostic value across different molecular subtypes of brain tumours. The Kruskal-Wallis Test (a non-parametric alternative to one-way ANOVA test) followed by Pairwise Wilcoxon Rank Sum Tests were performed to evaluate significant differences between the groups.

To examine prognosis, clinical data from the TGCA was accessed with the *TCGAbiolinks* R package (Script 3).

(Script 3) GDCquery_clinic(project = "TCGA-LGG", type = "clinical")

Samples were kept only if there was information about their gene expression and clinical status. Samples were classified based on their molecular subtype (IDH-wt, IDH-mt alone and IDH-mt with 1p/19 codeletion) and their expression of *PSG8-AS1*. For each molecular subtype, the samples at the top 25% of *PSG8-AS1* expression were termed “High”, and the bottom 25% were termed “Low”. Kaplan Meier curves and Log-rank tests were then created and visualized with *survival* and *survminer* R packages.

4.5.8 Statistical analyses

Statistical analyses were performed with the *stats* package in R.

Chapter 5: General discussion

The initial release of the human genome drafts in 2001 marked the start of a new era in disease genetics, catalysing significant advancements (Lander et al., 2001; Venter et al., 2001). Within a mere two decades, the vast genomic data amassed from millions of individuals has enabled a profound understanding of the genetics behind numerous traits, encompassing complex and polygenic characteristics like height and body mass index (BMI), as well as clinical phenotypes such as the activation of somatic mutations driving tumour development and germline genomic rearrangements associated with developmental delay and psychiatric traits (Grotzinger et al., 2022; Lello et al., 2018; Ostroverkhova et al., 2023; Yengo et al., 2018). Nevertheless, our current understanding of the genome remains incomplete. Despite the existence of an enormous amount of publicly available data, we continue to grapple with achieving the anticipated outputs (Ebbert et al., 2019).

Significant advances have been made in disease treatment. For example, the identification of HLA alleles associated with agranulocytosis risk following treatment with the most effective antipsychotic drug in clinical use (Clozapine) (Konte et al., 2021). Other areas of research, such as the study of the contribution of CNVs to schizophrenia, have not been able to report novel genome-wide significant loci despite increasing sample size substantially. The latest large-scale study of CNVs in schizophrenia identified 6 novel loci with nominal levels of significance that were validated with ddPCR. One of the other major observations of that study was that

unreported schizophrenia-relevant CNVs are predominantly mediated by NAHR. While they hypothesize novel CNVs are thus enriched in regions with flanking segmental duplications, CNVs overlapping 50% of their length with a segmental duplication were not analysed (Marshall et al., 2017). This issue frustrates a resolution of the challenge of the missing heritability in schizophrenia and possibly other psychiatric disorders (Owen and Williams, 2021). Camouflaged regions, exemplified by segmental duplications in this thesis, have been sequenced and captured in genomic arrays yet systematically overlooked in previous investigations with respect to their contribution to disease (Baba et al., 2019; Cooper et al., 2011; Kirov et al., 2009; Marshall et al., 2017; Mulle et al., 2014, 2010; Niarchou et al., 2019; Quintero-Rivera et al., 2010; Rees et al., 2014c; Rees and Kirov, 2021). Another constraint to the discovery of new genomic regions that contribute to susceptibility is the lack of functional characterization of non-coding regions. The majority of the genetic association signal for psychiatric disease is located to non-coding regions, and while techniques such as fine-mapping help to characterize the biological consequences of variants discovered through GWAS, the impact of non-coding regions in CNVs remains difficult to interpret (Hattori et al., 2003; Marshall et al., 2017; Trubetskoy et al., 2022; Umeton et al., 2022; Xiao et al., 2017).

The characterization of the molecular functions of genes within their cellular contexts are frequently instigated from their genetic association with disease (Grotzinger et al., 2022; Ostroverkhova et al., 2023). Because camouflaged regions are routinely overlooked in genetic studies, it is expected that the investigation of the molecular function of genes inside these regions is also neglected (Ebbert et al., 2019). Admittedly, discerning the molecular role of non-coding elements is not as

straightforward as the exploration of the function of coding genes. As a consequence, this perpetuates the cycle of scientific disregard, whereby "difficult genetic regions" (represented by the camouflaged and the non-coding genome in this thesis), are systematically neglected across many fields of study, including genetics and molecular biology.

The work reported in this thesis attempted to address this problem using three different approaches, as presented in chapters 2 - 4. (Figure 26). In chapter 2, camouflaged regions, represented by CNVs overlapping a segmental duplication by more than 50% of their length, were examined for the first time in the context of schizophrenia genetic association studies. A 500 Kb recurrent duplication encompassing the dopamine receptor 5 gene (*DRD5*), which had only been reported once before in a pedigree study, was found in 7 schizophrenia cases and 0 controls (from a total of 8,261 schizophrenia cases and 16,410 controls). Then, in chapter 3, the concept of constraint against genetic variation, which represents the biological significance of a genomic region, is used to determine whether non-coding regions disrupted in CNVs in schizophrenia cases are more highly constrained than CNVs in non-psychiatric controls (Chen and The gnomAD consortium, 2022). Only genic-centric measures of constraint, such us the LoF-intolerant gene classification, had been examined previously in schizophrenia. Using a dataset comprising 14,155 schizophrenia cases and 21,660 controls, non-coding regions were shown to be more highly constrained in CNVs in schizophrenia cases compared to controls, and non-genic CNVs, which were believed to be unimportant, were found to show higher constraint, especially in regulatory regions reported by ENCODE (cCREs) (Marshall et al., 2017; Singh et al., 2017). Overall, the findings of this thesis offer several new opportunities in the field of psychiatric genetics.

First, the results from chapter 2 indicate that the development of new techniques that allow for the detection of all structural variants will facilitate the discovery of novel rare variants implicated in schizophrenia. In chapter 3, non-coding functional regions disrupted by CNVs, even those without coding genes, are shown to be important contributors to risk, which suggests that these elements should be accounted for in future studies. These new perspectives should also be applied to other psychiatric disorders, which may shed more light into the shared and differential pathophysiology of these disorders and promote the development of new therapeutics.

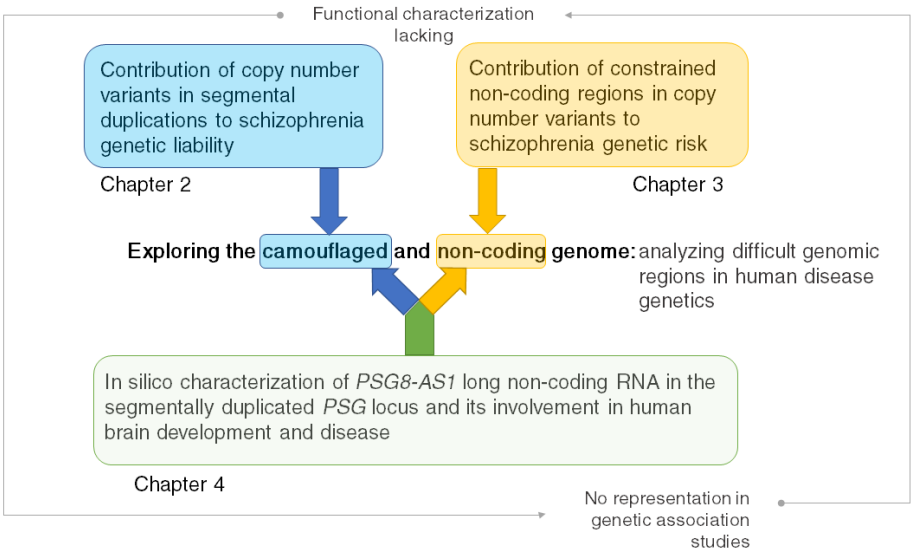


Figure 26. Overview of thesis structure.

Chapters 2 and 3 focused on genetic associations with schizophrenia; chapter 4 reports the analysis of a lncRNA with a putative function in myelination in the human brain. The themes of the three chapters cohere around the general themes of the camouflaged and the non-coding genome.

A different approach was undertaken in chapter 4. Segmental duplications contribute to the generation of novel genes in the human lineage essential for brain evolution (Florio et al., 2018; Mangan et al., 2022; Nuttle et al., 2016; Vollger et al., 2022). The segmentally duplicated *PSG* locus is probably the most rapidly evolving locus in the human genome (Sudmant et al., 2015), therefore, it was hypothesized that the previously uncharacterized *PSG8-AS1* lncRNA could be important for human brain function and disease. Using publicly available transcriptomic datasets, a human and oligodendrocyte specific pattern of expression was found for *PSG8-AS1*, with high correlation with the expression of myelin components. The levels of *PSG8-AS1* mRNA were found to have diagnostic value in gliomas and prognostic value in oligodendrogliomas. Myelination is a process linked with cognitive abilities, which happens in a distinct manner in humans compared to other hominids (Glasser et al., 2014). However, molecular characterization of *PSG8-AS1* could not be accomplished due to lack of suitable cell, animal or clinical models. Nevertheless, chapter 4 emphasises the need for the development of human models (such as patient-derived cell lines and organoids) in research on human brain development, which would complement the study of candidate genes identified in human disease genetics (Shaker et al., 2021).

Given that complications during pregnancy have been implicated as environmental risk factors for schizophrenia, one of the preliminary hypotheses explored in this thesis was whether the segmentally duplicated *Pregnancy-Specific Glycoprotein* (*PSG*) locus has CNVs that exhibited associations with schizophrenia (Abel et al., 2010; Cannon et al., 2002; Kotlicka-Antczak et al., 2014; Rubio-Abadal et al., 2015). Notably, this locus has already been associated with preeclampsia and other pregnancy complications (Arnold

et al., 1999; Chang et al., 2016; Gormley et al., 2017; Hertz and Schultz-Larsen, 1983; MacDonald et al., 1983; Rattila et al., 2019; Temur et al., 2020; Zhao et al., 2012). Despite these intriguing connections, the *PSG* locus exhibits frequent copy number variation and filtering for rare CNV events excluded many events from the *PSG* locus in chapters 2 and 3. Remaining events did not show association. While another possibility is that common CNVs in the *PSG* locus showed association with schizophrenia, this would have already been captured by GWAS SNPs, which has not been the case (Trubetskoy et al., 2022). Collectively, these findings indicate that the *PSG* locus, although holding promise in terms of its potential role in pregnancy-related disorders, does not appear to contribute substantially to the genetic architecture of schizophrenia through CNV associations.

In this thesis, the context and limitations of each chapter are reviewed in their respective discussions. More generally, the findings reported here should be replicated in more highly powered datasets, and in a variety of European and other populations. For genomics to have a significant impact on human health, the entire spectrum of human genetic diversity needs to be examined for disease associations (Peterson et al., 2019). Additionally, all bioinformatic findings, including those in this thesis, require independent experimental validation.

A human-specific perspective has been adopted in this thesis. Many of the recently evolved genomic elements that shape the unique characteristics that define our species are located in camouflaged regions, such as segmental duplications, and/or non-coding regions (Florio et al., 2018; Mangan et al., 2022; Nuttle et al., 2016; Vollger et al., 2022). In the field of psychiatry, exploring human-specific genes and non-coding

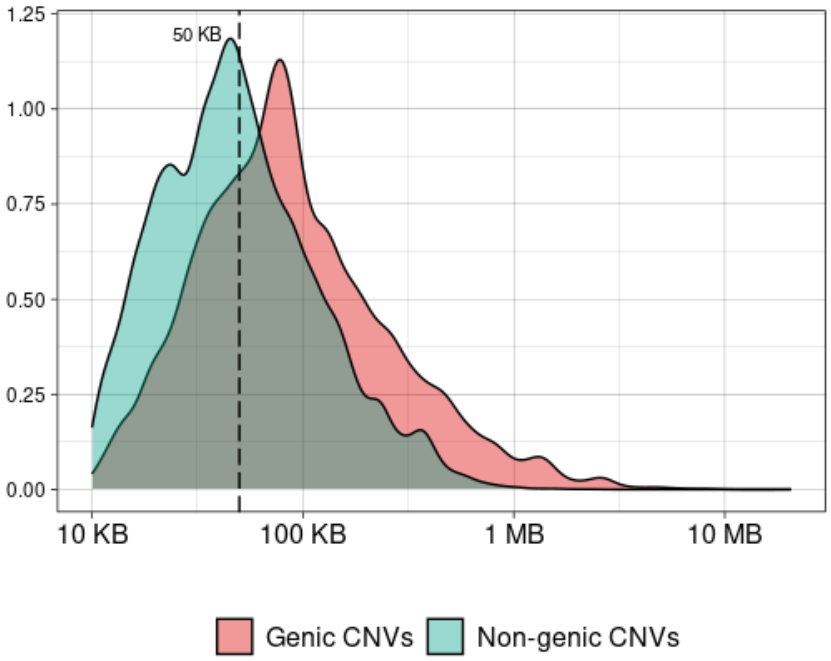
regions is particularly important as it enables us to investigate psychiatric traits that are exclusive to humans. While numerous psychiatric disorders exhibit similar symptoms across species, certain conditions such as schizophrenia and bipolar disorder are only observed in humans. However, it is worth noting that primates and other mammals do display psychiatric-like traits, including anxiety, depression, and abnormal social behaviors (Levchenko et al., 2023; Scheepers et al., 2018). By studying these shared and non-shared traits, we can gain valuable comparative perspectives on the underpinnings of the biological, social, and cognitive factors that shape our mental well-being.

The results in this thesis support the relevance of previously neglected regions to schizophrenia genetic liability. Furthermore, a human oligodendrocyte-specific lncRNA in such a camouflaged region was implicated in human brain development. Further investigation of the genetics and molecular biology of such camouflaged genomic regions may complement each other in the future, leading to novel findings relevant to disease.

Appendix: Supplementary material

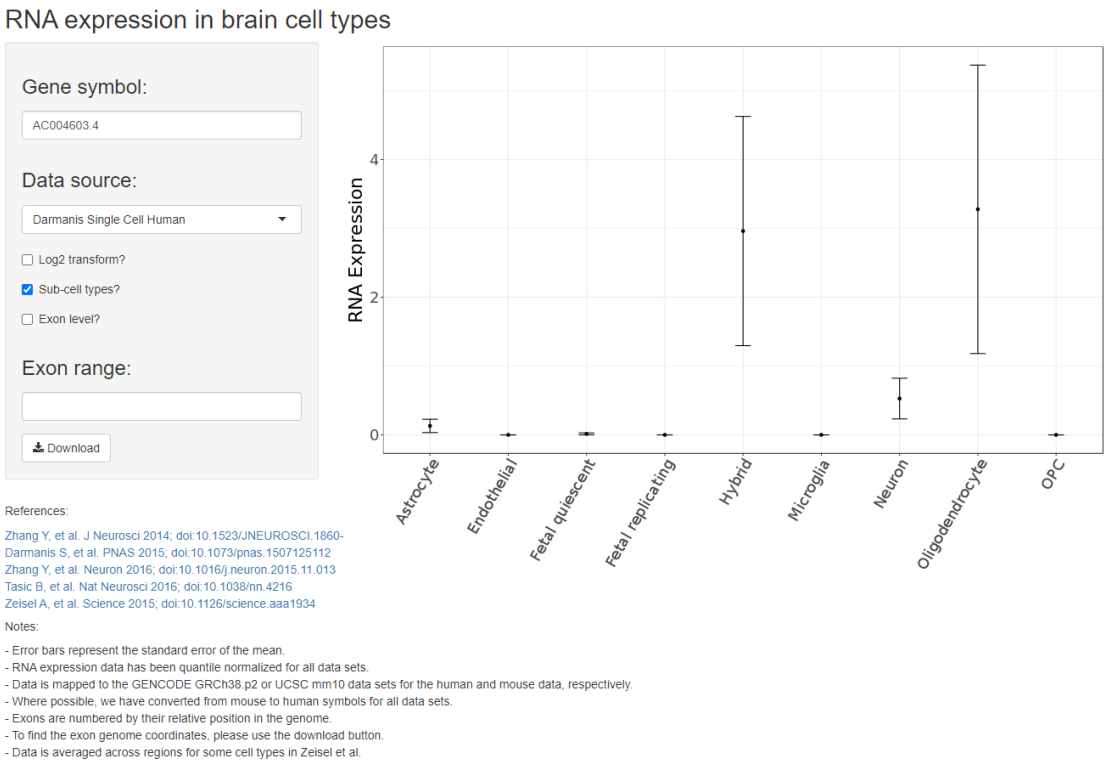
Supplementary Figures

Chapter 3



Supplementary Figure 1. Length distribution comparison between genic and non-genic CNVs.

The density area before the 50 kb threshold is substantially bigger for non-genic CNVs than for genic CNVs.



Supplementary Figure 2. Validation of *PSG8-AS1* expression specificity in oligodendrocytes

The Brain Cell Type Specific Gene Expression R/Shiny tool was used (http://celltypes.org/brain/) (McKenzie et al., 2018). AC004603.4 is a previous gene name for *PSG8-AS1*.

Supplementary Tables

Chapter 2

Supplementary Table 1. MGS dataset. Results of the gene-wise analysis of duplications.

Line in bold marks p value < 0.05. NAFF = number of affected individuals; NUNAFF = number of unaffected individuals (controls); FAFF frequency in affected individuals; FUNAFF = frequency in unaffected individuals (controls); OR = Odds ratio; CI1 – 95% lower limit confidence interval, CI2 – 95% upper limit confidence interval; PVAL = p value; No = number.

GENE	NAFF	NUNAFF	FAFF	FUNAFF	OR	CIL	CIU	PVAL	NO
DPY19L2	29	13	0.013	0.005	9.64	2.19	4.75E+01	2.44E-03	1
CKMT1A	5	0	0.002	0.000	285.60	1.70	2.22E+07	2.70E-02	2
SYT15	6	17	0.003	0.007	0.10	0.01	7.81E-01	2.70E-02	3
SYT15B	6	17	0.003	0.007	0.10	0.01	7.81E-01	2.70E-02	4
RNASE3	12	8	0.005	0.003	8.59	1.13	7.22E+01	3.74E-02	5
CATSPER2	7	1	0.003	0.000	35.44	1.01	6.52E+03	4.92E-02	6
NIPA1	0	4	0.000	0.002	0.01	0.00	1.01E+00	5.07E-02	7
TYW1B	3	11	0.001	0.004	0.08	0.00	1.13E+00	6.27E-02	8
DGCR6	17	9	0.008	0.004	5.54	0.89	3.94E+01	6.60E-02	9
ENSG00000283809	17	9	0.008	0.004	5.54	0.89	3.94E+01	6.60E-02	10
PRODH	17	9	0.008	0.004	5.54	0.89	3.94E+01	6.60E-02	11
EFL1	2	0	0.001	0.000	208.38	0.64	1.76E+07	7.19E-02	12
SAXO2	2	0	0.001	0.000	208.38	0.64	1.76E+07	7.19E-02	13
TEX101	4	0	0.002	0.000	95.60	0.55	7.37E+06	9.04E-02	14
ZDHHC11	21	15	0.010	0.006	3.69	0.79	1.81E+01	9.63E-02	15
DEFB130A	10	7	0.005	0.003	6.27	0.70	6.26E+01	1.01E-01	16
FAHD2B	0	4	0.000	0.002	0.01	0.00	2.97E+00	1.36E-01	17
DRD5	2	0	0.001	0.000	93.61	0.23	8.36E+06	1.47E-01	18
UGT2B28	2	0	0.001	0.000	93.61	0.23	8.36E+06	1.47E-01	19
FAM90A1	8	18	0.004	0.007	0.26	0.03	1.68E+00	1.61E-01	20
PSG8	2	6	0.001	0.002	0.10	0.00	2.51E+00	1.67E-01	21
DEFB103B	6	13	0.003	0.005	0.22	0.02	1.87E+00	1.70E-01	22
DEFB104B	6	13	0.003	0.005	0.22	0.02	1.87E+00	1.70E-01	23
DEFB105B	6	13	0.003	0.005	0.22	0.02	1.87E+00	1.70E-01	24
DEFB106B	6	13	0.003	0.005	0.22	0.02	1.87E+00	1.70E-01	25
DEFB107B	6	13	0.003	0.005	0.22	0.02	1.87E+00	1.70E-01	26
PRR23D1	6	13	0.003	0.005	0.22	0.02	1.87E+00	1.70E-01	27
SPAG11B	6	13	0.003	0.005	0.22	0.02	1.87E+00	1.70E-01	28
PDPR	5	2	0.002	0.001	10.37	0.35	5.82E+02	1.78E-01	29
USP17L3	3	9	0.001	0.004	0.15	0.01	2.33E+00	1.83E-01	30
USP17L8	3	9	0.001	0.004	0.15	0.01	2.33E+00	1.83E-01	31
ZNF705B	3	9	0.001	0.004	0.15	0.01	2.33E+00	1.83E-01	32

H3-7	1	5	0.000	0.002	0.07	0.00	3.34E+00	1.90E-01	33
DEFB4B	6	12	0.003	0.005	0.25	0.02	2.18E+00	2.12E-01	34
ZNF705G	6	12	0.003	0.005	0.25	0.02	2.18E+00	2.12E-01	35
CLEC18B	1	0	0.000	0.000	64.05	0.07	6.17E+06	2.25E-01	36
CTAGE15	1	0	0.000	0.000	64.05	0.07	6.17E+06	2.25E-01	37
NPIPB15	1	0	0.000	0.000	64.05	0.07	6.17E+06	2.25E-01	38
DEFB103A	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	39
DEFB104A	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	40
DEFB105A	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	41
DEFB106A	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	42
DEFB107A	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	43
FAM90A14	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	44
FAM90A16	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	45
FAM90A17	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	46
FAM90A18	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	47
FAM90A19	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	48
FAM90A22	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	49
FAM90A23	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	50
FAM90A7	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	51
FAM90A8	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	52
FAM90A9	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	53
PRR23D2	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	54
SPAG11A	8	16	0.004	0.006	0.32	0.04	2.13E+00	2.41E-01	55
FAM90A20P	5	10	0.002	0.004	0.25	0.02	2.68E+00	2.57E-01	56
USP17L1	5	10	0.002	0.004	0.25	0.02	2.68E+00	2.57E-01	57
USP17L4	5	10	0.002	0.004	0.25	0.02	2.68E+00	2.57E-01	58
DEFB4A	8	16	0.004	0.006	0.35	0.04	2.33E+00	2.81E-01	59
ACSM2A	0	2	0.000	0.001	0.03	0.00	1.32E+01	2.88E-01	60
GYPA	2	6	0.001	0.002	0.17	0.00	4.30E+00	2.91E-01	61
FAM90A11P	3	8	0.001	0.003	0.22	0.01	3.62E+00	2.99E-01	62
FAM90A12P	3	8	0.001	0.003	0.22	0.01	3.62E+00	2.99E-01	63
FAM90A24P	3	8	0.001	0.003	0.22	0.01	3.62E+00	2.99E-01	64
CD177	2	0	0.001	0.000	24.63	0.08	2.08E+06	3.05E-01	65
DTX2	0	1	0.000	0.000	0.05	0.00	4.10E+01	3.81E-01	66
FKBP6	0	1	0.000	0.000	0.05	0.00	4.10E+01	3.81E-01	67
NAT8B	0	1	0.000	0.000	0.05	0.00	4.10E+01	3.81E-01	68
NSUN5	0	1	0.000	0.000	0.05	0.00	4.10E+01	3.81E-01	69
PGM5	0	1	0.000	0.000	0.05	0.00	4.10E+01	3.81E-01	70
PRAMEF14	0	1	0.000	0.000	0.05	0.00	4.10E+01	3.81E-01	71
PRAMEF17	0	1	0.000	0.000	0.05	0.00	4.10E+01	3.81E-01	72
PRAMEF19	0	1	0.000	0.000	0.05	0.00	4.10E+01	3.81E-01	73
PRAMEF20	0	1	0.000	0.000	0.05	0.00	4.10E+01	3.81E-01	74
SIMC1	0	1	0.000	0.000	0.05	0.00	4.10E+01	3.81E-01	75
TRIM50	0	1	0.000	0.000	0.05	0.00	4.10E+01	3.81E-01	76

UPK3B	0	1	0.000	0.000	0.05	0.00	4.10E+01	3.81E-01	77
HNRNPCL1	2	5	0.001	0.002	0.22	0.00	6.35E+00	3.83E-01	78
POMZP3	6	10	0.003	0.004	0.36	0.03	3.50E+00	3.83E-01	79
SPDYE16	6	10	0.003	0.004	0.36	0.03	3.50E+00	3.83E-01	80
ANKRD36B	4	3	0.002	0.001	3.82	0.13	1.30E+02	4.27E-01	81
PSG11	15	20	0.007	0.008	0.59	0.12	2.79E+00	5.07E-01	82
POTED	0	2	0.000	0.001	0.12	0.00	4.04E+01	5.22E-01	83
TP53TG3	4	3	0.002	0.001	2.77	0.10	9.57E+01	5.46E-01	84
TP53TG3F	4	3	0.002	0.001	2.77	0.10	9.57E+01	5.46E-01	85
CST1	1	0	0.000	0.000	7.58	0.01	7.30E+05	5.67E-01	86
DEFA3	1	0	0.000	0.000	7.58	0.01	7.30E+05	5.67E-01	87
ENSG00000284337	1	0	0.000	0.000	7.58	0.01	7.30E+05	5.67E-01	88
FAM90A13P	1	0	0.000	0.000	7.58	0.01	7.30E+05	5.67E-01	89
FAM90A15P	1	0	0.000	0.000	7.58	0.01	7.30E+05	5.67E-01	90
FAM90A3P	1	0	0.000	0.000	7.58	0.01	7.30E+05	5.67E-01	91
FAM90A5P	1	0	0.000	0.000	7.58	0.01	7.30E+05	5.67E-01	92
GPR42	1	0	0.000	0.000	7.58	0.01	7.30E+05	5.67E-01	93
MALL	1	0	0.000	0.000	7.58	0.01	7.30E+05	5.67E-01	94
RGPD5	1	0	0.000	0.000	7.58	0.01	7.30E+05	5.67E-01	95
TP53TG3B	1	0	0.000	0.000	7.58	0.01	7.30E+05	5.67E-01	96
TP53TG3C	1	0	0.000	0.000	7.58	0.01	7.30E+05	5.67E-01	97
TP53TG3E	1	0	0.000	0.000	7.58	0.01	7.30E+05	5.67E-01	98
TRIM77	1	0	0.000	0.000	7.58	0.01	7.30E+05	5.67E-01	99
CROCC	18	18	0.008	0.007	1.55	0.33	7.28E+00	5.73E-01	100
PSG7	17	22	0.008	0.009	0.66	0.15	2.87E+00	5.79E-01	101
MGAM	6	4	0.003	0.002	2.20	0.13	4.48E+01	5.83E-01	102
GOLGA8A	8	11	0.004	0.004	0.63	0.07	5.06E+00	6.62E-01	103
GYPB	6	6	0.003	0.002	1.75	0.13	2.46E+01	6.70E-01	104
PSG2	14	17	0.006	0.007	0.70	0.13	3.63E+00	6.71E-01	105
PSG5	10	13	0.005	0.005	0.68	0.10	4.56E+00	6.91E-01	106
PSG6	17	21	0.008	0.008	0.75	0.16	3.33E+00	7.06E-01	107
PSG4	9	12	0.004	0.005	0.69	0.09	5.05E+00	7.15E-01	108
PSG9	9	12	0.004	0.005	0.69	0.09	5.05E+00	7.15E-01	109
PSG1	21	25	0.010	0.010	0.79	0.20	3.07E+00	7.33E-01	110
CKMT1B	2	1	0.001	0.000	1.96	0.02	4.83E+02	7.75E-01	111
ENSG00000284772	2	1	0.001	0.000	1.96	0.02	4.83E+02	7.75E-01	112
STRC	2	1	0.001	0.000	1.96	0.02	4.83E+02	7.75E-01	113
ANKRD36	4	7	0.002	0.003	0.68	0.03	1.03E+01	7.85E-01	114
POTEB2	2	2	0.001	0.001	1.75	0.02	1.47E+02	7.93E-01	115
ANKRD30B	0	1	0.000	0.000	0.40	0.00	3.46E+02	8.04E-01	116
C1QTNF9	0	1	0.000	0.000	0.40	0.00	3.46E+02	8.04E-01	117
C1QTNF9B	0	1	0.000	0.000	0.40	0.00	3.46E+02	8.04E-01	118
CCDC144A	0	1	0.000	0.000	0.40	0.00	3.46E+02	8.04E-01	119
ENSG00000266302	0	1	0.000	0.000	0.40	0.00	3.46E+02	8.04E-01	120

ENSG00000284393	0	1	0.000	0.000	0.40	0.00	3.46E+02	8.04E-01	121
MIPEP	0	1	0.000	0.000	0.40	0.00	3.46E+02	8.04E-01	122
MZT2A	0	1	0.000	0.000	0.40	0.00	3.46E+02	8.04E-01	123
NBPF1	0	1	0.000	0.000	0.40	0.00	3.46E+02	8.04E-01	124
NECAP1	0	1	0.000	0.000	0.40	0.00	3.46E+02	8.04E-01	125
TUBA3D	0	1	0.000	0.000	0.40	0.00	3.46E+02	8.04E-01	126
ZNF705A	0	1	0.000	0.000	0.40	0.00	3.46E+02	8.04E-01	127
ZNF716	0	1	0.000	0.000	0.40	0.00	3.46E+02	8.04E-01	128
RHD	2	2	0.001	0.001	0.60	0.01	4.45E+01	8.06E-01	129
SPATA31A1	2	2	0.001	0.001	0.60	0.01	4.45E+01	8.06E-01	130
TMEM50A	2	2	0.001	0.001	0.60	0.01	4.45E+01	8.06E-01	131
GYPE	6	7	0.003	0.003	1.35	0.10	1.67E+01	8.16E-01	132
ALDH3B2	1	1	0.000	0.000	1.75	0.00	6.76E+02	8.40E-01	133
CLEC18C	1	1	0.000	0.000	1.75	0.00	6.76E+02	8.40E-01	134
OR2T2	1	1	0.000	0.000	1.75	0.00	6.76E+02	8.40E-01	135
OR2T29	1	1	0.000	0.000	1.75	0.00	6.76E+02	8.40E-01	136
OR2T3	1	1	0.000	0.000	1.75	0.00	6.76E+02	8.40E-01	137
OR2T34	1	1	0.000	0.000	1.75	0.00	6.76E+02	8.40E-01	138
THOC3	1	1	0.000	0.000	0.60	0.00	1.96E+02	8.49E-01	139
FAM86B1	1	2	0.000	0.001	1.57	0.01	1.64E+02	8.51E-01	140
USP17L2	1	2	0.000	0.001	1.57	0.01	1.64E+02	8.51E-01	141
USP17L7	1	2	0.000	0.001	1.57	0.01	1.64E+02	8.51E-01	142
FAM86B2	3	4	0.001	0.002	0.81	0.02	2.30E+01	8.98E-01	143
AMY1A	7	9	0.003	0.004	1.13	0.11	1.10E+01	9.16E-01	144
AMY1B	7	9	0.003	0.004	1.13	0.11	1.10E+01	9.16E-01	145
CNTNAP3	3	3	0.001	0.001	0.84	0.02	3.20E+01	9.24E-01	146
AMY2A	14	19	0.006	0.007	1.08	0.21	5.36E+00	9.27E-01	147
AMY2B	14	19	0.006	0.007	1.08	0.21	5.36E+00	9.27E-01	148
DEFB131A	19	25	0.009	0.010	0.96	0.23	3.87E+00	9.53E-01	149
CYP2A6	2	3	0.001	0.001	0.96	0.01	4.65E+01	9.82E-01	150
SPIDR	2	2	0.001	0.001	1.01	0.01	8.26E+01	9.98E-01	151

Supplementary Table 2. ClozUK dataset. Results of the gene-wise analysis of duplications.

Line in bold marks p value < 0.05. NAFF = number of affected individuals; NUNAFF = number of unaffected individuals (controls); FAFF frequency in affected individuals; FUNAFF = frequency in unaffected individuals (controls); OR = Odds ratio; CI1 – 95% lower limit confidence interval, CI2 – 95% upper limit confidence interval; PVAL = p value; No = number.

GENE	NAFF	NUNAFF	FAFF	FUNAFF	OR	CIL	CIU	PVAL	NO
DRD5	5	0	0.0008	0.0000	1324.03	7.39E+00	1.05E+08	3.54E-03	1
DEFB131A	4	0	0.0007	0.0000	989.25	4.60E+00	8.09E+07	8.45E-03	2
OR4A5	9	6	0.0015	0.0004	13.67	1.29E+00	1.68E+02	3.01E-02	3
OR4A8	9	6	0.0015	0.0004	13.67	1.29E+00	1.68E+02	3.01E-02	4
OR4C46	9	6	0.0015	0.0004	13.67	1.29E+00	1.68E+02	3.01E-02	5
H2AC19	0	6	0.0000	0.0004	0.01	1.44E-07	1.62E+00	8.35E-02	6
H3C15	0	6	0.0000	0.0004	0.01	1.44E-07	1.62E+00	8.35E-02	7
H4C15	0	6	0.0000	0.0004	0.01	1.44E-07	1.62E+00	8.35E-02	8
CYP3A4	1	9	0.0002	0.0007	0.07	3.91E-04	2.21E+00	1.42E-01	9
MICA	0	3	0.0000	0.0002	0.04	5.44E-07	1.13E+01	3.08E-01	10
FOLH1	1	0	0.0002	0.0000	28.46	3.33E-02	2.74E+06	3.32E-01	11
HLA-B	1	0	0.0002	0.0000	28.46	3.33E-02	2.74E+06	3.32E-01	12
RHCE	1	0	0.0002	0.0000	28.46	3.33E-02	2.74E+06	3.32E-01	13
RHD	1	0	0.0002	0.0000	28.46	3.33E-02	2.74E+06	3.32E-01	14
TMEM50A	1	0	0.0002	0.0000	28.46	3.33E-02	2.74E+06	3.32E-01	15
ZNF761	1	5	0.0002	0.0004	0.15	7.91E-04	6.65E+00	3.50E-01	16
ZNF765	1	5	0.0002	0.0004	0.15	7.91E-04	6.65E+00	3.50E-01	17
EFL1	1	4	0.0002	0.0003	0.26	1.27E-03	1.39E+01	5.26E-01	18
TEX101	1	3	0.0002	0.0002	4.42	2.05E-02	2.83E+02	5.29E-01	19
ACSM5	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	20
CA5A	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	21
CCL3L3	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	22
CYP11B1	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	23
CYP4Z1	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	24
DPP6	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	25
ENSG00000025564	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	26
GPR89B	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	27
GRM5	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	28
KLRC1	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	29
KLRC2	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	30

KLRC3	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	31
PARP4	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	32
SLC7A5	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	33
TBC1D3B	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	34
TBC1D3C	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	35
TBC1D3D	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	36
TBC1D3E	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	37
TBC1D3H	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	38
TBC1D3I	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	39
ZNF813	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	40
PSG5	15	25	0.0025	0.0018	1.43	3.07E-01	6.26E+00	6.40E-01	41
UGT2B10	1	4	0.0002	0.0003	0.41	1.89E-03	2.33E+01	6.78E-01	42
CYP3A7	1	5	0.0002	0.0004	0.51	2.44E-03	2.47E+01	7.53E-01	43
CYP3A7-CYP3A51P	1	5	0.0002	0.0004	0.51	2.44E-03	2.47E+01	7.53E-01	44
SAXO2	1	3	0.0002	0.0002	0.53	2.30E-03	3.85E+01	7.76E-01	45
AIMP2	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	46
ANAPC1	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	47
ANKRD36	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	48
ANKRD36B	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	49
ASNS	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	50
C1QTNF9	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	51
CCZ1	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	52
ENSG00000273167	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	53
LYPD3	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	54
MFSD14B	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	55
NUTM2F	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	56
OCM	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	57
ODAD2	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	58
PMS2	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	59
RSPH10B	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	60
GYPB	2	5	0.0003	0.0004	1.39	2.39E-02	4.32E+01	8.58E-01	61
OR2G6	1	4	0.0002	0.0003	0.68	3.07E-03	4.05E+01	8.65E-01	62
OR2T2	1	4	0.0002	0.0003	0.68	3.07E-03	4.05E+01	8.65E-01	63

OR2T3	1	4	0.0002	0.0003	0.68	3.07E-03	4.05E+01	8.65E-01	64
OR2T5	1	4	0.0002	0.0003	0.68	3.07E-03	4.05E+01	8.65E-01	65
OR2T7	1	4	0.0002	0.0003	0.68	3.07E-03	4.05E+01	8.65E-01	66
ANTXRL	25	52	0.0041	0.0038	1.08	3.38E-01	3.25E+00	8.98E-01	67
ZDHHC11	2	4	0.0003	0.0003	0.83	1.42E-02	2.87E+01	9.17E-01	68
PSG1	24	47	0.0040	0.0034	1.06	3.22E-01	3.29E+00	9.24E-01	69
PSG11	24	47	0.0040	0.0034	1.06	3.22E-01	3.29E+00	9.24E-01	70
PSG2	24	47	0.0040	0.0034	1.06	3.22E-01	3.29E+00	9.24E-01	71
PSG6	24	47	0.0040	0.0034	1.06	3.22E-01	3.29E+00	9.24E-01	72
PSG7	24	47	0.0040	0.0034	1.06	3.22E-01	3.29E+00	9.24E-01	73
OCM2	0	2	0.0000	0.0001	0.76	9.04E-06	2.47E+02	9.39E-01	74

Supplementary Table 3. Combined dataset. Results of the gene-wise analysis of duplications.

Line in bold marks p value < 0.05. NAFF = number of affected individuals; NUNAFF = number of unaffected individuals (controls); FAFF frequency in affected individuals; FUNAFF = frequency in unaffected individuals (controls); OR = Odds ratio; CI1 – 95% lower limit confidence interval, CI2 – 95% upper limit confidence interval; PVAL = p value; No = number.

GENE	NAFF	NUNAFF	FAFF	FUNAFF	OR	CIL	CIU	PVAL	NO
DRD5	7	0	0.0008	0.0000	2235.53	1.55E+01	1.69E+08	6.20E-04	1
DPY19L2	29	13	0.0035	0.0008	10.05	2.25E+00	5.02E+01	2.26E-03	2
SYT15	6	17	0.0007	0.0010	0.10	9.85E-03	7.35E-01	2.32E-02	3
SYT15B	6	17	0.0007	0.0010	0.10	9.85E-03	7.35E-01	2.32E-02	4
RNASE3	12	8	0.0015	0.0005	10.20	1.34E+00	8.61E+01	2.53E-02	5
CKMT1A	5	0	0.0006	0.0000	290.55	1.65E+00	2.29E+07	2.83E-02	6
OR4A5	9	6	0.0011	0.0004	13.67	1.30E+00	1.67E+02	2.97E-02	7
OR4A8	9	6	0.0011	0.0004	13.67	1.30E+00	1.67E+02	2.97E-02	8
OR4C46	9	6	0.0011	0.0004	13.67	1.30E+00	1.67E+02	2.97E-02	9
NIPA1	0	4	0.0000	0.0002	0.00	6.25E-08	9.24E-01	4.60E-02	10
CATSPER2	7	1	0.0008	0.0001	33.28	9.26E-01	6.24E+03	5.57E-02	11
TYW1B	3	11	0.0004	0.0007	0.08	3.26E-03	1.11E+00	6.05E-02	12
DGCR6	17	9	0.0021	0.0005	5.58	8.84E-01	4.04E+01	6.77E-02	13
ENSG00000283809	17	9	0.0021	0.0005	5.58	8.84E-01	4.04E+01	6.77E-02	14
PRODH	17	9	0.0021	0.0005	5.58	8.84E-01	4.04E+01	6.77E-02	15
DEFB130A	10	7	0.0012	0.0004	7.22	7.89E-01	7.29E+01	7.98E-02	16
H2AC19	0	6	0.0000	0.0004	0.01	1.48E-07	1.65E+00	8.53E-02	17
H3C15	0	6	0.0000	0.0004	0.01	1.48E-07	1.65E+00	8.53E-02	18
H4C15	0	6	0.0000	0.0004	0.01	1.48E-07	1.65E+00	8.53E-02	19
TEX101	5	3	0.0006	0.0002	14.29	4.79E-01	5.33E+02	1.24E-01	20
ZDHHC11	23	19	0.0028	0.0012	3.02	7.19E-01	1.30E+01	1.31E-01	21
UGT2B28	2	0	0.0002	0.0000	108.17	2.50E-01	9.81E+06	1.38E-01	22
CYP3A4	1	9	0.0001	0.0005	0.07	3.99E-04	2.24E+00	1.44E-01	23
FAHD2B	0	4	0.0000	0.0002	0.02	1.91E-07	3.28E+00	1.45E-01	24
PSG8	2	6	0.0002	0.0004	0.09	1.81E-03	2.34E+00	1.53E-01	25
DEFB103B	6	13	0.0007	0.0008	0.21	1.99E-02	1.82E+00	1.60E-01	26
DEFB104B	6	13	0.0007	0.0008	0.21	1.99E-02	1.82E+00	1.60E-01	27
DEFB105B	6	13	0.0007	0.0008	0.21	1.99E-02	1.82E+00	1.60E-01	28
DEFB106B	6	13	0.0007	0.0008	0.21	1.99E-02	1.82E+00	1.60E-01	29
DEFB107B	6	13	0.0007	0.0008	0.21	1.99E-02	1.82E+00	1.60E-01	30
PRR23D1	6	13	0.0007	0.0008	0.21	1.99E-02	1.82E+00	1.60E-01	31

SPAG11B	6	13	0.0007	0.0008	0.21	1.99E-02	1.82E+00	1.60E-01	32
FAM90A1	8	18	0.0010	0.0011	0.26	3.40E-02	1.72E+00	1.67E-01	33
PDPR	5	2	0.0006	0.0001	11.00	3.59E-01	6.35E+02	1.73E-01	34
USP17L3	3	9	0.0004	0.0005	0.15	5.82E-03	2.37E+00	1.85E-01	35
USP17L8	3	9	0.0004	0.0005	0.15	5.82E-03	2.37E+00	1.85E-01	36
ZNF705B	3	9	0.0004	0.0005	0.15	5.82E-03	2.37E+00	1.85E-01	37
H3-7	1	5	0.0001	0.0003	0.07	3.50E-04	3.39E+00	1.90E-01	38
DEFB4B	6	12	0.0007	0.0007	0.24	2.16E-02	2.10E+00	1.98E-01	39
ZNF705G	6	12	0.0007	0.0007	0.24	2.16E-02	2.10E+00	1.98E-01	40
CLEC18B	1	0	0.0001	0.0000	80.63	9.44E-02	7.76E+06	1.99E-01	41
CTAGE15	1	0	0.0001	0.0000	80.63	9.44E-02	7.76E+06	1.99E-01	42
NPIPB15	1	0	0.0001	0.0000	80.63	9.44E-02	7.76E+06	1.99E-01	43
DEFB103A	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	44
DEFB104A	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	45
DEFB105A	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	46
DEFB106A	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	47
DEFB107A	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	48
FAM90A14	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	49
FAM90A16	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	50
FAM90A17	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	51
FAM90A18	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	52
FAM90A19	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	53
FAM90A22	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	54
FAM90A23	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	55
FAM90A7	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	56
FAM90A8	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	57
FAM90A9	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	58
PRR23D2	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	59
SPAG11A	8	16	0.0010	0.0010	0.31	3.93E-02	2.15E+00	2.42E-01	60
FAM90A20P	5	10	0.0006	0.0006	0.24	1.75E-02	2.59E+00	2.42E-01	61
USP17L1	5	10	0.0006	0.0006	0.24	1.75E-02	2.59E+00	2.42E-01	62
USP17L4	5	10	0.0006	0.0006	0.24	1.75E-02	2.59E+00	2.42E-01	63
ACSM2A	0	2	0.0000	0.0001	0.03	3.41E-07	1.34E+01	2.84E-01	64
DEFB4A	8	16	0.0010	0.0010	0.35	4.37E-02	2.40E+00	2.90E-01	65
GYPA	2	6	0.0002	0.0004	0.17	3.12E-03	4.42E+00	2.94E-01	66
FAM90A11P	3	8	0.0004	0.0005	0.22	8.26E-03	3.78E+00	3.09E-01	67
FAM90A12P	3	8	0.0004	0.0005	0.22	8.26E-03	3.78E+00	3.09E-01	68
FAM90A24P	3	8	0.0004	0.0005	0.22	8.26E-03	3.78E+00	3.09E-01	69
MICA	0	3	0.0000	0.0002	0.05	5.59E-07	1.15E+01	3.12E-01	70
FOLH1	1	0	0.0001	0.0000	29.59	3.47E-02	2.85E+06	3.26E-01	71
HLA-B	1	0	0.0001	0.0000	29.59	3.47E-02	2.85E+06	3.26E-01	72
RHCE	1	0	0.0001	0.0000	29.59	3.47E-02	2.85E+06	3.26E-01	73
CD177	2	0	0.0002	0.0000	21.26	6.57E-02	1.80E+06	3.30E-01	74
DTX2	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	75

FKBP6	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	76
NAT8B	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	77
NSUN5	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	78
PGM5	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	79
PRAMEF14	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	80
PRAMEF17	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	81
PRAMEF19	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	82
PRAMEF20	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	83
SIMC1	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	84
TRIM50	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	85
UPK3B	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	86
ZNF761	1	5	0.0001	0.0003	0.16	8.17E-04	6.84E+00	3.58E-01	87
ZNF765	1	5	0.0001	0.0003	0.16	8.17E-04	6.84E+00	3.58E-01	88
POMZP3	6	10	0.0007	0.0006	0.34	2.98E-02	3.37E+00	3.61E-01	89
SPDYE16	6	10	0.0007	0.0006	0.34	2.98E-02	3.37E+00	3.61E-01	90
SAXO2	3	3	0.0004	0.0002	4.93	1.30E-01	1.89E+02	3.74E-01	91
HNRNPCL1	2	5	0.0002	0.0003	0.21	3.72E-03	6.39E+00	3.78E-01	92
DEFB131A	23	25	0.0028	0.0015	1.70	4.38E-01	6.56E+00	4.39E-01	93
ANKRD36B	4	4	0.0005	0.0002	3.06	1.15E-01	8.05E+01	4.94E-01	94
TP53TG3	4	3	0.0005	0.0002	2.85	9.63E-02	1.02E+02	5.40E-01	95
TP53TG3F	4	3	0.0005	0.0002	2.85	9.63E-02	1.02E+02	5.40E-01	96
CROCC	18	18	0.0022	0.0011	1.58	3.34E-01	7.54E+00	5.60E-01	97
POTED	0	2	0.0000	0.0001	0.16	1.86E-06	5.09E+01	5.73E-01	98
CST1	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	99
DEFA3	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	100
ENSG00000284337	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	101
FAM90A13P	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	102
FAM90A15P	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	103
FAM90A3P	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	104
FAM90A5P	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	105
GPR42	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	106
MALL	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	107
RGPD5	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	108
TP53TG3B	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	109
TP53TG3C	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	110
TP53TG3E	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	111
TRIM77	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	112
EFL1	3	4	0.0004	0.0002	2.42	7.33E-02	6.69E+01	6.01E-01	113
MGAM	6	4	0.0007	0.0002	2.05	1.22E-01	4.28E+01	6.19E-01	114
ACSM5	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	115
CA5A	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	116
CCL3L3	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	117
CYP11B1	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	118
CYP4Z1	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	119

DPP6	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	120
ENSG00000255641	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	121
GPR89B	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	122
GRM5	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	123
KLRC1	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	124
KLRC2	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	125
KLRC3	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	126
PARP4	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	127
SLC7A5	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	128
TBC1D3B	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	129
TBC1D3C	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	130
TBC1D3D	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	131
TBC1D3E	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	132
TBC1D3H	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	133
TBC1D3I	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	134
ZNF813	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	135
GOLGA8A	8	11	0.0010	0.0007	0.61	6.85E-02	5.02E+00	6.46E-01	136
GYPB	8	11	0.0010	0.0007	1.57	1.70E-01	1.34E+01	6.83E-01	137
UGT2B10	1	4	0.0001	0.0002	0.41	1.93E-03	2.33E+01	6.83E-01	138
PSG4	9	12	0.0011	0.0007	0.68	8.48E-02	5.04E+00	7.02E-01	139
PSG9	9	12	0.0011	0.0007	0.68	8.48E-02	5.04E+00	7.02E-01	140
C1QTNF9	0	2	0.0000	0.0001	0.29	3.35E-06	1.05E+02	7.19E-01	141
PSG11	39	67	0.0047	0.0041	0.85	3.25E-01	2.15E+00	7.28E-01	142
CYP3A7	1	5	0.0001	0.0003	0.50	2.43E-03	2.40E+01	7.48E-01	143
CYP3A7-CYP3A51P	1	5	0.0001	0.0003	0.50	2.43E-03	2.40E+01	7.48E-01	144
ANKRD36	4	8	0.0005	0.0005	0.64	3.30E-02	9.27E+00	7.51E-01	145
SPATA31A1	2	2	0.0002	0.0001	0.52	7.06E-03	3.85E+01	7.52E-01	146
PSG7	41	69	0.0050	0.0042	0.87	3.41E-01	2.17E+00	7.71E-01	147
FAM86B1	1	2	0.0001	0.0001	1.98	8.02E-03	2.06E+02	7.77E-01	148
USP17L2	1	2	0.0001	0.0001	1.98	8.02E-03	2.06E+02	7.77E-01	149
USP17L7	1	2	0.0001	0.0001	1.98	8.02E-03	2.06E+02	7.77E-01	150
POTEB2	2	2	0.0002	0.0001	1.83	2.08E-02	1.61E+02	7.81E-01	151
GYPE	6	7	0.0007	0.0004	1.42	1.05E-01	1.80E+01	7.87E-01	152
THOC3	1	1	0.0001	0.0001	0.52	1.61E-03	1.69E+02	8.07E-01	153
RHD	3	2	0.0004	0.0001	1.55	3.44E-02	9.97E+01	8.20E-01	154
TMEM50A	3	2	0.0004	0.0001	1.55	3.44E-02	9.97E+01	8.20E-01	155
CKMT1B	2	1	0.0002	0.0001	1.69	1.62E-02	4.17E+02	8.24E-01	156
ENSG00000284772	2	1	0.0002	0.0001	1.69	1.62E-02	4.17E+02	8.24E-01	157
STRC	2	1	0.0002	0.0001	1.69	1.62E-02	4.17E+02	8.24E-01	158
AIMP2	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	159
ANAPC1	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	160
ASNS	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	161
CCZ1	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	162

ENSG00000273167	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	163
LYPD3	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	164
MFSD14B	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	165
NUTM2F	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	166
OCM	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	167
ODAD2	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	168
PMS2	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	169
RSPH10B	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	170
ALDH3B2	1	1	0.0001	0.0001	1.83	4.43E-03	7.55E+02	8.31E-01	171
CLEC18C	1	1	0.0001	0.0001	1.83	4.43E-03	7.55E+02	8.31E-01	172
OR2T29	1	1	0.0001	0.0001	1.83	4.43E-03	7.55E+02	8.31E-01	173
OR2T34	1	1	0.0001	0.0001	1.83	4.43E-03	7.55E+02	8.31E-01	174
PSG2	38	64	0.0046	0.0039	0.91	3.44E-01	2.34E+00	8.47E-01	175
ANKRD30B	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	176
C1QTNF9B	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	177
CCDC144A	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	178
ENSG00000266302	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	179
ENSG00000284393	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	180
MIPEP	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	181
MZT2A	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	182
NBPF1	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	183
NECAP1	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	184
TUBA3D	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	185
ZNF705A	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	186
ZNF716	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	187
AMY2A	14	19	0.0017	0.0012	1.16	2.19E-01	5.86E+00	8.58E-01	188
AMY2B	14	19	0.0017	0.0012	1.16	2.19E-01	5.86E+00	8.58E-01	189
PSG1	45	72	0.0054	0.0044	0.92	3.74E-01	2.23E+00	8.58E-01	190
PSG6	41	68	0.0050	0.0041	0.92	3.60E-01	2.30E+00	8.60E-01	191
OR2G6	1	4	0.0001	0.0002	0.68	3.08E-03	3.96E+01	8.62E-01	192
OR2T5	1	4	0.0001	0.0002	0.68	3.08E-03	3.96E+01	8.62E-01	193
OR2T7	1	4	0.0001	0.0002	0.68	3.08E-03	3.96E+01	8.62E-01	194
AMY1A	7	9	0.0008	0.0005	1.20	1.11E-01	1.19E+01	8.78E-01	195
AMY1B	7	9	0.0008	0.0005	1.20	1.11E-01	1.19E+01	8.78E-01	196
CNTNAP3	3	3	0.0004	0.0002	0.77	2.04E-02	2.98E+01	8.82E-01	197
ANTXRL	25	52	0.0030	0.0032	1.08	3.41E-01	3.26E+00	8.89E-01	198
FAM86B2	3	4	0.0004	0.0002	0.80	2.28E-02	2.39E+01	8.98E-01	199
PSG5	25	38	0.0030	0.0023	1.07	3.14E-01	3.49E+00	9.18E-01	200
OCM2	0	2	0.0000	0.0001	0.71	8.43E-06	2.30E+02	9.23E-01	201
OR2T2	2	5	0.0002	0.0003	0.85	1.41E-02	2.74E+01	9.31E-01	202
OR2T3	2	5	0.0002	0.0003	0.85	1.41E-02	2.74E+01	9.31E-01	203
SPIDR	2	2	0.0002	0.0001	0.94	1.15E-02	8.07E+01	9.77E-01	204
CYP2A6	2	3	0.0002	0.0002	1.02	1.42E-02	5.19E+01	9.92E-01	205

Supplementary Table 4. MGS dataset. Results of the gene-wise analysis of deletions.

Line in bold marks p value < 0.05. NAFF = number of affected individuals; NUNAFF = number of unaffected individuals (controls); FAFF frequency in affected individuals; FUNAFF = frequency in unaffected individuals (controls); OR = Odds ratio; CI1 – 95% lower limit confidence interval, CI2 – 95% upper limit confidence interval; PVAL = p value; No = number.

GENE	NAFF	NUNAFF	FAFF	FUNAFF	OR	CIL	CIU	PVAL	NO
OR4N2	29	13	0.013	0.005	26.08	4.32E+00	2.07E+02	2.31E-04	1
OR4M1	30	15	0.014	0.006	16.14	3.07E+00	1.04E+02	7.92E-04	2
TARP	11	1	0.005	0.0004	200.55	6.92E+00	3.44E+04	9.38E-04	3
OR4Q3	30	16	0.014	0.006	12.41	2.48E+00	7.37E+01	1.86E-03	4
OR11H2	20	11	0.009	0.004	10.59	1.56E+00	9.07E+01	1.52E-02	5
TRIM77	0	5	0.000	0.002	0.01	8.51E-08	1.12E+00	5.63E-02	6
CYP2A6	18	12	0.008	0.005	4.30	7.73E-01	2.63E+01	9.63E-02	7
GSTA1	0	3	0.000	0.001	0.01	1.37E-07	2.70E+00	1.18E-01	8
POMZP3	2	10	0.001	0.004	0.09	1.94E-03	1.82E+00	1.23E-01	9
SPDYE16	2	10	0.001	0.004	0.09	1.94E-03	1.82E+00	1.23E-01	10
RNASE3	2	8	0.001	0.003	0.09	1.83E-03	1.87E+00	1.24E-01	11
GSTM1	2	0	0.001	0.000	93.61	2.33E-01	8.36E+06	1.47E-01	12
GSTM2	2	0	0.001	0.000	93.61	2.33E-01	8.36E+06	1.47E-01	13
LINC02203	2	0	0.001	0.000	93.61	2.33E-01	8.36E+06	1.47E-01	14
OR4M2	2	0	0.001	0.000	93.61	2.33E-01	8.36E+06	1.47E-01	15
OR4M2-OT1	2	0	0.001	0.000	93.61	2.33E-01	8.36E+06	1.47E-01	16
OR4M2B	2	0	0.001	0.000	93.61	2.33E-01	8.36E+06	1.47E-01	17
OR4N4	2	0	0.001	0.000	93.61	2.33E-01	8.36E+06	1.47E-01	18
OR4N4C	2	0	0.001	0.000	93.61	2.33E-01	8.36E+06	1.47E-01	19
POTEM	0	2	0.000	0.001	0.01	1.75E-07	4.78E+00	1.65E-01	20
TPPP	3	1	0.001	0.0004	53.52	2.48E-01	4.27E+06	1.65E-01	21
FAM86B2	4	2	0.002	0.001	10.65	3.00E-01	6.32E+02	1.93E-01	22
DHRS4L2	0	4	0.000	0.002	0.02	2.59E-07	5.59E+00	2.00E-01	23
GPR42	9	4	0.004	0.002	4.90	3.94E-01	8.67E+01	2.22E-01	24
PUTATIVE	1	0	0.000	0.000	64.05	7.48E-02	6.17E+06	2.25E-01	25
SVIL	1	0	0.000	0.000	64.05	7.48E-02	6.17E+06	2.25E-01	26
THOC3	1	0	0.000	0.000	64.05	7.48E-02	6.17E+06	2.25E-01	27
ACSM2A	1	3	0.000	0.001	0.09	3.95E-04	5.47E+00	2.50E-01	28
GYPB	1	3	0.000	0.001	0.09	3.95E-04	5.47E+00	2.50E-01	29
POTEG	12	8	0.005	0.003	3.39	3.75E-01	3.74E+01	2.81E-01	30
GSTA2	0	2	0.000	0.001	0.03	3.68E-07	1.32E+01	2.88E-01	31
RPS15A	0	2	0.000	0.001	0.03	3.68E-07	1.32E+01	2.88E-01	32
POTED	2	0	0.001	0.000	24.63	7.61E-02	2.08E+06	3.05E-01	33
DEFB131A	13	11	0.006	0.004	2.55	3.79E-01	1.82E+01	3.35E-01	34

PDPR	5	3	0.002	0.001	5.27	1.89E-01	2.89E+02	3.36E-01	35
CLEC18C	3	2	0.001	0.001	7.95	1.11E-01	1.88E+03	3.49E-01	36
SPATA31A1	2	1	0.001	0.000	8.95	7.41E-02	2.39E+03	3.62E-01	37
TP53TG3	23	23	0.010	0.009	1.92	4.68E-01	7.92E+00	3.64E-01	38
TP53TG3F	23	23	0.010	0.009	1.92	4.68E-01	7.92E+00	3.64E-01	39
ENSG00000273756	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	40
ENSG00000289768	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	41
FCGR2B	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	42
FCGR2C	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	43
FCGR3A	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	44
FCGR3B	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	45
GOLGA8S	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	46
LINC02218	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	47
NOX4	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	48
TAF11L2	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	49
TP53TG3D	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	50
TRIM49B	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	51
TRIM51G	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	52
TRIM64C	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	53
ZNF761	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	54
ZNF813	0	1	0.000	0.000	0.05	4.99E-07	4.10E+01	3.81E-01	55
OR2T29	5	12	0.002	0.005	0.38	2.87E-02	3.88E+00	4.25E-01	56
OR2T34	5	12	0.002	0.005	0.38	2.87E-02	3.88E+00	4.25E-01	57
GYPE	1	2	0.000	0.001	0.19	7.53E-04	1.94E+01	4.71E-01	58
CNTNAP3	2	4	0.001	0.002	0.29	4.87E-03	1.02E+01	4.97E-01	59
DTX2	0	2	0.000	0.001	0.12	1.47E-06	4.04E+01	5.22E-01	60
UPK3B	0	2	0.000	0.001	0.12	1.47E-06	4.04E+01	5.22E-01	61
FAHD2B	14	15	0.006	0.006	1.76	2.87E-01	1.08E+01	5.38E-01	62
C1QTNF3-AMACR	1	0	0.000	0.000	7.58	8.87E-03	7.30E+05	5.67E-01	63
ENSG00000266728	1	0	0.000	0.000	7.58	8.87E-03	7.30E+05	5.67E-01	64
GYPA	1	0	0.000	0.000	7.58	8.87E-03	7.30E+05	5.67E-01	65
LGALS9	1	0	0.000	0.000	7.58	8.87E-03	7.30E+05	5.67E-01	66
NF1	1	0	0.000	0.000	7.58	8.87E-03	7.30E+05	5.67E-01	67
OR13C2	1	0	0.000	0.000	7.58	8.87E-03	7.30E+05	5.67E-01	68
POM121	1	0	0.000	0.000	7.58	8.87E-03	7.30E+05	5.67E-01	69
ZNF419	1	0	0.000	0.000	7.58	8.87E-03	7.30E+05	5.67E-01	70
FAM86B1	2	2	0.001	0.001	3.06	3.72E-02	2.44E+02	6.00E-01	71
USP17L2	2	2	0.001	0.001	3.06	3.72E-02	2.44E+02	6.00E-01	72
USP17L7	2	2	0.001	0.001	3.06	3.72E-02	2.44E+02	6.00E-01	73
ANKRD36B	4	6	0.002	0.002	0.44	1.43E-02	1.01E+01	6.09E-01	74
SYT15	3	6	0.001	0.002	0.44	1.43E-02	1.01E+01	6.09E-01	75
SYT15B	3	6	0.001	0.002	0.44	1.43E-02	1.01E+01	6.09E-01	76
TYW1	2	4	0.001	0.002	0.41	6.79E-03	1.48E+01	6.32E-01	77

NIPA1	2	3	0.001	0.001	0.40	6.21E-03	1.89E+01	6.39E-01	78
DEFA1B	1	2	0.000	0.001	0.34	1.29E-03	4.15E+01	6.58E-01	79
DGCR6	3	4	0.001	0.002	2.10	6.33E-02	5.70E+01	6.59E-01	80
ENSG00000283809	3	4	0.001	0.002	2.10	6.33E-02	5.70E+01	6.59E-01	81
PRODH	3	4	0.001	0.002	2.10	6.33E-02	5.70E+01	6.59E-01	82
CTAGE6	1	3	0.000	0.001	0.39	1.63E-03	2.77E+01	6.74E-01	83
TCAF2	1	3	0.000	0.001	0.39	1.63E-03	2.77E+01	6.74E-01	84
ANKRD36	14	16	0.006	0.006	1.42	2.37E-01	8.33E+00	6.99E-01	85
POTEB2	3	5	0.001	0.002	0.58	1.85E-02	1.34E+01	7.33E-01	86
DPY19L2	11	11	0.005	0.004	1.30	1.83E-01	9.24E+00	7.92E-01	87
CLEC18B	0	1	0.000	0.000	0.40	4.20E-06	3.46E+02	8.04E-01	88
DEFB131B	0	1	0.000	0.000	0.40	4.20E-06	3.46E+02	8.04E-01	89
MZT2B	0	1	0.000	0.000	0.40	4.20E-06	3.46E+02	8.04E-01	90
NPIPB15	0	1	0.000	0.000	0.40	4.20E-06	3.46E+02	8.04E-01	91
NPIPB3	0	1	0.000	0.000	0.40	4.20E-06	3.46E+02	8.04E-01	92
SMPD4	0	1	0.000	0.000	0.40	4.20E-06	3.46E+02	8.04E-01	93
TUBA3E	0	1	0.000	0.000	0.40	4.20E-06	3.46E+02	8.04E-01	94
XNDC1N	0	1	0.000	0.000	0.40	4.20E-06	3.46E+02	8.04E-01	95
NOVEL	16	18	0.007	0.007	1.22	2.34E-01	6.21E+00	8.12E-01	96
RRP7A	4	6	0.002	0.002	0.71	3.43E-02	1.21E+01	8.14E-01	97
CCDC144A	1	1	0.000	0.000	1.75	4.54E-03	6.76E+02	8.40E-01	98
ENSG00000266302	1	1	0.000	0.000	1.75	4.54E-03	6.76E+02	8.40E-01	99
NBPF1	1	1	0.000	0.000	1.75	4.54E-03	6.76E+02	8.40E-01	100
TP53TG3B	1	1	0.000	0.000	0.60	1.86E-03	1.96E+02	8.49E-01	101
TP53TG3C	1	1	0.000	0.000	0.60	1.86E-03	1.96E+02	8.49E-01	102
TP53TG3E	1	1	0.000	0.000	0.60	1.86E-03	1.96E+02	8.49E-01	103
DEFB130A	2	3	0.001	0.001	0.96	1.39E-02	4.65E+01	9.82E-01	104
TYW1B	2	3	0.001	0.001	0.96	1.39E-02	4.65E+01	9.82E-01	105

Supplementary Table 5. ClozUK dataset. Results of the gene-wise analysis of deletions.

Line in bold marks p value < 0.05. NAFF = number of affected individuals; NUNAFF = number of unaffected individuals (controls); FAFF frequency in affected individuals; FUNAFF = frequency in unaffected individuals (controls); OR = Odds ratio; CI1 – 95% lower limit confidence interval, CI2 – 95% upper limit confidence interval; PVAL = p value; No = number.

GENE	NAFF	NUNAFF	FAFF	FUNAFF	OR	CIL	CIU	PVAL	NO
MICA	2	19	0.0003	0.0014	0.05	1.24E-03	7.58E-01	2.86E-02	1
CA5A	1	0	0.0002	0.0000	390.94	4.58E-01	3.76E+07	8.12E-02	2
SLC7A5	1	0	0.0002	0.0000	390.94	4.58E-01	3.76E+07	8.12E-02	3
GSTA1	6	9	0.0010	0.0007	4.28	3.47E-01	4.59E+01	2.46E-01	4
GSTA2	6	9	0.0010	0.0007	4.28	3.47E-01	4.59E+01	2.46E-01	5
ALG1	1	0	0.0002	0.0000	28.46	3.33E-02	2.74E+06	3.32E-01	6
C1QTNF9	1	0	0.0002	0.0000	28.46	3.33E-02	2.74E+06	3.32E-01	7
EEF2KMT	1	0	0.0002	0.0000	28.46	3.33E-02	2.74E+06	3.32E-01	8
ENSG00000273167	1	0	0.0002	0.0000	28.46	3.33E-02	2.74E+06	3.32E-01	9
TARP	0	4	0.0000	0.0003	0.07	8.72E-07	1.53E+01	3.90E-01	10
CCDC146	1	1	0.0002	0.0001	8.40	1.99E-02	3.54E+03	4.59E-01	11
ODAD2	0	2	0.0000	0.0001	0.14	1.51E-06	6.11E+01	5.60E-01	12
ANTXRL	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	13
CYP3A4	0	1	0.0000	0.0001	0.18	1.88E-06	1.54E+02	6.31E-01	14
ZNF761	2	5	0.0003	0.0004	2.12	3.62E-02	6.42E+01	6.88E-01	15
ZNF765	2	5	0.0003	0.0004	2.12	3.62E-02	6.42E+01	6.88E-01	16
ZNF813	2	5	0.0003	0.0004	2.12	3.62E-02	6.42E+01	6.88E-01	17
UGT2B10	6	11	0.0010	0.0008	1.36	1.21E-01	1.28E+01	7.93E-01	18
AAMDC	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	19
ANKRD18A	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	20
AREG	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	21
CEACAM8	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	22
DPY19L2	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	23
ENSG00000286070	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	24
GGT1	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	25
GRM8	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	26
GYPB	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	27
INTS4	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	28
TPPP	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	29
ZDHHC11	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	30
ZDHHC11B	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	31
BANP	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	32
NPIPB3	0	1	0.0000	0.0001	2.48	2.58E-05	2.12E+03	8.16E-01	33
NBPF4	0	2	0.0000	0.0001	0.76	9.04E-06	2.47E+02	9.39E-01	34

Supplementary Table 6. Combined dataset. Results for the gene-wise analysis of deletions.

Line in bold marks p value < 0.05. NAFF = number of affected individuals; NUNAFF = number of unaffected individuals (controls); FAFF frequency in affected individuals; FUNAFF = frequency in unaffected individuals (controls); OR = Odds ratio; CI1 – 95% lower limit confidence interval, CI2 – 95% upper limit confidence interval; PVAL = p value; No = number.

GENE	NAFF	NUNAFF	FAFF	FUNAFF	OR	CIL	CIU	PVAL	NO
OR4N2	26	7	0.0031	0.0004	26.88	4.38E+00	2.17E+02	2.33E-04	1
OR4M1	27	9	0.0033	0.0005	16.50	3.08E+00	1.07E+02	8.14E-04	2
OR4Q3	27	10	0.0033	0.0006	12.57	2.47E+00	7.57E+01	1.96E-03	3
OR11H2	18	7	0.0022	0.0004	10.68	1.54E+00	9.30E+01	1.59E-02	4
TARP	11	5	0.0013	0.0003	17.93	1.64E+00	2.54E+02	1.76E-02	5
MICA	2	19	0.0002	0.0012	0.05	1.25E-03	7.63E-01	2.90E-02	6
TRIM77	0	5	0.0000	0.0003	0.01	8.24E-08	1.14E+00	5.75E-02	7
CA5A	1	0	0.0001	0.0000	364.20	4.27E-01	3.50E+07	8.48E-02	8
SLC7A5	1	0	0.0001	0.0000	364.20	4.27E-01	3.50E+07	8.48E-02	9
CYP2A6	18	11	0.0022	0.0007	4.36	7.71E-01	2.71E+01	9.63E-02	10
RNASE3	2	8	0.0002	0.0005	0.09	1.77E-03	1.89E+00	1.25E-01	11
POMZP3	2	9	0.0002	0.0005	0.10	1.97E-03	1.94E+00	1.33E-01	12
SPDYE16	2	9	0.0002	0.0005	0.10	1.97E-03	1.94E+00	1.33E-01	13
GSTM1	2	0	0.0002	0.0000	108.17	2.50E-01	9.81E+06	1.38E-01	14
GSTM2	2	0	0.0002	0.0000	108.17	2.50E-01	9.81E+06	1.38E-01	15
LINC02203	2	0	0.0002	0.0000	108.17	2.50E-01	9.81E+06	1.38E-01	16
OR4M2	2	0	0.0002	0.0000	108.17	2.50E-01	9.81E+06	1.38E-01	17
OR4M2-OT1	2	0	0.0002	0.0000	108.17	2.50E-01	9.81E+06	1.38E-01	18
OR4M2B	2	0	0.0002	0.0000	108.17	2.50E-01	9.81E+06	1.38E-01	19
OR4N4	2	0	0.0002	0.0000	108.17	2.50E-01	9.81E+06	1.38E-01	20
OR4N4C	2	0	0.0002	0.0000	108.17	2.50E-01	9.81E+06	1.38E-01	21
POTEM	0	2	0.0000	0.0001	0.01	1.51E-07	4.13E+00	1.49E-01	22
FAM86B2	4	2	0.0005	0.0001	12.28	3.35E-01	7.42E+02	1.72E-01	23
GYPB	1	4	0.0001	0.0002	0.07	3.14E-04	3.77E+00	1.93E-01	24
PUTATIVE	1	0	0.0001	0.0000	80.63	9.44E-02	7.76E+06	1.99E-01	25
SVIL	1	0	0.0001	0.0000	80.63	9.44E-02	7.76E+06	1.99E-01	26
THOC3	1	0	0.0001	0.0000	80.63	9.44E-02	7.76E+06	1.99E-01	27
DHRS4L2	0	3	0.0000	0.0002	0.02	2.55E-07	6.01E+00	2.06E-01	28
ACSM2A	1	3	0.0001	0.0002	0.07	3.42E-04	4.72E+00	2.22E-01	29
GPR42	9	4	0.0011	0.0002	4.54	3.59E-01	8.17E+01	2.48E-01	30
TPPP	3	1	0.0004	0.0001	13.39	1.39E-01	4.25E+03	2.77E-01	31
RPS15A	0	2	0.0000	0.0001	0.03	3.41E-07	1.34E+01	2.84E-01	32
POTEG	11	6	0.0013	0.0004	3.15	3.44E-01	3.53E+01	3.14E-01	33
ALG1	1	0	0.0001	0.0000	29.59	3.47E-02	2.85E+06	3.26E-01	34
C1QTNF9	1	0	0.0001	0.0000	29.59	3.47E-02	2.85E+06	3.26E-01	35

EEF2KMT	1	0	0.0001	0.0000	29.59	3.47E-02	2.85E+06	3.26E-01	36
ENSG00000273167	1	0	0.0001	0.0000	29.59	3.47E-02	2.85E+06	3.26E-01	37
POTED	2	0	0.0002	0.0000	21.26	6.57E-02	1.80E+06	3.30E-01	38
DEFB131A	13	10	0.0016	0.0006	2.57	3.75E-01	1.88E+01	3.36E-01	39
SPATA31A1	2	1	0.0002	0.0001	10.36	8.11E-02	2.84E+03	3.36E-01	40
TP53TG3	22	21	0.0027	0.0013	2.00	4.80E-01	8.37E+00	3.40E-01	41
TP53TG3F	22	21	0.0027	0.0013	2.00	4.80E-01	8.37E+00	3.40E-01	42
ENSG00000273756	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	43
ENSG00000289768	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	44
FCGR2B	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	45
FCGR2C	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	46
FCGR3A	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	47
FCGR3B	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	48
GOLGA8S	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	49
LINC02218	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	50
NOX4	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	51
TAF11L2	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	52
TP53TG3D	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	53
TRIM49B	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	54
TRIM51G	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	55
TRIM64C	0	1	0.0000	0.0001	0.04	4.31E-07	3.54E+01	3.58E-01	56
CLEC18C	3	1	0.0004	0.0001	7.80	1.04E-01	1.92E+03	3.60E-01	57
PDPR	5	2	0.0006	0.0001	4.87	1.71E-01	2.74E+02	3.64E-01	58
GYPE	1	2	0.0001	0.0001	0.16	6.51E-04	1.67E+01	4.33E-01	59
OR2T29	5	11	0.0006	0.0007	0.41	2.97E-02	4.22E+00	4.59E-01	60
OR2T34	5	11	0.0006	0.0007	0.41	2.97E-02	4.22E+00	4.59E-01	61
CCDC146	1	1	0.0001	0.0001	8.27	2.01E-02	3.41E+03	4.60E-01	62
CNTNAP3	2	4	0.0002	0.0002	0.27	4.51E-03	1.00E+01	4.81E-01	63
FAHD2B	13	13	0.0016	0.0008	1.83	2.94E-01	1.14E+01	5.12E-01	64
GSTA2	6	11	0.0007	0.0007	2.09	1.80E-01	2.00E+01	5.40E-01	65
FAM86B1	2	2	0.0002	0.0001	3.56	4.15E-02	2.91E+02	5.55E-01	66
USP17L2	2	2	0.0002	0.0001	3.56	4.15E-02	2.91E+02	5.55E-01	67
USP17L7	2	2	0.0002	0.0001	3.56	4.15E-02	2.91E+02	5.55E-01	68
ODAD2	0	2	0.0000	0.0001	0.14	1.54E-06	6.06E+01	5.63E-01	69
DTX2	0	2	0.0000	0.0001	0.16	1.86E-06	5.09E+01	5.73E-01	70
UPK3B	0	2	0.0000	0.0001	0.16	1.86E-06	5.09E+01	5.73E-01	71
DGCR6	3	4	0.0004	0.0002	2.52	7.44E-02	6.91E+01	5.86E-01	72
ENSG00000283809	3	4	0.0004	0.0002	2.52	7.44E-02	6.91E+01	5.86E-01	73
PRODH	3	4	0.0004	0.0002	2.52	7.44E-02	6.91E+01	5.86E-01	74
C1QTNF3-AMACR	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	75
ENSG00000266728	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	76
GYPA	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	77
LGALS9	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	78

NF1	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	79
OR13C2	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	80
POM121	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	81
ZNF419	1	0	0.0001	0.0000	6.55	7.67E-03	6.30E+05	5.96E-01	82
ANKRD36B	3	5	0.0004	0.0003	0.42	1.35E-02	1.01E+01	5.97E-01	83
SYT15	3	5	0.0004	0.0003	0.42	1.35E-02	1.01E+01	5.97E-01	84
SYT15B	3	5	0.0004	0.0003	0.42	1.35E-02	1.01E+01	5.97E-01	85
NIPA1	2	3	0.0002	0.0002	0.37	5.58E-03	1.77E+01	6.06E-01	86
TYW1	2	4	0.0002	0.0002	0.41	6.63E-03	1.56E+01	6.38E-01	87
ANTXRL	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	88
CYP3A4	0	1	0.0000	0.0001	0.19	1.95E-06	1.60E+02	6.39E-01	89
DEFA1B	1	2	0.0001	0.0001	0.32	1.18E-03	4.13E+01	6.43E-01	90
ANKRD36	13	14	0.0016	0.0009	1.46	2.40E-01	8.75E+00	6.77E-01	91
ZNF765	2	5	0.0002	0.0003	2.05	3.55E-02	6.16E+01	6.99E-01	92
CTAGE6	1	3	0.0001	0.0002	0.43	1.74E-03	3.21E+01	7.12E-01	93
TCAF2	1	3	0.0001	0.0002	0.43	1.74E-03	3.21E+01	7.12E-01	94
NPIPB3	0	2	0.0000	0.0001	0.29	3.35E-06	1.05E+02	7.19E-01	95
POTEB2	3	5	0.0004	0.0003	0.59	1.82E-02	1.41E+01	7.44E-01	96
GSTA1	6	12	0.0007	0.0007	1.42	1.24E-01	1.32E+01	7.68E-01	97
UGT2B10	6	11	0.0007	0.0007	1.38	1.23E-01	1.28E+01	7.85E-01	98
NOVEL	15	17	0.0018	0.0010	1.25	2.36E-01	6.46E+00	7.93E-01	99
TP53TG3B	1	1	0.0001	0.0001	0.52	1.61E-03	1.69E+02	8.07E-01	100
TP53TG3C	1	1	0.0001	0.0001	0.52	1.61E-03	1.69E+02	8.07E-01	101
TP53TG3E	1	1	0.0001	0.0001	0.52	1.61E-03	1.69E+02	8.07E-01	102
RRP7A	4	6	0.0005	0.0004	0.73	3.41E-02	1.28E+01	8.28E-01	103
BANP	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	104
AAMDC	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	105
ANKRD18A	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	106
AREG	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	107
CEACAM8	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	108
ENSG00000286070	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	109
GGT1	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	110
GRM8	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	111
INTS4	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	112
ZDHHC11	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	113
ZDHHC11B	0	1	0.0000	0.0001	2.31	2.40E-05	1.97E+03	8.29E-01	114
CCDC144A	1	1	0.0001	0.0001	1.83	4.43E-03	7.55E+02	8.31E-01	115
ENSG00000266302	1	1	0.0001	0.0001	1.83	4.43E-03	7.55E+02	8.31E-01	116
NBPF1	1	1	0.0001	0.0001	1.83	4.43E-03	7.55E+02	8.31E-01	117
CLEC18B	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	118
DEFB131B	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	119
MZT2B	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	120
NPIPB15	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	121
SMPD4	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	122

TUBA3E	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	123
XNDC1N	0	1	0.0000	0.0001	0.51	5.31E-06	4.36E+02	8.55E-01	124
DPY19L2	11	12	0.0013	0.0007	1.18	1.66E-01	8.22E+00	8.69E-01	125
ZNF761	2	6	0.0002	0.0004	0.81	1.38E-02	2.32E+01	9.07E-01	126
ZNF813	2	6	0.0002	0.0004	0.81	1.38E-02	2.32E+01	9.07E-01	127
NBPF4	0	2	0.0000	0.0001	0.71	8.43E-06	2.30E+02	9.23E-01	128
DEFB130A	2	3	0.0002	0.0002	1.02	1.42E-02	5.19E+01	9.92E-01	129
TYW1B	2	3	0.0002	0.0002	1.02	1.42E-02	5.19E+01	9.92E-01	130

Supplementary Table 7. Burden by number of CNVs.

Results are organized by study and type of CNV (CNV = deletions and duplications together, DUP = duplications, DEL = deletions). OR = Odds ratio, CI1 – 95% lower limit confidence interval, CI2 – 95% upper limit confidence interval, SE = standard error, NCONTROLS = number of controls with variant and average number of variants per individual between brackets, NCASES = number of cases with variant and average number of variants per individual between brackets.

	TYPE	OR	CI1	CI2	SE	PVAL	NCONTROLS	NCASES
MGS	CNV	1.11	0.8	1.55	0.19	5.19E-01	395 (1.08)	335 (1.11)
	DUP	1.09	0.76	1.55	0.2	6.46E-01	345 (1.08)	296 (1.10)
	DEL	1.42	0.95	2.12	0.3	8.96E-02	276 (1.05)	257 (1.07)
ClozUK	CNV	0.61	0.3	1.22	0.24	1.62E-01	166 (1)	60 (1)
	DUP	0.72	0.35	1.48	0.29	3.73E-01	140 (1)	60 (1)
	DEL	0.76	0.32	1.78	0.37	5.27E-01	114 (1)	41 (1)
Combined	CNV	1	0.74	1.34	0.15	9.77E-01	561 (1.06)	395 (1.09)
	DUP	1	0.73	1.38	0.17	9.90E-01	485 (1.05)	356 (1.08)
	DEL	1.26	0.88	1.81	0.24	2.12E-01	390 (1.03)	298 (1.06)

Supplementary Table 8. Burden by length (Kb) of CNVs.

Results are organized by study and type of CNV (CNV = deletions and duplications together, DUP = duplications, DEL = deletions). OR = Odds ratio, CI1 – 95% lower limit confidence interval, CI2 – 95% upper limit confidence interval, SE = standard error, NCONTROLS = number of controls with variant and average length in KB of variants per individual between brackets, NCASES = number of cases with variant and average length in KB of variants per individual between brackets.

	TYPE	OR	CI1	CI2	SE	PVAL	NCONTROLS	NCASES
MGS	CNV	1.00	1.00	1.00	0.00	3.93E-01	395 (195)	335 (175)
	DUP	1.00	1.00	1.00	0.00	3.35E-01	345 (224)	296 (200)
	DEL	1.00	1.00	1.00	0.00	2.95E-01	276 (234)	257 (232)
ClozUK	CNV	1.00	1.00	1.01	0.00	6.11E-01	166 (80.3)	60 (109)
	DUP	1.00	1.00	1.00	0.00	5.05E-01	140 (152)	60 (199)
	DEL	1.00	0.98	1.01	0.01	4.93E-01	114 (51.2)	41 (44.3)
Combined	CNV	1.00	1.00	1.00	0.00	4.91E-01	561 (161)	395 (165)
	DUP	1.00	1.00	1.00	0.00	4.99E-01	485 (203)	356 (200)
	DEL	1.00	1.00	1.00	0.00	3.18E-01	390 (181)	298 (206)

Supplementary Table 9. Burden by number of genes affected by the CNV.

Results are organized by study and type of CNV (CNV = deletions and duplications together, DUP = duplications, DEL = deletions). OR = Odds ratio, CI1 – 95% lower limit confidence interval, CI2 – 95% upper limit confidence interval, SE = standard error, NCONTROLS = number of controls with variant and average number of genes affected by CNV per individual between brackets, NCASES = number of cases with variant and average number of genes affected by CNV per individual between brackets.

	TYPE	OR	CI1	CI2	SE	PVAL	NCONTROLS	NCASES
MGS	CNV	1.11	0.97	1.26	0.07	1.30E-01	395 (4.62)	335 (4.51)
	DUP	1.02	0.94	1.12	0.05	6.13E-01	345 (6.49)	296 (6.00)
	DEL	1.14	0.97	1.34	0.09	1.18E-01	276 (5.52)	257 (5.68)
ClozUK	**CNV**	**0.68**	**0.48**	**0.98**	**0.13**	**3.66E-02**	**166 (3.31)**	**60 (3.01)**
	DUP	**0.62**	**0.40**	**0.96**	**0.14**	**3.03E-02**	**140 (4.8)**	**60 (4.98)**
	DEL	0.79	0.47	1.34	0.22	3.91E-01	114 (1.81)	41 (1.34)
Combined	CNV	1.04	0.92	1.17	0.06	5.42E-01	561 (4.23)	395 (4.28)
	DUP	0.99	0.91	1.07	0.04	7.32E-01	485 (6.00)	356 (5.83)
	DEL	1.11	0.95	1.29	0.09	1.82E-01	390 (4.43)	298 (5.08)

Chapter 3

Supplementary Table 10. Primary findings of Z score constraint.

Results are separated by the type of constraint tested. For a full explanation of different categories, see 3.5 Methods. CNV = duplications and deletions together; DUP = duplications; DEL = deletions; KB = kilobases; N CNV UN/AFF = number of variants in controls versus cases. Statistically significant results ($p < 0.05$) in both sets of CNVs (categorized by different filtering thresholds) are highlighted as following: */* significant in both; -/- not significant in both; */- or -/* significant in one but not the other.

		>50 KB >50 PROBES			**>10 KB >10 PROBES >1 PROBE/5KB DENSITY**		
STUDY	TYPE	N CNV UN/AFF	OR (95% CI)	P VAL	N CNV UN/AFF	OR (95% CI)	P VAL
GENOMIC CONSTRAINT							
MGS	CNV */*	2305/2337	1.16 (1.08-1.24)	4.32E-05	6752/6641	1.1 (1.05-1.14)	2.65E-06
	DUP -/-	1315/1296	1.08 (0.98-1.19)	1.30E-01	2978/2950	1.01 (0.95-1.07)	7.26E-01
	DEL */*	990/1041	1.3 (1.16-1.46)	4.63E-06	3774/3691	1.17 (1.11-1.24)	3.48E-09
CLOZUK1	CNV -/-	2660/1858	1.07 (1-1.14)	5.55E-02	9395/5802	0.99 (0.96-1.03)	6.30E-01
	DUP -/-	1919/1332	1.03 (0.94-1.12)	5.20E-01	4823/3081	0.96 (0.91-1.01)	1.16E-01
	DEL */-	741/526	1.14 (1.01-1.27)	2.78E-02	4572/2721	1.01 (0.96-1.06)	7.24E-01
CLOZUK2	CNV -/*	3095/2045	1.04 (0.98-1.11)	1.79E-01	13464/8531	1.04 (1.01-1.07)	8.04E-03
	DUP -/-	2103/1347	1.04 (0.95-1.14)	3.44E-01	6287/3965	1.03 (0.98-1.08)	2.15E-01
	DEL -/*	992/698	1.08 (0.98-1.19)	1.30E-01	7177/4566	1.05 (1.01-1.1)	9.41E-03
META-ANALYSIS	CNV */*	8060/6240	1.08 (1.04-1.13)	2.53E-04	29611/20974	1.03 (1.01-1.06)	7.28E-04
	DUP -/-	5337/3975	1.05(0.99-1.11)	9.70E-02	14088/9996	1 (0.97-1.03)	9.40E-01
	DEL */*	2723/2265	1.14 (1.07-1.23)	2.08E-04	15523/10978	1.06 (1.03-1.09)	3.55E-05
CODING CONSTRAINT							
MGS	CNV */*	1259/1438	1.12 (1.04-1.21)	2.96E-03	2501/2811	1.13 (1.07-1.2)	1.33E-05
	DUP -/-	875/894	1.09 (1-1.2)	6.29E-02	1499/1558	1.05 (0.97-1.13)	2.33E-01
	DEL */*	384/544	1.23 (1.09-1.4)	1.17E-03	1002/1253	1.3 (1.19-1.42)	7.33E-09
CLOZUK1	CNV */-	2017/1479	1.13 (1.05-1.21)	6.55E-04	4839/2995	1.03 (0.99-1.08)	1.92E-01
	DUP */-	1551/1089	1.1 (1.02-1.2)	2.06E-02	2995/1880	1.03 (0.97-1.1)	3.31E-01
	DEL */-	466/390	1.22 (1.07-1.38)	2.86E-03	1844/1115	1.03 (0.96-1.11)	4.15E-01
CLOZUK2	CNV	2079/1407	1.07 (1-1.14)	5.85E-02	6227/4056	1.08 (1.03-1.12)	3.67E-04

	-/*						
	DUP -/-	1556/1000	1.03 (0.95-1.12)	4.61E-01	3530/2296	1.02 (0.97-1.08)	4.05E-01
	DEL */*	523/407	1.16 (1.04-1.3)	8.28E-03	2697/1760	1.14 (1.07-1.21)	2.04E-05
META-ANALYSIS	CNV */*	5355/4324	1.1 (1.05-1.15)	2.29E-05	13567/9862	1.07 (1.04-1.1)	3.17E-06
	DUP */-	3982/2983	1.07 (1.02-1.13)	1.08E-02	8024/5734	1.03 (0.99-1.07)	1.03E-01
	DEL */*	1373/1341	1.2 (1.1-1.31)	2.68E-05	5543/4128	1.12 (1.07-1.17)	2.00E-06
NON-CODING CONSTRAINT							
MGS	CNV */*	2305/2337	1.12 (1.03-1.21)	7.90E-03	6723/6620	1.07 (1.03-1.12)	5.87E-04
	DUP -/-	1315/1296	1.02 (0.91-1.15)	6.90E-01	2973/2941	0.99 (0.93-1.06)	8.76E-01
	DEL */*	990/1041	1.24 (1.1-1.4)	5.69E-04	3750/3679	1.14 (1.08-1.21)	4.29E-06
CLOZUK1	CNV -/-	2660/1858	1.01 (0.94-1.09)	7.65E-01	9368/5788	0.99 (0.96-1.03)	7.81E-01
	DUP -/-	1919/1332	0.94 (0.85-1.04)	2.34E-01	4818/3079	0.97 (0.92-1.03)	3.42E-01
	DEL -/-	741/526	1.12 (0.99-1.26)	7.26E-02	4550/2709	1 (0.95-1.05)	9.47E-01
CLOZUK2	CNV -/-	3095/2045	1.04 (0.97-1.11)	2.94E-01	13446/8522	1.03 (1-1.07)	5.12E-02
	DUP -/-	2103/1347	1.03 (0.92-1.14)	6.39E-01	6285/3964	1.02 (0.97-1.08)	4.49E-01
	DEL -/*	992/698	1.08 (0.98-1.2)	1.30E-01	7161/4558	1.04 (1-1.09)	4.50E-02
META-ANALYSIS	CNV */*	8060/6240	1.05 (1-1.1)	4.53E-02	29537/20930	1.03 (1.01-1.05)	9.90E-03
	DUP -/-	5337/3975	0.99 (0.93-1.05)	7.29E-01	14076/9984	1 (0.96-1.03)	8.34E-01
	DEL */*	2723/2265	1.13 (1.05-1.22)	1.39E-03	15461/10946	1.05 (1.02-1.08)	1.61E-03
CCRE CONSTRAINT							
MGS	CNV */*	2284/2317	1.15 (1.07-1.24)	1.52E-04	6283/6199	1.08 (1.04-1.12)	1.04E-04
	DUP -/-	1313/1293	1.06 (0.96-1.18)	2.31E-01	2834/2806	1.01 (0.95-1.07)	8.26E-01
	DEL */*	971/1024	1.26 (1.13-1.4)	2.88E-05	3449/3393	1.13 (1.08-1.19)	1.45E-06
CLOZUK1	CNV -/-	2660/1858	1.06 (0.99-1.14)	8.33E-02	9155/5635	1.02 (0.98-1.05)	3.32E-01
	DUP -/-	1919/1332	1.01 (0.92-1.12)	7.65E-01	4773/3039	1 (0.95-1.05)	9.90E-01
	DEL */-	741/526	1.13 (1.01-1.26)	2.66E-02	4382/2596	1.02 (0.98-1.07)	3.26E-01
CLOZUK2	CNV -/*	3092/2044	1.04 (0.97-1.11)	2.58E-01	12980/8251	1.03 (1-1.06)	4.44E-02
	DUP -/-	2100/1347	1 (0.91-1.1)	9.81E-01	6161/3876	1.01 (0.96-1.06)	7.97E-01
	DEL */*	992/697	1.11 (1.01-1.22)	3.40E-02	6819/4375	1.05 (1.01-1.09)	9.65E-03
META-ANALYSIS	CNV	8036/6219	1.07 (1.03-1.12)	9.24E-04	28418/20085	1.04 (1.02-1.06)	3.60E-04

	/						
	DUP -/-	5332/3972	1.02 (0.97-1.08)	4.35E-01	13768/9721	1 (0.97-1.04)	7.91E-01
	DEL */*	2704/2247	1.15 (1.08-1.23)	5.46E-05	14650/10364	1.06 (1.03-1.09)	2.88E-05
NON-CODING EXON CONSTRAINT							
MGS	CNV -/-	1627/1618	1.01 (0.93-1.1)	7.74E-01	2953/2910	1.03 (0.98-1.09)	2.42E-01
	DUP -/-	1009/956	1 (0.91-1.11)	9.35E-01	1564/1502	1.03 (0.96-1.11)	4.00E-01
	DEL -/-	618/662	1.05 (0.9-1.21)	5.45E-01	1389/1408	1.05 (0.96-1.14)	3.07E-01
CLOZUK1	CNV -/-	2298/1614	1.05 (0.97-1.14)	1.92E-01	4365/2809	1 (0.95-1.05)	8.72E-01
	DUP -/-	1694/1166	0.99 (0.89-1.09)	7.94E-01	2371/1597	0.95 (0.88-1.03)	2.09E-01
	DEL */-	604/448	1.21 (1.05-1.39)	6.65E-03	1994/1212	1.02 (0.95-1.1)	5.36E-01
CLOZUK2	CNV -/*	2542/1694	1.03 (0.95-1.1)	4.96E-01	6023/3834	1.08 (1.03-1.12)	1.64E-03
	DUP -/-	1784/1151	0.96 (0.88-1.06)	4.03E-01	3131/1936	1.02 (0.95-1.09)	5.96E-01
	DEL */*	758/543	1.15 (1.02-1.3)	2.02E-02	2892/1898	1.13 (1.07-1.2)	5.71E-05
META-ANALYSIS	CNV */*	6467/4926	1.03 (0.98-1.08)	2.10E-01	13341/9553	1.04 (1.01-1.07)	2.03E-02
	DUP -/-	4487/3273	0.98 (0.93-1.04)	5.18E-01	7066/5035	1 (0.96-1.04)	9.96E-01
	DEL */*	1980/1653	1.13 (1.04-1.24)	5.56E-03	6275/4518	1.07 (1.03-1.12)	1.77E-03
INTRONIC CONSTRAINT							
MGS	CNV -/*	1627/1618	1.01 (0.93-1.1)	7.74E-01	4759/4781	1.08 (1.03-1.13)	6.65E-04
	DUP -/-	1009/956	1 (0.91-1.11)	9.35E-01	2247/2222	1.03 (0.96-1.11)	3.53E-01
	DEL -/*	618/662	1.05 (0.9-1.21)	5.45E-01	2512/2559	1.13 (1.06-1.21)	1.40E-04
CLOZUK1	CNV -/-	2298/1614	1.05 (0.97-1.14)	1.92E-01	7155/4395	1.01 (0.97-1.06)	5.95E-01
	DUP -/-	1694/1166	0.99 (0.89-1.09)	7.94E-01	3717/2390	1.02 (0.95-1.09)	5.81E-01
	DEL */-	604/448	1.21 (1.05-1.39)	6.65E-03	3438/2005	0.99 (0.94-1.05)	8.03E-01
CLOZUK2	CNV -/-	2542/1694	1.03 (0.95-1.1)	4.96E-01	9807/6221	1.03 (0.99-1.07)	1.18E-01
	DUP -/*	1784/1151	0.96 (0.88-1.06)	4.03E-01	4739/3040	1.06 (1-1.13)	4.58E-02
	DEL */-	758/543	1.15 (1.02-1.3)	2.02E-02	5068/3181	1 (0.95-1.06)	8.60E-01
META-ANALYSIS	CNV -/*	6467/4926	1.03 (0.98-1.08)	2.10E-01	21721/15397	1.04 (1.01-1.06)	4.22E-03
	DUP -/-	4487/3273	0.98 (0.93-1.04)	5.18E-01	10703/7652	1.04 (1-1.08)	5.26E-02
	DEL */-	1980/1653	1.13 (1.04-1.24)	5.56E-03	11018/7745	1.03 (0.99-1.06)	1.17E-01
NON-FUNCTIONAL REGION CONSTRAINT							

MGS	CNV -/-	1812/1776	0.99 (0.9-1.11)	9.23E-01	4253/4170	0.99 (0.93-1.05)	7.16E-01
	DUP -/-	1143/1090	0.96 (0.85-1.1)	5.90E-01	2153/2088	0.98 (0.91-1.06)	6.59E-01
	DEL -/-	669/686	1.04 (0.86-1.25)	6.94E-01	2100/2082	1 (0.92-1.09)	9.98E-01
CLOZUK1	CNV -/-	2215/1594	0.98 (0.89-1.07)	6.54E-01	6181/3892	1.04 (0.99-1.1)	1.08E-01
	DUP -/-	1668/1178	0.93 (0.84-1.04)	1.91E-01	3613/2355	1.03 (0.96-1.11)	4.32E-01
	DEL -/-	547/416	1.15 (0.95-1.39)	1.55E-01	2568/1537	1.05 (0.97-1.13)	2.52E-01
CLOZUK2	CNV -/-	2495/1664	1.02 (0.93-1.12)	6.16E-01	8532/5328	1.03 (0.98-1.08)	2.28E-01
	DUP -/-	1844/1171	0.98 (0.88-1.09)	6.95E-01	4522/2728	1 (0.94-1.07)	9.32E-01
	DEL -/*	651/493	1.19 (1-1.43)	5.56E-02	4010/2600	1.07 (1-1.15)	4.20E-02
META-ANALYSIS	CNV -/-	6522/5034	1 (0.94-1.06)	9.64E-01	18966/13390	1.02 (0.99-1.05)	1.61E-01
	DUP -/-	4655/3439	0.96 (0.9-1.02)	1.61E-01	10288/7171	1.01 (0.96-1.05)	8.10E-01
	DEL -/-	1867/1595	1.12 (0.99-1.26)	7.12E-02	8678/6219	1.04 (1-1.09)	7.20E-02

Supplementary Table 11. Sensitivity analysis results removing previously implicated loci.

Results are separated by the type of constraint tested. For a full explanation of different categories, see 3.5 Methods. CNV = duplications and deletions together; DUP = duplications; DEL = deletions; KB = kilobases; N CNV UN/AFF = number of variants in controls versus cases. Statistically significant results ($p < 0.05$) in both sets of CNVs (categorized by different filtering thresholds) are highlighted as following: */* significant in both; -/- not significant in both; */- or -/* significant in one but not the other.

		>50 KB >50 PROBES			**>10 KB >10 PROBES >1 PROBE/5KB DENSITY**		
STUDY	TYPE	N CNV UN/AFF	OR (95% CI)	P VAL	N CNV UN/AFF	OR (95% CI)	P VAL
GENOMIC CONSTRAINT							
MGS	CNV */*	2296/2295	1.15 (1.07-1.23)	1.69E-04	6727/6591	1.09 (1.05-1.14)	3.62E-06
	DUP -/-	1312/1282	1.07 (0.97-1.17)	1.99E-01	2971/2935	1.01 (0.95-1.07)	7.78E-01
	DEL */*	984/1013	1.29 (1.15-1.45)	9.49E-06	3756/3656	1.17 (1.11-1.24)	2.92E-09
CLOZUK1	CNV -/-	2619/1792	1.05 (0.98-1.12)	1.71E-01	9303/5730	0.99 (0.96-1.03)	6.31E-01
	DUP -/-	1885/1298	1.01 (0.92-1.1)	8.34E-01	4784/3066	0.96 (0.91-1.01)	9.49E-02
	DEL *	734/494	1.12 (1-1.25)	6.00E-02	4519/2664	1.01 (0.96-1.06)	7.43E-01
CLOZUK2	CNV -/*	3059/2000	1.03 (0.97-1.1)	2.96E-01	13353/8459	1.04 (1.01-1.07)	5.80E-03
	DUP -/-	2076/1319	1.03 (0.94-1.12)	5.63E-01	6249/3939	1.03 (0.98-1.08)	2.31E-01
	DEL -/*	983/681	1.08 (0.98-1.19)	1.21E-01	7104/4520	1.06 (1.02-1.1)	5.11E-03
META-ANALYSIS	CNV **	7974/6087	1.07 (1.02-1.11)	2.10E-03	29383/20780	1.04 (1.02-1.06)	6.16E-04
	DUP -/-	5273/3899	1.03 (0.98-1.09)	2.73E-01	14004/9940	1 (0.97-1.03)	8.43E-01
	DEL **	2701/2188	1.14 (1.06-1.22)	3.77E-04	15379/10840	1.06 (1.03-1.09)	2.05E-05
CODING CONSTRAINT							
MGS	CNV */*	1252/1397	1.1 (1.02-1.19)	9.66E-03	2485/2768	1.13 (1.07-1.19)	2.73E-05
	DUP -/-	872/880	1.08 (0.98-1.19)	1.06E-01	1492/1544	1.04 (0.97-1.12)	2.70E-01
	DEL */*	380/517	1.22 (1.07-1.38)	2.63E-03	993/1224	1.3 (1.19-1.42)	1.10E-08
CLOZUK1	CNV */-	1977/1413	1.11 (1.04-1.19)	3.17E-03	4773/2928	1.03 (0.99-1.08)	1.71E-01
	DUP */-	1517/1055	1.09 (1-1.19)	4.86E-02	2963/1865	1.03 (0.97-1.1)	3.43E-01

	DEL */-	460/358	1.2 (1.05-1.37)	6.51E-03	1810/1063	1.03 (0.96-1.1)	4.56E-01
CLOZUK2	CNV -/*	2044/1363	1.05 (0.98-1.12)	1.56E-01	6136/3998	1.08 (1.03-1.12)	4.74E-04
	DUP -/-	1530/972	1.01 (0.93-1.1)	7.88E-01	3496/2272	1.02 (0.97-1.08)	4.21E-01
	DEL */*	514/391	1.16 (1.04-1.3)	9.67E-03	2640/1726	1.14 (1.07-1.21)	2.09E-05
META-ANALYSIS	CNV */*	5273/4173	1.08 (1.04-1.13)	3.27E-04	13394/9694	1.07 (1.04-1.1)	4.02E-06
	DUP -/*	3919/2907	1.06 (1-1.12)	4.61E-02	7951/5681	1.03 (0.99-1.07)	1.18E-01
	DEL */*	1354/1266	1.19 (1.09-1.29)	6.60E-05	5443/4013	1.12 (1.07-1.17)	2.49E-06
NON-CODING CONSTRAINT							
MGS	CNV */*	2296/2295	1.11 (1.02-1.2)	1.21E-02	6698/6570	1.08 (1.03-1.12)	5.16E-04
	DUP -/-	1312/1282	1.02 (0.91-1.14)	7.23E-01	2966/2926	1 (0.93-1.06)	8.81E-01
	DEL */*	984/1013	1.23 (1.09-1.39)	8.01E-04	3732/3644	1.14 (1.08-1.21)	2.78E-06
CLOZUK1	CNV -/-	2619/1792	1 (0.93-1.08)	9.49E-01	9276/5716	1 (0.96-1.03)	8.11E-01
	DUP -/-	1885/1298	0.94 (0.85-1.03)	1.92E-01	4779/3064	0.97 (0.92-1.03)	2.98E-01
	DEL -/-	734/494	1.1 (0.98-1.24)	1.15E-01	4497/2652	1 (0.95-1.05)	9.01E-01
CLOZUK2	CNV -/*	3059/2000	1.04 (0.97-1.12)	2.83E-01	13336/8450	1.04 (1-1.07)	3.20E-02
	DUP -/-	2076/1319	1.03 (0.92-1.14)	6.38E-01	6247/3938	1.02 (0.97-1.08)	4.68E-01
	DEL -/*	983/681	1.09 (0.98-1.21)	1.04E-01	7089/4512	1.05 (1.01-1.1)	2.08E-02
META-ANALYSIS	CNV */*	7974/6087	1.04 (1-1.09)	6.91E-02	29310/20736	1.03 (1.01-1.05)	6.16E-03
	DUP -/-	5273/3899	0.99 (0.93-1.05)	6.49E-01	13992/9928	1 (0.96-1.03)	7.77E-01
	DEL */*	2701/2188	1.13 (1.05-1.22)	1.74E-03	15318/10808	1.05 (1.02-1.08)	6.53E-04
CCRE CONSTRAINT							
MGS	CNV */*	2275/2275	1.14 (1.06-1.23)	3.02E-04	6258/6149	1.08 (1.04-1.12)	9.68E-05
	DUP -/-	1310/1279	1.06 (0.96-1.18)	2.55E-01	2827/2791	1.01 (0.95-1.07)	8.25E-01
	DEL */*	965/996	1.25 (1.12-1.39)	4.74E-05	3431/3358	1.14 (1.08-1.2)	1.01E-06
CLOZUK1	CNV -/-	2619/1792	1.05 (0.98-1.13)	1.35E-01	9063/5563	1.02 (0.98-1.05)	3.18E-01
	DUP -/-	1885/1298	1.01 (0.92-1.11)	8.46E-01	4734/3024	1 (0.95-1.05)	9.45E-01
	DEL */-	734/494	1.12 (1-1.24)	4.59E-02	4329/2539	1.02 (0.98-1.07)	3.38E-01
CLOZUK2	CNV -/*	3056/1999	1.04 (0.97-1.11)	2.44E-01	12870/8179	1.04 (1-1.07)	2.37E-02
	DUP -/-	2073/1319	1 (0.91-1.1)	9.78E-01	6123/3850	1.01 (0.96-1.06)	8.14E-01

	DEL */*	983/680	1.12 (1.01-1.23)	2.46E-02	6747/4329	1.06 (1.02-1.1)	3.15E-03
META-ANALYSIS	CNV */*	7950/6066	1.07 (1.03-1.12)	1.64E-03	28191/19891	1.04 (1.02-1.06)	1.79E-04
	DUP -/-	5268/3896	1.02 (0.96-1.08)	4.86E-01	13684/9665	1 (0.97-1.04)	8.27E-01
	DEL */*	2682/2170	1.15 (1.07-1.23)	7.15E-05	14507/10226	1.06 (1.04-1.09)	1.01E-05
NON-CODING EXON CONSTRAINT							
MGS	CNV -/-	1620/1579	1.01 (0.93-1.1)	8.12E-01	2946/2871	1.03 (0.98-1.09)	2.60E-01
	DUP -/-	1006/943	1 (0.9-1.11)	9.86E-01	1561/1490	1.03 (0.96-1.11)	4.28E-01
	DEL -/-	614/636	1.05 (0.91-1.22)	5.02E-01	1385/1381	1.05 (0.96-1.14)	2.99E-01
CLOZUK1	CNV -/-	2261/1551	1.05 (0.97-1.14)	2.22E-01	4338/2775	1 (0.95-1.05)	9.83E-01
	DUP -/-	1663/1134	0.98 (0.89-1.09)	7.34E-01	2349/1590	0.95 (0.88-1.03)	1.92E-01
	DEL */-	598/417	1.21 (1.05-1.39)	8.05E-03	1989/1185	1.03 (0.95-1.11)	4.67E-01
CLOZUK2	CNV */-	2508/1654	1.02 (0.95-1.1)	5.60E-01	5983/3803	1.07 (1.03-1.12)	1.70E-03
	DUP -/-	1759/1127	0.96 (0.87-1.05)	3.49E-01	3102/1925	1.02 (0.95-1.09)	6.30E-01
	DEL */*	749/527	1.15 (1.02-1.3)	1.83E-02	2881/1878	1.13 (1.07-1.2)	6.03E-05
META-ANALYSIS	CNV -/*	6389/4784	1.03 (0.98-1.08)	2.61E-01	13267/9449	1.04 (1.01-1.07)	1.74E-02
	DUP -/-	4428/3204	0.98 (0.92-1.03)	4.28E-01	7012/5005	1 (0.96-1.04)	9.30E-01
	DEL */*	1961/1580	1.13 (1.04-1.24)	5.14E-03	6255/4444	1.08 (1.03-1.12)	1.42E-03
INTRONIC CONSTRAINT							
MGS	CNV -/*	1919/1957	1.08 (0.99-1.18)	8.91E-02	4734/4732	1.08 (1.03-1.13)	7.60E-04
	DUP -/-	1154/1118	0.98 (0.87-1.12)	7.99E-01	2240/2208	1.03 (0.96-1.11)	3.70E-01
	DEL */*	765/839	1.22 (1.07-1.4)	2.76E-03	2494/2524	1.13 (1.06-1.21)	1.13E-04
CLOZUK1	CNV -/-	2531/1740	1.02 (0.94-1.11)	6.40E-01	7063/4323	1.01 (0.97-1.06)	5.87E-01
	DUP -/-	1845/1267	0.96 (0.86-1.07)	4.81E-01	3678/2375	1.02 (0.95-1.09)	6.25E-01
	DEL -/-	686/473	1.11 (0.97-1.26)	1.25E-01	3385/1948	0.99 (0.94-1.05)	8.14E-01
CLOZUK2	CNV -/-	2840/1860	1.08 (1-1.16)	6.10E-02	9702/6151	1.03 (0.99-1.07)	1.06E-01
	DUP -/-	1932/1235	1.09 (0.97-1.21)	1.42E-01	4703/3015	1.06 (1-1.13)	5.00E-02
	DEL -/-	908/625	1.09 (0.98-1.21)	1.29E-01	4999/3136	1.01 (0.96-1.06)	7.54E-01
META-ANALYSIS	CNV */*	7290/5557	1.06 (1-1.11)	3.13E-02	21499/15206	1.04 (1.01-1.07)	3.88E-03
	DUP -/-	4931/3620	1.01 (0.94-1.08)	8.70E-01	10621/7598	1.04 (1-1.08)	6.21E-02

	DEL */-	2359/1937	1.13 (1.04-1.22)	3.75E-03	10878/7608	1.03 (1-1.07)	9.13E-02

Supplementary Table 12. Sensitivity analysis results adding the number of LoF intolerant genes affected.

Results are separated by the type of constraint tested. For a full explanation of different categories, see 3.5 Methods. CNV = duplications and deletions together; DUP = duplications; DEL = deletions; KB = kilobases; N CNV UN/AFF = number of variants in controls versus cases. Statistically significant results ($p < 0.05$) in both sets of CNVs (categorized by different filtering thresholds) are highlighted as following: */* significant in both; -/- not significant in both; */- or -/* significant in one but not the other.

		>50 KB >50 PROBES			**>10 KB >10 PROBES >1 PROBE/5KB DENSITY**		
STUDY	TYPE	N CNV UN/AFF	OR (95% CI)	P VAL	N CNV UN/AFF	OR (95% CI)	P VAL
GENOMIC CONSTRAINT							
MGS	CNV */*	2305/2337	1.11 (1.03-1.19)	5.56E-03	6752/6641	1.07 (1.03-1.12)	3.38E-04
	DUP -/-	1315/1296	1.04 (0.94-1.15)	4.35E-01	2978/2950	0.99 (0.93-1.05)	7.14E-01
	DEL */*	990/1041	1.25 (1.11-1.4)	1.61E-04	3774/3691	1.15 (1.09-1.22)	1.45E-07
CLOZUK1	CNV -/-	2660/1858	1.02 (0.95-1.09)	5.73E-01	9395/5802	0.99 (0.96-1.02)	5.29E-01
	DUP -/-	1919/1332	0.98 (0.89-1.07)	6.23E-01	4823/3081	0.96 (0.91-1.01)	1.30E-01
	DEL -/-	741/526	1.09 (0.97-1.22)	1.54E-01	4572/2721	1.01 (0.96-1.06)	6.66E-01
CLOZUK2	CNV -/*	3095/2045	1.03 (0.97-1.1)	3.19E-01	13464/8531	1.04 (1.01-1.07)	2.20E-02
	DUP -/-	2103/1347	1.05 (0.95-1.15)	3.42E-01	6287/3965	1.03 (0.98-1.08)	2.12E-01
	DEL -/*	992/698	1.06 (0.96-1.17)	2.45E-01	7177/4566	1.04 (1-1.09)	3.69E-02
META-ANALYSIS	CNV */*	8060/6240	1.05 (1.01-1.09)	2.78E-02	29611/20974	1.03 (1.01-1.05)	8.87E-03
	DUP -/-	5337/3975	1.02 (0.96-1.08)	5.49E-01	14088/9996	0.99 (0.96-1.02)	6.92E-01
	DEL */*	2723/2265	1.11 (1.04-1.19)	3.36E-03	15523/10978	1.05 (1.02-1.08)	2.47E-04
CODING CONSTRAINT							
MGS	CNV */*	1259/1438	1.08 (1-1.17)	5.33E-02	2501/2811	1.1 (1.04-1.17)	8.40E-04
	DUP -/-	875/894	1.07 (0.97-1.18)	1.90E-01	1499/1558	1.02 (0.95-1.11)	5.31E-01

	DEL */*	384/544	1.18 (1.04-1.35)	1.22E-02	1002/1253	1.26 (1.15-1.38)	4.36E-07
CLOZUK1	CNV */-	2017/1479	1.09 (1.01-1.17)	2.46E-02	4839/2995	1.03 (0.98-1.07)	2.43E-01
	DUP */-	1551/1089	1.06 (0.97-1.16)	1.70E-01	2995/1880	1.04 (0.97-1.1)	2.82E-01
	DEL */-	466/390	1.17 (1.02-1.33)	2.07E-02	1844/1115	1.04 (0.97-1.11)	3.03E-01
CLOZUK2	CNV -/*	2079/1407	1.06 (0.99-1.13)	9.49E-02	6227/4056	1.07 (1.03-1.12)	1.11E-03
	DUP -/-	1556/1000	1.04 (0.95-1.13)	4.34E-01	3530/2296	1.03 (0.97-1.09)	3.69E-01
	DEL */*	523/407	1.15 (1.03-1.29)	1.64E-02	2697/1760	1.12 (1.06-1.19)	1.54E-04
META-ANALYSIS	CNV */*	5355/4324	1.07 (1.03-1.12)	1.82E-03	13567/9862	1.06 (1.03-1.09)	5.44E-05
	DUP -/-	3982/2983	1.05 (1-1.11)	6.24E-02	8024/5734	1.03 (0.99-1.07)	1.41E-01
	DEL */*	1373/1341	1.16 (1.07-1.26)	3.73E-04	5543/4128	1.11 (1.06-1.16)	7.48E-06
NON-CODING CONSTRAINT							
MGS	CNV */*	2305/2337	1.07 (0.99-1.17)	8.50E-02	6723/6620	1.06 (1.02-1.1)	7.26E-03
	DUP -/-	1315/1296	0.99 (0.88-1.11)	8.84E-01	2973/2941	0.98 (0.92-1.04)	5.17E-01
	DEL */*	990/1041	1.2 (1.07-1.36)	2.91E-03	3750/3679	1.13 (1.07-1.19)	2.19E-05
CLOZUK1	CNV -/-	2660/1858	0.97 (0.9-1.05)	4.80E-01	9368/5788	0.99 (0.96-1.03)	7.37E-01
	DUP */-	1919/1332	0.9 (0.81-1)	4.44E-02	4818/3079	0.97 (0.92-1.03)	3.63E-01
	DEL -/-	741/526	1.07 (0.95-1.21)	2.41E-01	4550/2709	1.01 (0.96-1.06)	7.66E-01
CLOZUK2	CNV -/-	3095/2045	1.03 (0.96-1.1)	4.38E-01	13446/8522	1.03 (1-1.06)	8.72E-02
	DUP -/-	2103/1347	1.02 (0.92-1.14)	6.51E-01	6285/3964	1.02 (0.97-1.08)	4.50E-01
	DEL -/-	992/698	1.07 (0.96-1.19)	2.03E-01	7161/4558	1.04 (1-1.08)	8.10E-02
META-ANALYSIS	CNV -/*	8060/6240	1.02 (0.97-1.07)	4.07E-01	29537/20930	1.02 (1-1.05)	3.72E-02
	DUP -/-	5337/3975	0.96 (0.9-1.02)	1.98E-01	14076/9984	0.99 (0.96-1.03)	6.52E-01
	DEL */*	2723/2265	1.1 (1.03-1.19)	8.80E-03	15461/10946	1.05 (1.02-1.08)	2.51E-03
CCRE CONSTRAINT							
MGS	CNV */*	2284/2317	1.11 (1.03-1.19)	5.72E-03	6283/6199	1.06 (1.02-1.1)	2.19E-03
	DUP -/-	1313/1293	1.03 (0.93-1.15)	5.59E-01	2834/2806	0.99 (0.93-1.05)	7.67E-01
	DEL */*	971/1024	1.22 (1.1-1.36)	2.62E-04	3449/3393	1.12 (1.07-1.18)	1.36E-05
CLOZUK1	CNV -/-	2660/1858	1.03 (0.96-1.11)	3.54E-01	9155/5635	1.02 (0.98-1.05)	3.65E-01
	DUP -/-	1919/1332	0.98 (0.89-1.08)	6.91E-01	4773/3039	1 (0.95-1.06)	9.81E-01

	DEL -/-	741/526	1.1 (0.99-1.23)	7.43E-02	4382/2596	1.03 (0.98-1.08)	2.30E-01
CLOZUK2	CNV -/-	3092/2044	1.03 (0.97-1.1)	3.71E-01	12980/8251	1.03 (1-1.06)	7.43E-02
	DUP -/-	2100/1347	1 (0.91-1.1)	9.99E-01	6161/3876	1.01 (0.96-1.06)	8.02E-01
	DEL -/*	992/697	1.1 (1-1.21)	5.07E-02	6819/4375	1.05 (1.01-1.09)	1.95E-02
META-ANALYSIS	CNV */*	8036/6219	1.05 (1.01-1.1)	1.94E-02	28418/20085	1.03 (1.01-1.05)	2.31E-03
	DUP -/-	5332/3972	1 (0.94-1.06)	9.63E-01	13768/9721	1 (0.97-1.03)	9.88E-01
	DEL */*	2704/2247	1.13 (1.06-1.21)	3.36E-04	14650/10364	1.06 (1.03-1.09)	6.64E-05
NON-CODING EXON CONSTRAINT							
MGS	CNV -/-	1627/1618	0.99 (0.91-1.08)	7.95E-01	2953/2910	1.02 (0.97-1.08)	4.53E-01
	DUP -/-	1009/956	0.98 (0.89-1.09)	7.53E-01	1564/1502	1.02 (0.95-1.1)	6.00E-01
	DEL -/-	618/662	1.03 (0.89-1.2)	6.52E-01	1389/1408	1.04 (0.95-1.14)	3.76E-01
CLOZUK1	CNV -/-	2298/1614	1.03 (0.95-1.12)	4.31E-01	4365/2809	1 (0.95-1.05)	8.68E-01
	DUP -/-	1694/1166	0.97 (0.88-1.08)	5.89E-01	2371/1597	0.95 (0.88-1.03)	2.09E-01
	DEL */-	604/448	1.17 (1.01-1.34)	3.07E-02	1994/1212	1.02 (0.95-1.1)	5.26E-01
CLOZUK2	CNV */-	2542/1694	1.02 (0.95-1.1)	5.78E-01	6023/3834	1.07 (1.03-1.12)	2.06E-03
	DUP -/-	1784/1151	0.96 (0.88-1.06)	4.12E-01	3131/1936	1.02 (0.95-1.09)	5.92E-01
	DEL */*	758/543	1.13 (1-1.27)	4.33E-02	2892/1898	1.13 (1.06-1.2)	1.25E-04
META-ANALYSIS	CNV -/*	6467/4926	1.01 (0.97-1.06)	5.38E-01	13341/9553	1.03 (1-1.06)	3.80E-02
	DUP -/-	4487/3273	0.97 (0.92-1.03)	3.11E-01	7066/5035	1 (0.96-1.04)	8.69E-01
	DEL */*	1980/1653	1.11 (1.02-1.21)	1.87E-02	6275/4518	1.07 (1.02-1.12)	2.86E-03
INTRONIC CONSTRAINT							
MGS	CNV -/*	1928/1999	1.06 (0.97-1.16)	2.32E-01	4759/4781	1.07 (1.02-1.12)	4.63E-03
	DUP -/-	1157/1132	0.96 (0.84-1.1)	5.55E-01	2247/2222	1.02 (0.95-1.09)	5.80E-01
	DEL */*	771/867	1.21 (1.06-1.38)	5.49E-03	2512/2559	1.12 (1.06-1.2)	3.02E-04
CLOZUK1	CNV -/-	2572/1806	0.98 (0.91-1.07)	7.13E-01	7155/4395	1.01 (0.97-1.06)	6.26E-01
	DUP -/-	1879/1301	0.92 (0.82-1.03)	1.66E-01	3717/2390	1.02 (0.95-1.09)	5.48E-01
	DEL -/-	693/505	1.07 (0.94-1.22)	3.13E-01	3438/2005	1 (0.94-1.06)	9.98E-01
CLOZUK2	CNV -/-	2875/1905	1.07 (0.99-1.16)	7.28E-02	9807/6221	1.03 (0.99-1.07)	1.44E-01
	DUP */*	1958/1263	1.1 (0.98-1.23)	9.62E-02	4739/3040	1.07 (1-1.13)	3.97E-02

	DEL -/-	917/642	1.08 (0.96-1.2)	1.88E-01	5068/3181	1 (0.95-1.06)	8.70E-01
META-ANALYSIS	CNV -/*	7375/5710	1.04 (0.99-1.09)	1.71E-01	21721/15397	1.03 (1.01-1.06)	1.14E-02
	DUP -/-	4994/3696	0.98 (0.92-1.05)	6.39E-01	10703/7652	1.04 (1-1.08)	7.22E-02
	DEL */-	2381/2014	1.1 (1.02-1.19)	1.36E-02	11018/7745	1.03 (0.99-1.07)	1.00E-01

Supplementary Table 13. Genic CNVs results.

Results are separated by the type of constraint tested. For a full explanation of different categories, see 3.5 Methods. CNV = duplications and deletions together; DUP = duplications; DEL = deletions; KB = kilobases; N CNV UN/AFF = number of variants in controls versus cases. Statistically significant results ($p < 0.05$) in both sets of CNVs (categorized by different filtering thresholds) are highlighted as following: */* significant in both; -/- not significant in both; */- or -/* significant in one but not the other.

		>50 KB >50 PROBES			**>10 KB >10 PROBES >1 PROBE/5KB DENSITY**		
STUDY	TYPE	N CNV UN/AFF	OR (95% CI)	P VAL	N CNV UN/AFF	OR (95% CI)	P VAL
GENOMIC CONSTRAINT							
MGS	CNV */*	1328/1509	1.14 (1.04-1.26)	5.88E-03	2667/2968	1.13 (1.06-1.2)	5.83E-05
	DUP -/-	902/928	1.08 (0.96-1.22)	2.05E-01	1582/1631	1.04 (0.96-1.13)	3.38E-01
	DEL */*	426/581	1.35 (1.14-1.59)	3.52E-04	1085/1337	1.31 (1.19-1.44)	3.72E-08
CLOZUK1	CNV -/-	2046/1498	1.1 (1.01-1.19)	2.90E-02	5083/3145	1.01 (0.96-1.06)	7.80E-01
	DUP -/-	1568/1102	1.05 (0.94-1.16)	4.00E-01	3160/1975	1.01 (0.94-1.09)	7.72E-01
	DEL */-	478/396	1.27 (1.09-1.49)	2.89E-03	1923/1170	1 (0.92-1.09)	9.81E-01
CLOZUK2	CNV -/*	2124/1456	1.04 (0.96-1.13)	3.47E-01	6527/4235	1.09 (1.04-1.14)	6.20E-04
	DUP -/-	1577/1026	1.01 (0.91-1.13)	8.36E-01	3700/2408	1.03 (0.96-1.1)	3.64E-01
	DEL -/*	547/430	1.13 (0.99-1.3)	7.13E-02	2827/1827	1.15 (1.07-1.24)	9.87E-05
META-ANALYSIS	CNV */*	5498/4463	1.09 (1.03-1.15)	3.21E-03	14277/10348	1.07 (1.03-1.1)	1.21E-04
	DUP -/-	4047/3056	1.04 (0.97-1.11)	2.28E-01	8442/6014	1.03 (0.98-1.07)	2.34E-01
	DEL */*	1451/1407	1.22 (1.1-1.36)	2.86E-04	5835/4334	1.11 (1.06-1.17)	7.52E-05
NON-CODING CONSTRAINT							
MGS	CNV -/*	1328/1509	1.08 (0.96-1.21)	2.01E-01	2638/2947	1.1 (1.03-1.18)	5.09E-03
	DUP -/-	902/928	1.01 (0.87-1.17)	8.91E-01	1577/1622	1.03 (0.95-1.13)	4.68E-01
	DEL */*	426/581	1.25 (1.04-1.5)	1.79E-02	1061/1325	1.24 (1.11-1.38)	8.72E-05
CLOZUK1	CNV -/-	2046/1498	1.02 (0.93-1.13)	6.21E-01	5056/3131	1.02 (0.96-1.08)	5.89E-01
	DUP -/-	1568/1102	0.93 (0.82-1.05)	2.18E-01	3155/1973	1.05 (0.97-1.14)	2.47E-01
	DEL	478/396	1.28 (1.08-1.52)	4.82E-03	1901/1158	0.98 (0.9-1.07)	6.22E-01

	*/-						
CLOZUK2	CNV -/*	2124/1456	1.03 (0.94-1.13)	5.13E-01	6509/4226	1.07 (1.01-1.13)	1.58E-02
	DUP -/-	1577/1026	0.97 (0.86-1.1)	6.65E-01	3698/2407	1.02 (0.94-1.1)	6.84E-01
	DEL */*	547/430	1.17 (1-1.36)	4.48E-02	2811/1819	1.12 (1.04-1.21)	4.55E-03
META-ANALYSIS	CNV -/*	5498/4463	1.04 (0.98-1.1)	2.05E-01	14203/10304	1.06 (1.02-1.1)	2.88E-03
	DUP -/-	4047/3056	0.96 (0.89-1.03)	2.88E-01	8430/6002	1.03 (0.98-1.09)	2.10E-01
	DEL */*	1451/1407	1.22 (1.08-1.37)	9.72E-04	5773/4302	1.08 (1.02-1.14)	9.49E-03
CCRE CONSTRAINT							
MGS	CNV */*	1320/1501	1.1 (1-1.22)	6.19E-02	2542/2871	1.09 (1.02-1.16)	8.05E-03
	DUP -/-	901/926	1.04 (0.91-1.19)	5.58E-01	1529/1588	1.02 (0.93-1.11)	6.97E-01
	DEL */*	419/575	1.24 (1.06-1.44)	7.55E-03	1013/1283	1.22 (1.11-1.35)	6.62E-05
CLOZUK1	CNV -/-	2046/1498	1.08 (1-1.18)	6.33E-02	4971/3065	1.04 (0.98-1.09)	1.82E-01
	DUP -/-	1568/1102	1.02 (0.91-1.14)	7.71E-01	3134/1952	1.07 (0.99-1.16)	8.75E-02
	DEL */-	478/396	1.26 (1.08-1.46)	2.71E-03	1837/1113	1.01 (0.93-1.09)	8.74E-01
CLOZUK2	CNV -/*	2121/1456	1.02 (0.94-1.11)	6.53E-01	6321/4123	1.06 (1.01-1.12)	1.65E-02
	DUP -/-	1574/1026	0.95 (0.85-1.07)	4.12E-01	3642/2366	1 (0.93-1.08)	9.87E-01
	DEL */*	547/430	1.15 (1.01-1.32)	3.34E-02	2679/1757	1.13 (1.05-1.21)	1.07E-03
META-ANALYSIS	CNV */*	5487/4455	1.06 (1-1.12)	3.34E-02	13834/10059	1.06 (1.02-1.1)	7.28E-04
	DUP -/-	4043/3054	1 (0.93-1.07)	9.33E-01	8305/5906	1.03 (0.98-1.08)	2.66E-01
	DEL */*	1444/1401	1.21 (1.09-1.33)	2.97E-04	5529/4153	1.09 (1.04-1.15)	9.31E-04
NON-CODING EXON CONSTRAINT							
MGS	CNV -/-	971/1059	0.96 (0.86-1.08)	5.32E-01	1442/1525	1.04 (0.96-1.12)	3.18E-01
	DUP -/-	695/707	0.98 (0.85-1.12)	7.14E-01	933/954	1.02 (0.93-1.12)	6.37E-01
	DEL -/-	276/352	0.96 (0.76-1.21)	7.28E-01	509/571	1.1 (0.95-1.26)	2.13E-01
CLOZUK1	CNV -/-	1745/1299	1.05 (0.96-1.16)	3.11E-01	2519/1677	1.04 (0.96-1.12)	3.14E-01
	DUP -/-	1363/958	0.98 (0.87-1.1)	7.38E-01	1621/1100	0.95 (0.86-1.05)	3.56E-01
	DEL */*	382/341	1.28 (1.08-1.53)	5.17E-03	898/577	1.17 (1.04-1.32)	1.13E-02
CLOZUK2	CNV */-	1740/1185	1.03 (0.94-1.13)	5.66E-01	3120/1976	1.02 (0.94-1.09)	6.80E-01
	DUP -/-	1344/868	0.93 (0.83-1.05)	2.44E-01	1948/1188	0.97 (0.88-1.07)	5.38E-01
	DEL	396/317	1.3 (1.09-1.55)	3.78E-03	1172/788	1.1 (0.98-1.23)	1.16E-01

	/						
META-ANALYSIS	CNV -/*	4456/3543	1.02 (0.96-1.08)	5.71E-01	7081/5178	1.03 (0.99-1.08)	1.84E-01
	DUP -/-	3402/2533	0.96 (0.9-1.03)	2.42E-01	4502/3242	0.98 (0.93-1.04)	5.11E-01
	DEL */*	1054/1010	1.17 (1.02-1.33)	2.07E-02	2579/1936	1.12 (1.03-1.21)	6.26E-03
INTRONIC CONSTRAINT							
MGS	CNV -/*	1291/1476	1.06 (0.93-1.2)	4.01E-01	2505/2811	1.06 (1-1.13)	5.15E-02
	DUP -/-	884/909	0.96 (0.82-1.13)	6.25E-01	1504/1557	1.02 (0.94-1.11)	5.91E-01
	DEL */*	407/567	1.3 (1.06-1.61)	1.36E-02	1001/1254	1.16 (1.05-1.28)	4.90E-03
CLOZUK1	CNV -/-	2035/1496	1.05 (0.95-1.16)	3.63E-01	4724/2935	1.05 (0.98-1.11)	1.51E-01
	DUP -/*	1563/1101	0.96 (0.85-1.1)	5.71E-01	2877/1812	1.11 (1.01-1.21)	2.59E-02
	DEL */-	472/395	1.28 (1.06-1.54)	8.79E-03	1847/1123	0.99 (0.9-1.08)	7.77E-01
CLOZUK2	CNV */*	2112/1447	1.14 (1.03-1.27)	9.86E-03	6071/3856	1.07 (1.01-1.13)	2.65E-02
	DUP -/-	1570/1019	1.11 (0.97-1.27)	1.21E-01	3422/2200	1.08 (1-1.17)	5.59E-02
	DEL */-	542/428	1.24 (1.05-1.47)	1.09E-02	2649/1656	1.04 (0.96-1.13)	3.53E-01
META-ANALYSIS	CNV */*	5438/4419	1.08 (1.01-1.16)	2.32E-02	13300/9602	1.06 (1.02-1.1)	2.27E-03
	DUP -/*	4017/3029	1.01 (0.93-1.09)	8.98E-01	7803/5569	1.07 (1.01-1.12)	1.56E-02
	DEL */-	1421/1390	1.27 (1.11-1.45)	6.04E-04	5497/4033	1.04 (0.99-1.1)	1.33E-01

Supplementary Table 14. Non-genic CNVs results.

Results are separated by the type of constraint tested. For a full explanation of different categories, see 3.5 Methods. CNV = duplications and deletions together; DUP = duplications; DEL = deletions; KB = kilobases; N CNV UN/AFF = number of variants in controls versus cases. Statistically significant results ($p < 0.05$) in both sets of CNVs (categorized by different filtering thresholds) are highlighted as following: */* significant in both; -/- not significant in both; */- or -/* significant in one but not the other.

		>50 KB >50 PROBES			**>10 KB >10 PROBES >1 PROBE/5KB DENSITY**		
STUDY	TYPE	N CNV UN/AFF	OR (95% CI)	P VAL	N CNV UN/AFF	OR (95% CI)	P VAL
GENOMIC CONSTRAINT							
MGS	CNV -/-	977/828	1.04 (0.92-1.18)	5.06E-01	4085/3673	1.01 (0.96-1.07)	6.66E-01
	DUP -/-	413/368	0.98 (0.81-1.19)	8.47E-01	1396/1319	0.91 (0.82-1)	5.87E-02
	DEL -/-	564/460	1.09 (0.92-1.3)	3.29E-01	2689/2354	1.06 (0.99-1.14)	9.55E-02
CLOZUK1	CNV -/-	614/360	0.95 (0.84-1.08)	4.42E-01	4312/2657	1 (0.95-1.05)	9.54E-01
	DUP -/*	351/230	0.97 (0.8-1.18)	7.82E-01	1663/1106	0.92 (0.84-1)	4.67E-02
	DEL -/-	263/130	0.86 (0.71-1.03)	1.04E-01	2649/1551	1.04 (0.97-1.1)	2.65E-01
CLOZUK2	CNV -/-	971/589	1.03 (0.92-1.15)	6.50E-01	6937/4296	1.01 (0.97-1.06)	5.26E-01
	DUP -/-	526/321	1.13 (0.93-1.37)	2.16E-01	2587/1557	1.01 (0.93-1.09)	8.49E-01
	DEL -/-	445/268	0.97 (0.83-1.13)	6.78E-01	4350/2739	1.02 (0.97-1.08)	4.04E-01
META-ANALYSIS	CNV -/-	2562/1777	1.01 (0.94-1.08)	8.85E-01	15334/10626	1.01 (0.98-1.04)	5.16E-01
	DUP -/*	1290/919	1.02 (0.91-1.14)	7.77E-01	5646/3982	0.95 (0.9-0.99)	2.33E-02
	DEL -/-	1272/858	0.95 (0.87-1.05)	3.39E-01	9688/6644	1.04 (1-1.07)	5.54E-02
NON-CODING CONSTRAINT							
MGS	CNV -/-	977/828	1.04 (0.92-1.18)	5.06E-01	4085/3673	1.01 (0.96-1.07)	6.63E-01
	DUP -/-	413/368	0.98 (0.81-1.19)	8.47E-01	1396/1319	0.91 (0.82-1)	5.87E-02
	DEL -/-	564/460	1.09 (0.92-1.3)	3.29E-01	2689/2354	1.06 (0.99-1.14)	9.44E-02
CLOZUK1	CNV -/-	614/360	0.95 (0.84-1.08)	4.42E-01	4312/2657	1 (0.95-1.05)	9.60E-01
	DUP -/*	351/230	0.97 (0.8-1.18)	7.82E-01	1663/1106	0.92 (0.84-1)	4.67E-02
	DEL	263/130	0.86 (0.71-1.03)	1.04E-01	2649/1551	1.04 (0.97-1.1)	2.69E-01

	-/-						
CLOZUK2	CNV -/-	971/589	1.03 (0.92-1.15)	6.34E-01	6937/4296	1.01 (0.97-1.06)	5.19E-01
	DUP -/-	526/321	1.13 (0.94-1.37)	2.02E-01	2587/1557	1.01 (0.93-1.09)	8.38E-01
	DEL -/-	445/268	0.97 (0.83-1.13)	6.78E-01	4350/2739	1.02 (0.97-1.08)	4.01E-01
META-ANALYSIS	CNV -/-	2562/1777	1.01 (0.94-1.08)	8.76E-01	15334/10626	1.01 (0.98-1.04)	5.13E-01
	DUP -/*	1290/919	1.02 (0.91-1.14)	7.68E-01	5646/3982	0.95 (0.9-0.99)	2.37E-02
	DEL -/-	1272/858	0.95 (0.87-1.05)	3.39E-01	9688/6644	1.04 (1-1.07)	5.53E-02
CCRE CONSTRAINT							
MGS	CNV -/-	964/816	1.1 (0.98-1.24)	1.17E-01	3741/3328	1.02 (0.97-1.08)	4.36E-01
	DUP -/-	412/367	1.03 (0.86-1.24)	7.40E-01	1305/1218	0.94 (0.86-1.03)	2.14E-01
	DEL -/-	552/449	1.15 (0.98-1.35)	8.40E-02	2436/2110	1.06 (0.99-1.13)	1.03E-01
CLOZUK1	CNV -/-	614/360	0.98 (0.87-1.11)	7.68E-01	4184/2570	1.03 (0.98-1.08)	2.65E-01
	DUP -/-	351/230	1 (0.82-1.21)	9.71E-01	1639/1087	0.95 (0.88-1.04)	2.64E-01
	DEL -/-	263/130	0.86 (0.72-1.03)	1.02E-01	2545/1483	1.05 (0.99-1.12)	9.48E-02
CLOZUK2	CNV -/-	971/588	1.05 (0.94-1.17)	4.12E-01	6659/4128	1.01 (0.97-1.05)	5.37E-01
	DUP -/-	526/321	1.1 (0.92-1.32)	3.04E-01	2519/1510	0.99 (0.92-1.07)	8.41E-01
	DEL -/-	445/267	1.02 (0.88-1.18)	8.22E-01	4140/2618	1.03 (0.98-1.08)	2.59E-01
META-ANALYSIS	CNV -/-	2549/1764	1.04 (0.97-1.11)	2.89E-01	14584/10026	1.02 (0.99-1.05)	1.65E-01
	DUP -/-	1289/918	1.04 (0.93-1.16)	4.91E-01	5463/3815	0.97 (0.92-1.01)	1.35E-01
	DEL -/*	1260/846	0.99 (0.9-1.09)	8.05E-01	9121/6211	1.04 (1.01-1.08)	1.76E-02
NON-CODING EXON CONSTRAINT							
MGS	CNV -/-	656/559	0.98 (0.86-1.12)	8.09E-01	1511/1385	1 (0.92-1.09)	9.25E-01
	DUP -/-	314/249	0.96 (0.81-1.14)	6.25E-01	631/548	1.01 (0.89-1.15)	8.69E-01
	DEL -/-	342/310	1.04 (0.85-1.27)	6.98E-01	880/837	1.01 (0.9-1.13)	9.14E-01
CLOZUK1	CNV -/-	553/315	1.01 (0.86-1.17)	9.26E-01	1846/1132	0.96 (0.89-1.04)	2.91E-01
	DUP -/-	331/208	0.98 (0.79-1.21)	8.40E-01	750/497	0.94 (0.83-1.07)	3.53E-01
	DEL -/-	222/107	0.93 (0.73-1.19)	5.82E-01	1096/635	0.95 (0.86-1.05)	3.47E-01
CLOZUK2	CNV -/*	802/509	1.03 (0.91-1.16)	6.17E-01	2903/1858	1.14 (1.07-1.2)	2.64E-05
	DUP -/-	440/283	1.03 (0.86-1.23)	7.48E-01	1183/748	1.11 (1-1.23)	5.17E-02
	DEL	362/226	1.02 (0.86-1.21)	8.17E-01	1720/1110	1.15 (1.07-1.24)	1.09E-04

	-/*						
META-ANALYSIS	CNV -/-	2011/1383	1.01 (0.93-1.09)	8.41E-01	6260/4375	1.04 (1-1.09)	7.48E-02
	DUP -/-	1085/740	0.99 (0.89-1.1)	8.06E-01	2564/1793	1.02 (0.95-1.1)	5.57E-01
	DEL -/-	926/643	1 (0.89-1.13)	9.56E-01	3696/2582	1.05 (0.99-1.11)	9.53E-02
INTRONIC CONSTRAINT							
MGS	CNV -/-	637/523	1.02 (0.89-1.17)	7.57E-01	2254/1970	1.05 (0.98-1.13)	1.66E-01
	DUP -/-	273/223	0.97 (0.77-1.22)	7.81E-01	743/665	1 (0.87-1.15)	9.87E-01
	DEL -/-	364/300	1.07 (0.9-1.28)	4.31E-01	1511/1305	1.07 (0.98-1.16)	1.11E-01
CLOZUK1	CNV -/-	537/310	0.93 (0.81-1.08)	3.52E-01	2431/1460	0.99 (0.93-1.06)	7.52E-01
	DUP -/-	316/200	0.94 (0.74-1.2)	6.18E-01	840/578	0.94 (0.84-1.04)	2.27E-01
	DEL -/-	221/110	0.83 (0.68-1.02)	7.29E-02	1591/882	1 (0.93-1.09)	9.19E-01
CLOZUK2	CNV -/-	763/458	0.98 (0.86-1.11)	7.11E-01	3736/2365	1.01 (0.96-1.07)	6.62E-01
	DUP -/-	388/244	1.07 (0.85-1.35)	5.71E-01	1317/840	1.06 (0.96-1.17)	2.76E-01
	DEL -/-	375/214	0.92 (0.78-1.07)	2.68E-01	2419/1525	0.99 (0.93-1.06)	8.13E-01
META-ANALYSIS	CNV -/-	1937/1291	0.98 (0.9-1.05)	5.28E-01	8421/5795	1.01 (0.98-1.05)	4.59E-01
	DUP -/-	977/667	0.99 (0.86-1.13)	8.45E-01	2900/2083	0.99 (0.93-1.06)	8.44E-01
	DEL -/-	960/624	0.92 (0.84-1.02)	1.02E-01	5521/3712	1.01 (0.97-1.06)	5.22E-01

Supplementary Table 15. Genic CNVs results, correcting for the maximum Z score from coding exons.

Results are separated by the type of constraint tested. For a full explanation of different categories, see 3.5 Methods. CNV = duplications and deletions together; DUP = duplications; DEL = deletions; KB = kilobases; N CNV UN/AFF = number of variants in controls versus cases. Statistically significant results ($p < 0.05$) in both sets of CNVs (categorized by different filtering thresholds) are highlighted as following: */* significant in both; -/- not significant in both; */- or -/* significant in one but not the other.

		>50 KB >50 PROBES			**>10 KB >10 PROBES >1 PROBE/5KB DENSITY**		
STUDY	TYPE	N CNV UN/AFF	OR (95% CI)	P VAL	N CNV UN/AFF	OR (95% CI)	P VAL
NON-CODING CONSTRAINT							
MGS	CNV -/-	1328/1509	0.92 (0.79-1.09)	3.39E-01	2638/2947	1.01 (0.91-1.12)	8.72E-01
	DUP -/-	902/928	0.9 (0.74-1.09)	2.85E-01	1577/1622	0.97 (0.85-1.11)	6.75E-01
	DEL -/-	426/581	0.96 (0.71-1.29)	7.72E-01	1061/1325	1.06 (0.88-1.28)	5.38E-01
CLOZUK1	CNV */-	2046/1498	0.88 (0.78-1)	4.50E-02	5056/3131	0.97 (0.89-1.06)	5.08E-01
	DUP */	1568/1102	0.81 (0.7-0.93)	3.03E-03	3155/1973	1.02 (0.92-1.13)	7.40E-01
	DEL -/-	478/396	1.17 (0.91-1.52)	2.16E-01	1901/1158	0.89 (0.78-1.03)	1.14E-01
CLOZUK2	CNV -/-	2124/1456	0.99 (0.87-1.12)	8.63E-01	6509/4226	1 (0.93-1.08)	9.77E-01
	DUP -/-	1577/1026	0.92 (0.79-1.07)	2.71E-01	3698/2407	1 (0.9-1.11)	9.68E-01
	DEL -/-	547/430	1.15 (0.91-1.46)	2.46E-01	2811/1819	0.98 (0.87-1.1)	7.39E-01
META-ANALYSIS	CNV */-	5498/4463	0.93 (0.86-0.99)	3.55E-02	14203/10304	0.99 (0.94-1.04)	7.46E-01
	DUP */	4047/3056	0.86 (0.79-0.93)	1.79E-04	8430/6002	1 (0.94-1.07)	9.66E-01
	DEL -/-	1451/1407	1.09 (0.92-1.29)	3.17E-01	5773/4302	0.96 (0.89-1.03)	2.72E-01
CCRE CONSTRAINT							
MGS	CNV -/-	1320/1501	0.95 (0.82-1.1)	4.92E-01	2542/2871	0.99 (0.9-1.1)	9.13E-01
	DUP -/-	901/926	0.92 (0.77-1.11)	3.96E-01	1529/1588	0.94 (0.82-1.07)	3.43E-01
	DEL -/-	419/575	0.95 (0.75-1.22)	7.09E-01	1013/1283	1.06 (0.9-1.25)	4.84E-01
CLOZUK1	CNV -/-	2046/1498	0.97 (0.87-1.09)	6.34E-01	4971/3065	1.01 (0.94-1.09)	7.72E-01
	DUP -/-	1568/1102	0.91 (0.8-1.04)	1.87E-01	3134/1952	1.05 (0.95-1.16)	3.25E-01

	DEL -/-	478/396	1.17 (0.93-1.47)	1.73E-01	1837/1113	0.96 (0.85-1.08)	4.52E-01
CLOZUK2	CNV -/-	2121/1456	0.96 (0.86-1.08)	5.02E-01	6321/4123	1 (0.93-1.07)	9.77E-01
	DUP -/-	1574/1026	0.88 (0.76-1.02)	9.93E-02	3642/2366	0.97 (0.88-1.07)	5.55E-01
	DEL -/-	547/430	1.12 (0.91-1.38)	2.99E-01	2679/1757	1.01 (0.91-1.12)	8.20E-01
META-ANALYSIS	CNV -/-	5487/4455	0.96 (0.9-1.03)	2.79E-01	13834/10059	1 (0.96-1.05)	9.17E-01
	DUP */-	4043/3054	0.91 (0.84-0.98)	1.32E-02	8305/5906	0.99 (0.93-1.05)	7.33E-01
	DEL -/	1444/1401	1.07 (0.93-1.23)	3.50E-01	5529/4153	1 (0.93-1.07)	9.46E-01
NON-CODING EXON CONSTRAINT							
MGS	CNV -/-	971/1059	0.88 (0.76-1.02)	8.71E-02	1442/1525	0.98 (0.87-1.12)	7.97E-01
	DUP -/-	695/707	0.88 (0.75-1.04)	1.44E-01	933/954	0.95 (0.82-1.1)	4.97E-01
	DEL -/-	276/352	0.88 (0.65-1.19)	4.09E-01	509/571	1.07 (0.84-1.36)	5.75E-01
CLOZUK1	CNV -/-	1745/1299	0.94 (0.84-1.06)	3.14E-01	2519/1677	1.01 (0.91-1.11)	8.88E-01
	DUP -/-	1363/958	0.9 (0.79-1.03)	1.25E-01	1621/1100	0.92 (0.81-1.04)	1.76E-01
	DEL -/-	382/341	1.09 (0.87-1.36)	4.69E-01	898/577	1.16 (0.99-1.36)	7.44E-02
CLOZUK2	CNV -/-	1740/1185	0.93 (0.82-1.05)	2.20E-01	3120/1976	0.94 (0.86-1.03)	1.96E-01
	DUP */-	1344/868	0.86 (0.75-0.99)	3.70E-02	1948/1188	0.93 (0.83-1.04)	1.82E-01
	DEL -/-	396/317	1.11 (0.88-1.41)	3.65E-01	1172/788	0.95 (0.82-1.1)	4.84E-01
META-ANALYSIS	CNV */-	4456/3543	0.92 (0.86-0.98)	1.50E-02	7081/5178	0.97 (0.92-1.03)	3.39E-01
	DUP */*	3402/2533	0.88 (0.82-0.95)	8.45E-04	4502/3242	0.93 (0.87-0.99)	3.19E-02
	DEL -/-	1054/1010	1.03 (0.88-1.19)	7.23E-01	2579/1936	1.03 (0.93-1.14)	6.08E-01
INTRONIC CONSTRAINT							
MGS	CNV -/-	1291/1476	0.89 (0.75-1.05)	1.51E-01	2505/2811	0.95 (0.85-1.05)	3.15E-01
	DUP -/-	884/909	0.85 (0.69-1.04)	1.12E-01	1504/1557	0.95 (0.83-1.08)	4.52E-01
	DEL -/-	407/567	0.92 (0.68-1.24)	5.88E-01	1001/1254	0.93 (0.78-1.11)	4.27E-01
CLOZUK1	CNV -/-	2035/1496	0.9 (0.79-1.02)	1.08E-01	4724/2935	1.01 (0.93-1.1)	8.38E-01
	DUP */-	1563/1101	0.84 (0.72-0.97)	2.20E-02	2877/1812	1.09 (0.98-1.22)	1.21E-01
	DEL -/-	472/395	1.11 (0.85-1.44)	4.60E-01	1847/1123	0.9 (0.79-1.03)	1.13E-01
CLOZUK2	CNV */-	2112/1447	1.14 (1-1.3)	4.51E-02	6071/3856	1.01 (0.94-1.09)	7.07E-01
	DUP */-	1570/1019	1.1 (0.94-1.29)	2.42E-01	3422/2200	1.09 (0.98-1.2)	1.05E-01

	DEL -/-	542/428	1.22 (0.96-1.55)	1.08E-01	2649/1656	0.93 (0.83-1.03)	1.74E-01
	CNV -/-	5438/4419	0.96 (0.88-1.03)	2.56E-01	13300/9602	1 (0.95-1.05)	8.71E-01
META-ANALYSIS	DUP */-	4017/3029	0.9 (0.82-0.98)	2.11E-02	7803/5569	1.05 (0.98-1.12)	1.96E-01
	DEL -/-	1421/1390	1.07 (0.9-1.26)	4.39E-01	5497/4033	0.92 (0.86-0.98)	1.62E-02

Supplementary Table 16. Sum Z score.

CNV = duplications and deletions together; DUP = duplications; DEL = deletions; KB = kilobases; N CNV UN/AFF = number of variants in controls versus cases. Statistically significant results ($p < 0.05$) in both sets of CNVs (categorized by different filtering thresholds) are highlighted as following: */* significant in both; -/- not significant in both; */- or -/* significant in one but not the other.

		>50 KB >50 PROBES			**>10 KB >10 PROBES >1 PROBE/5KB DENSITY**		
STUDY	TYPE	N CNV UN/AFF	OR (95% CI)	P VAL	N CNV UN/AFF	OR (95% CI)	P VAL
GENOMIC CONSTRAINT							
MGS	CNV */*	2305/2337	1 (1-1)	2.60E-03	6752/6641	1 (1-1)	2.67E-04
	DUP -/-	1315/1296	1 (1-1)	3.38E-01	2978/2950	1 (1-1)	3.63E-01
	DEL */*	990/1041	1 (1-1)	2.57E-03	3774/3691	1 (1-1)	1.22E-05
CLOZUK1	CNV */-	2660/1858	1 (1-1)	3.31E-05	9395/5802	1 (1-1)	9.18E-02
	DUP */-	1919/1332	1 (1-1)	2.96E-04	4823/3081	1 (1-1)	5.20E-01
	DEL */*	741/526	1 (1-1)	9.34E-03	4572/2721	1 (1-1)	3.77E-02
CLOZUK2	CNV */*	3095/2045	1 (1-1)	9.73E-05	13464/8531	1 (1-1)	1.13E-03
	DUP */-	2103/1347	1 (1-1)	2.06E-02	6287/3965	1 (1-1)	2.20E-01
	DEL */*	992/698	1 (1-1)	1.20E-04	7177/4566	1 (1-1)	5.64E-05
META-ANALYSIS	CNV */*	8060/6240	1 (1-1)	1.28E-10	29611/20974	1 (1-1)	8.78E-07
	DUP */-	5337/3975	1 (1-1)	2.94E-05	14088/9996	1 (1-1)	1.02E-01
	DEL */*	2723/2265	1 (1-1)	1.03E-07	15523/10978	1 (1-1)	4.21E-09

Supplementary Table 17. Mean Z score.

CNV = duplications and deletions together; DUP = duplications; DEL = deletions; KB = kilobases; N CNV UN/AFF = number of variants in controls versus cases. Statistically significant results ($p < 0.05$) in both sets of CNVs (categorized by different filtering thresholds) are highlighted as following: */* significant in both; -/- not significant in both; */- or -/* significant in one but not the other.

		>50 KB >50 PROBES			**>10 KB >10 PROBES >1 PROBE/5KB DENSITY**		
STUDY	TYPE	N CNV UN/AFF	OR (95% CI)	P VAL	N CNV UN/AFF	OR (95% CI)	P VAL
GENOMIC CONSTRAINT							
MGS	CNV */*	2305/2337	1.18 (1.09-1.27)	3.64E-05	6752/6641	1.11 (1.07-1.16)	9.03E-07
	DUP -/-	1315/1296	1.09 (0.98-1.21)	1.31E-01	2978/2950	1.02 (0.96-1.09)	5.16E-01
	DEL */*	990/1041	1.28 (1.14-1.44)	3.47E-05	3774/3691	1.17 (1.11-1.23)	2.69E-08
CLOZUK1	CNV */-	2660/1858	1.08 (1-1.17)	4.86E-02	9395/5802	1.01 (0.98-1.05)	4.95E-01
	DUP -/-	1919/1332	1.03 (0.93-1.14)	5.61E-01	4823/3081	0.98 (0.92-1.03)	4.14E-01
	DEL */-	741/526	1.15 (1.02-1.3)	2.38E-02	4572/2721	1.03 (0.98-1.08)	2.23E-01
CLOZUK2	CNV -/-	3095/2045	1.05 (0.98-1.13)	1.48E-01	13464/8531	1.03 (1-1.07)	5.05E-02
	DUP -/-	2103/1347	1.02 (0.92-1.13)	7.17E-01	6287/3965	1.03 (0.98-1.09)	2.73E-01
	DEL */-	992/698	1.12 (1.01-1.25)	2.91E-02	7177/4566	1.04 (1-1.08)	7.68E-02
META-ANALYSIS	CNV */*	8060/6240	1.09 (1.04-1.15)	1.92E-04	29611/20974	1.04 (1.02-1.07)	2.10E-04
	DUP -/-	5337/3975	1.04 (0.98-1.11)	1.98E-01	14088/9996	1.01 (0.97-1.04)	6.42E-01
	DEL */*	2723/2265	1.17 (1.09-1.27)	5.74E-05	15523/10978	1.06 (1.03-1.09)	5.97E-05

Supplementary Table 18. Median Z score.

CNV = duplications and deletions together; DUP = duplications; DEL = deletions; KB = kilobases; N CNV UN/AFF = number of variants in controls versus cases. Statistically significant results ($p < 0.05$) in both sets of CNVs (categorized by different filtering thresholds) are highlighted as following: */* significant in both; -/- not significant in both; */- or -/* significant in one but not the other.

		>50 KB >50 PROBES			**>10 KB >10 PROBES >1 PROBE/5KB DENSITY**		
STUDY	TYPE	N CNV UN/AFF	OR (95% CI)	P VAL	N CNV UN/AFF	OR (95% CI)	P VAL
GENOMIC CONSTRAINT							
MGS	CNV */*	2305/2337	1.17 (1.08-1.27)	6.93E-05	6752/6641	1.11 (1.06-1.16)	1.54E-06
	DUP -/-	1315/1296	1.08 (0.97-1.21)	1.47E-01	2978/2950	1.03 (0.96-1.1)	4.76E-01
	DEL */*	990/1041	1.27 (1.13-1.42)	6.81E-05	3774/3691	1.16 (1.1-1.23)	8.32E-08
CLOZUK1	CNV */-	2660/1858	1.08 (1-1.17)	4.49E-02	9395/5802	1.01 (0.98-1.05)	5.45E-01
	DUP -/-	1919/1332	1.03 (0.93-1.15)	5.38E-01	4823/3081	0.97 (0.92-1.03)	3.66E-01
	DEL */-	741/526	1.15 (1.02-1.29)	2.46E-02	4572/2721	1.03 (0.98-1.08)	2.46E-01
CLOZUK2	CNV -/-	3095/2045	1.05 (0.98-1.13)	1.43E-01	13464/8531	1.03 (1-1.07)	5.89E-02
	DUP -/-	2103/1347	1.02 (0.92-1.13)	6.60E-01	6287/3965	1.03 (0.98-1.09)	2.50E-01
	DEL */-	992/698	1.12 (1.01-1.24)	3.52E-02	7177/4566	1.04 (0.99-1.08)	1.01E-01
META-ANALYSIS	CNV */*	8060/6240	1.09 (1.04-1.15)	2.20E-04	29611/20974	1.04 (1.02-1.06)	3.58E-04
	DUP -/-	5337/3975	1.04 (0.98-1.11)	1.86E-01	14088/9996	1.01 (0.97-1.04)	6.41E-01
	DEL */*	2723/2265	1.17 (1.08-1.26)	8.69E-05	15523/10978	1.06 (1.03-1.09)	1.34E-04

Chapter 4

Supplementary Table 19. Transcription factor binding to cCREs in proximity to PSG8-AS1 genomic location.

Transcription factor binding information was gathered from ENCODE and ORegAnno. Distance is shown in base pairs relative to the start of *PSG8-AS1* exonic sequence in hg38. TF mentioned in the text are highlighted.

CCRE ENCODE NAME	TYPE OF CCRE	TF NAME	FUNCTION	DISTANCE
EH38E1955821	Distal-enhancer	ATF2	General Gene Regulation	-22634
EH38E1955811	Distal-enhancer	ATF3	General Gene Regulation	-9538
EH38E1955812	Distal-enhancer	ATF3	General Gene Regulation	-9908
EH38E1955811	Distal-enhancer	BCL3	General Gene Regulation	-9538
EH38E1955812	Distal-enhancer	BCL3	General Gene Regulation	-9908
EH38E1955800	CTCF-only	BCOR	General Gene Regulation	30521
EH38E1955804	Distal-enhancer	C11orf30	General Gene Regulation	21523
EH38E1955805	Proximal-enhancer	CEBPB	General Gene Regulation	1784
EH38E1955811	Distal-enhancer	CEBPB	General Gene Regulation	-9538
EH38E1955811	Distal-enhancer	CREB1	General Gene Regulation	-9538
EH38E1955812	Distal-enhancer	CREB1	General Gene Regulation	-9908
EH38E1955811	Distal-enhancer	CTBP1	General Gene Regulation	-9538
EH38E1955799	CTCF-only	CTCF	General Gene Regulation	34202
EH38E1955800	CTCF-only	CTCF	General Gene Regulation	30521
EH38E1955801	Distal-enhancer	CTCF	General Gene Regulation	28208
EH38E1955802	Distal-enhancer	CTCF	General Gene Regulation	27952
EH38E1955811	Distal-enhancer	CTCF	General Gene Regulation	-9538
EH38E1955811	Distal-enhancer	CUX	CNS Gene Regulation	-9538
EH38E1955812	Distal-enhancer	CUX1	CNS Gene Regulation	-9908
EH38E1955800	CTCF-only	DACH1	General Gene Regulation	30521
EH38E1955811	Distal-enhancer	DPF2	General Gene Regulation	-9538

EH38E1955812	Distal-enhancer	DPF2	General Gene Regulation	-9908
EH38E1955811	Distal-enhancer	ELF1	General Gene Regulation	-9538
EH38E1955812	Distal-enhancer	ELF1	General Gene Regulation	-9908
EH38E1955811	Distal-enhancer	ESRRA	General Gene Regulation	-9538
EH38E1955812	Distal-enhancer	ESRRA	General Gene Regulation	-9908
EH38E1955805	Proximal-enhancer	FOS	General Gene Regulation	1784
EH38E1955811	Distal-enhancer	FOS	General Gene Regulation	-9538
EH38E1955812	Distal-enhancer	FOS	General Gene Regulation	-9908
EH38E1955814	Distal-enhancer	FOS	General Gene Regulation	-11060
EH38E1955815	Distal-enhancer	FOS	General Gene Regulation	-11389
EH38E1955811	Distal-enhancer	FOSL2	General Gene Regulation	-9538
EH38E1955811	Distal-enhancer	FOSL2	General Gene Regulation	-9677
EH38E1955810	Distal-enhancer	FOXA1	General Gene Regulation	-8011
EH38E1955811	Distal-enhancer	FOXA1	General Gene Regulation	-9538
EH38E1955811	Distal-enhancer	FOXA2	General Gene Regulation	-9538
EH38E1955810	Distal-enhancer	GATA3	General Gene Regulation	-8011
EH38E1955811	Distal-enhancer	GATA3	General Gene Regulation	-9538
EH38E1955813	Distal-enhancer	GATA3	General Gene Regulation	-10623
EH38E1955811	Distal-enhancer	GATAD2B	CNS Gene Regulation	-9538
EH38E1955811	Distal-enhancer	JUN	General Gene Regulation	-9538
EH38E1955805	Proximal-enhancer	JUND	General Gene Regulation	1784
EH38E1955811	Distal-enhancer	JUND	General Gene Regulation	-9538
EH38E1955805	Proximal-enhancer	MAFK	General Gene Regulation	1784
EH38E1955806	Proximal-enhancer	MAFK	General Gene Regulation	1445
EH38E1955811	Distal-enhancer	MAFK	General Gene Regulation	-9538
EH38E1955813	Distal-enhancer	MAFK	General Gene Regulation	-10623
EH38E1955811	Distal-enhancer	MTA1	General Gene Regulation	-9538
EH38E1955808	Promoter	NA	NA	63
EH38E1955809	Proximal-enhancer	NA	NA	-1434
EH38E1955818	Distal-enhancer	NA	NA	-13012

EH38E1955822	Distal-enhancer	NA	NA	-23951
EH38E1955823	Distal-enhancer	NA	NA	-27964
EH38E1955811	Distal-enhancer	NCOA3	General Gene Regulation	-9538
EH38E1955805	Proximal-enhancer	NFE2L2	General Gene Regulation	1784
EH38E1955811	Distal-enhancer	NFIB	CNS Gene Regulation	-9538
EH38E1955812	Distal-enhancer	NFIB	CNS Gene Regulation	-9908
EH38E1955801	Distal-enhancer	NFIC	General Gene Regulation	28208
EH38E1955811	Distal-enhancer	NR3C1	General Gene Regulation	-9538
EH38E1955811	Distal-enhancer	PKNOX1	CNS Gene Regulation	-9538
EH38E1955814	Distal-enhancer	PKNOX1	CNS Gene Regulation	-11060
EH38E1955815	Distal-enhancer	PKNOX1	CNS Gene Regulation	-11389
EH38E1955800	CTCF-only	RAD21	General Gene Regulation	30521
EH38E1955801	Distal-enhancer	RAD21	General Gene Regulation	28208
EH38E1955811	Distal-enhancer	RAD21	General Gene Regulation	-9538
EH38E1955800	CTCF-only	RBM25	General Gene Regulation	30521
EH38E1955811	Distal-enhancer	RCOR1	REST complex	-9538
EH38E1955812	Distal-enhancer	RCOR1	REST complex	-9908
EH38E1955800	CTCF-only	REST	REST complex	30521
EH38E1955801	Distal-enhancer	REST	REST complex	28208
EH38E1955802	Distal-enhancer	REST	REST complex	27952
EH38E1955811	Distal-enhancer	REST	REST complex	-9538
EH38E1955812	Distal-enhancer	REST	REST complex	-9908
EH38E1955811	Distal-enhancer	RFX5	General Gene Regulation	-9538
EH38E1955812	Distal-enhancer	RFX5	General Gene Regulation	-9908
EH38E1955811	Distal-enhancer	SIN3A	REST complex	-9538
EH38E1955812	Distal-enhancer	SIN3A	REST complex	-9908
EH38E1955813	Distal-enhancer	SIN3A	REST complex	-10623
EH38E1955800	CTCF-only	SMARCA5	General Gene Regulation	30521
EH38E1955811	Distal-enhancer	SMARCE1	REST complex	-9538
EH38E1955801	Distal-enhancer	SMC3	NA	28208
EH38E1955801	Distal-enhancer	SP1	General Gene Regulation	28208
EH38E1955802	Distal-enhancer	SP1	General Gene Regulation	27952
EH38E1955811	Distal-enhancer	SP1	General Gene Regulation	-9538
EH38E1955812	Distal-enhancer	SP1	General Gene Regulation	-9908
EH38E1955813	Distal-enhancer	SP1	General Gene Regulation	-10623

EH38E1955814	Distal-enhancer	SP1	General Gene Regulation	-11060
EH38E1955815	Distal-enhancer	SP1	General Gene Regulation	-11389
EH38E1955811	Distal-enhancer	TCF12	CNS Gene Regulation	-9538
EH38E1955813	Distal-enhancer	TCF12	CNS Gene Regulation	-10623
EH38E1955813	**Distal-enhancer**	**TCF7L2**	**Oligodendrocyte Development and Myelination**	**-10777**
EH38E1955805	**Proximal-enhancer**	**TP53**	**General Gene Regulation**	**1461**
EH38E1955806	**Proximal-enhancer**	**TP53**	**General Gene Regulation**	**1330**
EH38E1955806	**Proximal-enhancer**	**TP63**	**General Gene Regulation**	**1271**
EH38E1955801	Distal-enhancer	TRIM22	General Gene Regulation	28208
EH38E1955801	Distal-enhancer	USF2	General Gene Regulation	28208
EH38E1955802	Distal-enhancer	USF2	General Gene Regulation	27952
EH38E1955811	Distal-enhancer	ZBTB33	General Gene Regulation	-9538
EH38E1955812	Distal-enhancer	ZBTB33	General Gene Regulation	-9908
EH38E1955811	Distal-enhancer	ZKSCAN1	General Gene Regulation	-9538
EH38E1955812	Distal-enhancer	ZKSCAN1	General Gene Regulation	-9908
EH38E1955800	CTCF-only	ZMYM3	CNS Gene Regulation	30521
EH38E1955811	Distal-enhancer	ZNF217	General Gene Regulation	-9538
EH38E1955812	Distal-enhancer	ZNF217	General Gene Regulation	-9908
EH38E1955813	**Distal-enhancer**	**ZNF24**	**Oligodendrocyte Development and Myelination**	**-10623**
EH38E1955811	Distal-enhancer	ZNF687	General Gene Regulation	-9538
EH38E1955812	Distal-enhancer	ZNF687	General Gene Regulation	-9908

Supplementary Table 20. Cis-eQTLs for *PSG8-AS1* in the brain.

eQTL information was gathered from GTEx and psychENCODE. Distance is shown in base pairs relative to the start of *PSG8-AS1* exonic sequence in hg38.

LOCATION (HG38)	SOURCE	DISTANCE
chr19:41991305	pshycENCODE	830468
chr19:42069100	pshycENCODE	752673
chr19:42094437	pshycENCODE	727336
chr19:42096832	pshycENCODE	724941
chr19:42107672	pshycENCODE	714101
chr19:42174477	GTEx	647296
chr19:42243381	GTEx	578392
chr19:42370749	GTEx	451024
chr19:42398019	GTEx	423754
chr19:42398020	GTEx	423753
chr19:42398025	GTEx	423748
chr19:42485003	GTEx	336770
chr19:42505685	pshycENCODE	316088
chr19:42526240	pshycENCODE	295533
chr19:42554888	GTEx	266885
chr19:42578891	GTEx	242882
chr19:42638794	pshycENCODE	182979
chr19:42698527	GTEx	123246
chr19:42698618	GTEx	123155
chr19:42699355	GTEx	122418
chr19:42700632	GTEx	121141
chr19:42705963	pshycENCODE	115810
chr19:42707734	GTEx	114039
chr19:42710459	GTEx	111314
chr19:42711349	pshycENCODE	110424
chr19:42713713	GTEx	108060
chr19:42714083	pshycENCODE	107690
chr19:42715583	GTEx	106190
chr19:42715585	GTEx	106188
chr19:42715822	GTEx	105951
chr19:42716214	pshycENCODE	105559
chr19:42717067	pshycENCODE	104706
chr19:42717298	GTEx	104475
chr19:42718403	pshycENCODE	103370
chr19:42719000	pshycENCODE	102773
chr19:42721138	GTEx	100635
chr19:42721654	GTEx/pshycENCODE	100119

chr19:42721884	pshycENCODE	99889
chr19:42724888	pshycENCODE	96885
chr19:42726668	pshycENCODE	95105
chr19:42729152	pshycENCODE	92621
chr19:42732111	pshycENCODE	89662
chr19:42734779	pshycENCODE	86994
chr19:42739413	pshycENCODE	82360
chr19:42740067	pshycENCODE	81706
chr19:42740333	GTEx	81440
chr19:42741311	pshycENCODE	80462
chr19:42741460	pshycENCODE	80313
chr19:42741609	pshycENCODE	80164
chr19:42744873	pshycENCODE	76900
chr19:42745022	GTEx	76751
chr19:42747514	pshycENCODE	74259
chr19:42747693	pshycENCODE	74080
chr19:42749331	GTEx	72442
chr19:42749348	GTEx	72425
chr19:42749538	GTEx	72235
chr19:42750807	pshycENCODE	70966
chr19:42751942	pshycENCODE	69831
chr19:42752081	pshycENCODE	69692
chr19:42752339	pshycENCODE	69434
chr19:42753287	pshycENCODE	68486
chr19:42753667	pshycENCODE	68106
chr19:42754190	pshycENCODE	67583
chr19:42755032	pshycENCODE	66741
chr19:42755394	pshycENCODE	66379
chr19:42755615	pshycENCODE	66158
chr19:42756149	pshycENCODE	65624
chr19:42756402	pshycENCODE	65371
chr19:42761442	GTEx	60331
chr19:42761883	pshycENCODE	59890
chr19:42762384	pshycENCODE	59389
chr19:42763917	GTEx	57856
chr19:42764317	GTEx	57456
chr19:42764337	GTEx	57436
chr19:42765980	GTEx	55793
chr19:42766413	GTEx	55360
chr19:42766786	pshycENCODE	54987
chr19:42766893	GTEx	54880
chr19:42767707	GTEx	54066
chr19:42768287	GTEx	53486
chr19:42768613	GTEx	53160

chr19:42768812	GTEx	52961
chr19:42770145	GTEx	51628
chr19:42770214	GTEx	51559
chr19:42770369	GTEx	51404
chr19:42770716	GTEx	51057
chr19:42771437	GTEx	50336
chr19:42771740	GTEx	50033
chr19:42771957	GTEx	49816
chr19:42772038	GTEx	49735
chr19:42772913	GTEx	48860
chr19:42773287	GTEx	48486
chr19:42774435	GTEx	47338
chr19:42774778	GTEx	46995
chr19:42777920	GTEx	43853
chr19:42778522	pshycENCODE	43251
chr19:42778803	GTEx	42970
chr19:42782560	GTEx	39213
chr19:42785947	GTEx	35826
chr19:42786453	GTEx	35320
chr19:42786586	GTEx	35187
chr19:42786621	GTEx	35152
chr19:42787083	GTEx	34690
chr19:42787517	GTEx	34256
chr19:42804104	GTEx	17669
chr19:42808708	GTEx	13065
chr19:42813818	GTEx	7955
chr19:42814427	GTEx	7346
chr19:42815321	GTEx_DOG	6452
chr19:42815704	GTEx_DOG	6069
chr19:42816533	GTEx	5240
chr19:42817434	GTEx	4339
chr19:42817737	GTEx	4036
chr19:42817914	GTEx	3859
chr19:42818067	GTEx	3706
chr19:42818309	GTEx	3464
chr19:42819098	GTEx	2675
chr19:42819231	GTEx	2542
chr19:42820146	GTEx	1627
chr19:42820427	GTEx	1346
chr19:42820805	GTEx	968
chr19:42821162	GTEx_DOG	611
chr19:42821220	GTEx	553
chr19:42821547	GTEx	226
chr19:42821662	GTEx	111

chr19:42821781	GTEx	-8
chr19:42822081	GTEx	-308
chr19:42822106	GTEx	-333
chr19:42822836	GTEx	-1063
chr19:42824093	GTEx	-2320
chr19:42824189	GTEx	-2416
chr19:42824473	GTEx	-2700
chr19:42824633	GTEx	-2860
chr19:42825528	GTEx	-3755
chr19:42826270	GTEx	-4497
chr19:42826492	GTEx	-4719
chr19:42827445	GTEx	-5672
chr19:42827604	GTEx	-5831
chr19:42827643	GTEx	-5870
chr19:42828102	GTEx	-6329
chr19:42830836	GTEx_DOG	-9063
chr19:42830896	GTEx_DOG	-9123
chr19:42831047	GTEx	-9274
chr19:42831156	GTEx	-9383
chr19:42831423	GTEx	-9650
chr19:42831712	GTEx	-9939
chr19:42832253	GTEx	-10480
chr19:42832592	GTEx	-10819
chr19:42832780	GTEx	-11007
chr19:42832849	GTEx	-11076
chr19:42833011	GTEx	-11238
chr19:42833278	GTEx	-11505
chr19:42833360	GTEx	-11587
chr19:42833369	GTEx	-11596
chr19:42833798	GTEx	-12025
chr19:42833953	GTEx	-12180
chr19:42833994	GTEx	-12221
chr19:42834229	GTEx	-12456
chr19:42834446	GTEx	-12673
chr19:42835396	GTEx	-13623
chr19:42837282	GTEx	-15509
chr19:42837867	GTEx	-16094
chr19:42838329	GTEx	-16556
chr19:42838973	GTEx	-17200
chr19:42840210	GTEx	-18437
chr19:42840495	GTEx	-18722
chr19:42843915	GTEx	-22142
chr19:42844698	GTEx	-22925
chr19:42844917	GTEx	-23144

chr19:42845738	GTEx	-23965
chr19:42847012	GTEx	-25239
chr19:42847870	GTEx	-26097
chr19:42850702	GTEx/pshycENCODE	-28929
chr19:42850796	GTEx	-29023
chr19:42851163	GTEx	-29390
chr19:42851191	GTEx	-29418
chr19:42851639	GTEx	-29866
chr19:42852484	GTEx	-30711
chr19:42853312	GTEx	-31539
chr19:42854123	GTEx	-32350
chr19:42854231	GTEx	-32458
chr19:42855033	pshycENCODE	-33260
chr19:42855446	GTEx	-33673
chr19:42855719	GTEx	-33946
chr19:42855769	GTEx	-33996
chr19:42856373	GTEx	-34600
chr19:42856468	GTEx	-34695
chr19:42857094	GTEx	-35321
chr19:42858337	GTEx	-36564
chr19:42859381	GTEx	-37608
chr19:42859678	GTEx	-37905
chr19:42862513	GTEx	-40740
chr19:42865404	GTEx	-43631
chr19:42867885	GTEx	-46112
chr19:42868841	GTEx	-47068
chr19:42870320	GTEx	-48547
chr19:42870449	GTEx	-48676
chr19:42871718	GTEx	-49945
chr19:42879499	GTEx	-57726
chr19:42879528	GTEx/pshycENCODE	-57755
chr19:42879623	GTEx	-57850
chr19:42879649	GTEx	-57876
chr19:42880599	GTEx	-58826
chr19:42884428	GTEx	-62655
chr19:42885826	GTEx	-64053
chr19:42886393	GTEx	-64620
chr19:42894684	GTEx	-72911
chr19:42894981	GTEx	-73208
chr19:42895581	GTEx	-73808
chr19:42896384	GTEx	-74611
chr19:42897414	GTEx	-75641
chr19:42898669	GTEx	-76896
chr19:42898800	GTEx	-77027

chr19:42900144	GTEx	-78371
chr19:42902280	GTEx	-80507
chr19:42905883	GTEx	-84110
chr19:42908380	GTEx	-86607
chr19:42908666	GTEx	-86893
chr19:42909159	GTEx	-87386
chr19:42909612	GTEx	-87839
chr19:42909843	GTEx	-88070
chr19:42910207	GTEx	-88434
chr19:42915777	GTEx	-94004
chr19:42918541	GTEx	-96768
chr19:42919067	GTEx	-97294
chr19:42919069	GTEx	-97296
chr19:42919070	GTEx	-97297
chr19:42919236	GTEx	-97463
chr19:42919520	GTEx	-97747
chr19:42921903	GTEx	-100130
chr19:42922567	GTEx	-100794
chr19:42927011	GTEx	-105238
chr19:42928608	GTEx	-106835
chr19:42935017	GTEx	-113244
chr19:42935851	GTEx	-114078
chr19:42951436	GTEx	-129663
chr19:42956598	pshycENCODE	-134825
chr19:42957809	pshycENCODE	-136036
chr19:42958288	pshycENCODE	-136515
chr19:42958348	pshycENCODE	-136575
chr19:42959297	pshycENCODE	-137524
chr19:42960054	pshycENCODE	-138281
chr19:42960105	pshycENCODE	-138332
chr19:43002194	pshycENCODE	-180421
chr19:43007594	GTEx	-185821
chr19:43011320	GTEx	-189547
chr19:43011947	pshycENCODE	-190174
chr19:43043695	pshycENCODE	-221922
chr19:43044400	pshycENCODE	-222627
chr19:43044441	pshycENCODE	-222668
chr19:43044469	pshycENCODE	-222696
chr19:43044829	pshycENCODE	-223056
chr19:43048050	pshycENCODE	-226277
chr19:43049368	pshycENCODE	-227595
chr19:43054085	pshycENCODE	-232312
chr19:43061748	pshycENCODE	-239975
chr19:43062146	pshycENCODE	-240373

chr19:43063274	pshycENCODE	-241501
chr19:43064031	pshycENCODE	-242258
chr19:43064698	pshycENCODE	-242925
chr19:43065399	pshycENCODE	-243626
chr19:43067066	pshycENCODE	-245293
chr19:43067611	pshycENCODE	-245838
chr19:43067817	pshycENCODE	-246044
chr19:43068886	pshycENCODE	-247113
chr19:43070242	pshycENCODE	-248469
chr19:43071315	pshycENCODE	-249542
chr19:43072380	pshycENCODE	-250607
chr19:43074311	pshycENCODE	-252538
chr19:43074939	pshycENCODE	-253166
chr19:43077525	pshycENCODE	-255752
chr19:43079443	pshycENCODE	-257670
chr19:43079711	pshycENCODE	-257938
chr19:43100045	pshycENCODE	-278272
chr19:43100241	pshycENCODE	-278468
chr19:43101794	pshycENCODE	-280021
chr19:43102446	pshycENCODE	-280673
chr19:43102509	pshycENCODE	-280736
chr19:43104310	pshycENCODE	-282537
chr19:43104369	pshycENCODE	-282596
chr19:43105158	pshycENCODE	-283385
chr19:43105332	pshycENCODE	-283559
chr19:43105616	pshycENCODE	-283843
chr19:43105911	pshycENCODE	-284138
chr19:43108626	pshycENCODE	-286853
chr19:43108628	pshycENCODE	-286855
chr19:43108863	pshycENCODE	-287090
chr19:43109554	GTEx	-287781
chr19:43111141	pshycENCODE	-289368
chr19:43112318	pshycENCODE	-290545
chr19:43114943	pshycENCODE	-293170
chr19:43117170	pshycENCODE	-295397
chr19:43117665	pshycENCODE	-295892
chr19:43117760	pshycENCODE	-295987
chr19:43121359	pshycENCODE	-299586
chr19:43133855	GTEx	-312082
chr19:43161395	pshycENCODE	-339622
chr19:43161502	pshycENCODE	-339729
chr19:43167751	pshycENCODE	-345978
chr19:43177816	pshycENCODE	-356043
chr19:43180253	pshycENCODE	-358480

chr19:43180255	pshycENCODE	-358482
chr19:43180479	pshycENCODE	-358706
chr19:43183173	pshycENCODE	-361400
chr19:43260786	pshycENCODE	-439013
chr19:43265192	pshycENCODE	-443419
chr19:43266131	GTEx	-444358
chr19:43270604	pshycENCODE	-448831
chr19:43274289	pshycENCODE	-452516
chr19:43274698	pshycENCODE	-452925
chr19:43275671	pshycENCODE	-453898
chr19:43328661	GTEx	-506888
chr19:43667978	GTEx	-846205
chr19:43683354	GTEx	-861581

Supplementary Table 21. Genes in *PSG8-AS1* WGCNA module.

Genes are ordered by p-value of correlation.

ENSEMBL ID	GENE NAME	PEARSONCOR	PVALUE
ENSG00000197971	*MBP*	0.912060755	4.46E-203
ENSG00000184481	*FOXO4*	0.904772053	1.57E-194
ENSG00000129250	*KIF1C*	0.901444422	7.44E-191
ENSG00000227495	*KIF1C-AS1*	0.891445127	1.49E-180
ENSG00000078804	*TP53INP2*	0.876481841	6.89E-167
ENSG00000064787	*BCAS1*	0.868435507	2.95E-160
ENSG00000198121	*LPAR1*	0.857134868	1.20E-151
ENSG00000168314	*MOBP*	0.856654684	2.68E-151
ENSG00000144283	*PKP4*	0.845818494	9.70E-144
ENSG00000146063	*TRIM41*	0.837752881	1.75E-138
ENSG00000021300	*PLEKHB1*	0.826824188	8.40E-132
ENSG00000154945	*ANKRD40*	0.822518043	2.68E-129
ENSG00000229670	*PKP4P1*	0.81474794	6.03E-125
ENSG00000204278	*TMEM235*	0.813925686	1.69E-124
ENSG00000146122	*DAAM2*	0.813691294	2.27E-124
ENSG00000204644	*ZFP57*	0.807184662	6.65E-121
ENSG00000180739	*S1PR5*	0.806771042	1.09E-120
ENSG00000124374	*PAIP2B*	0.802391492	1.97E-118
ENSG00000176485	*PLAAT3*	0.784784555	6.55E-110
ENSG00000158315	*RHBDL2*	0.784381265	1.00E-109
ENSG00000247049	ENSG00000247049	0.775010529	1.61E-105
ENSG00000108387	*SEPTIN4*	0.774011929	4.41E-105
ENSG00000235033	*DAAM2-AS1*	0.77297418	1.25E-104
ENSG00000113578	*FGF1*	0.769139595	5.51E-103
ENSG00000203586	ENSG00000203586	0.767891078	1.86E-102
ENSG00000236239	ENSG00000236239	0.767837065	1.96E-102
ENSG00000184208	*C22orf46*	0.76340105	1.40E-100
ENSG00000242861	ENSG00000242861	0.762407871	3.58E-100
ENSG00000012171	*SEMA3B*	0.761930459	5.63E-100
ENSG00000204380	*PKP4-AS1*	0.761119369	1.21E-99
ENSG00000135773	*CAPN9*	0.757016246	5.51E-98
ENSG00000064393	*HIPK2*	0.755658911	1.92E-97
ENSG00000233708	ENSG00000233708	0.754314408	6.54E-97
ENSG00000182851	ENSG00000182851	0.753242275	1.73E-96
ENSG00000255953	ENSG00000255953	0.752161258	4.59E-96
ENSG00000176381	*PRR18*	0.749689565	4.19E-95
ENSG00000224795	*NTM-AS1*	0.749570809	4.66E-95
ENSG00000109846	*CRYAB*	0.749055291	7.36E-95
ENSG00000114993	*RTKN*	0.743271071	1.16E-92

ENSG00000225140	*LOC101930421*	0.74073737	1.02E-91
ENSG00000163113	*OTUD7B*	0.73761879	1.44E-90
ENSG00000148218	*ALAD*	0.737569174	1.50E-90
ENSG00000180354	*MTURN*	0.736188641	4.76E-90
ENSG00000108381	*ASPA*	0.736030906	5.43E-90
ENSG00000084453	*SLCO1A2*	0.733180808	5.77E-89
ENSG00000245293	*CYP2U1-AS1*	0.730076696	7.33E-88
ENSG00000236859	*NIFK-AS1*	0.728681068	2.27E-87
ENSG00000227544	*LINC03013*	0.726055212	1.87E-86
ENSG00000172005	*MAL*	0.724033436	9.32E-86
ENSG00000259357	ENSG00000259357	0.721900437	5.00E-85
ENSG00000233552	ENSG00000233552	0.720707788	1.27E-84
ENSG00000197892	*KIF13B*	0.72030224	1.74E-84
ENSG00000132702	*HAPLN2*	0.719732111	2.71E-84
ENSG00000047849	*MAP4*	0.718832277	5.45E-84
ENSG00000115468	*EFHD1*	0.718085882	9.70E-84
ENSG00000244274	*DBNDD2*	0.712075067	9.40E-82
ENSG00000135678	*CPM*	0.709773789	5.25E-81
ENSG00000255782	ENSG00000255782	0.708515496	1.33E-80
ENSG00000141524	*TMC6*	0.707147807	3.66E-80
ENSG00000049245	*VAMP3*	0.705466194	1.26E-79
ENSG00000198835	*GJC2*	0.704150888	3.27E-79
ENSG00000130958	*SLC35D2*	0.703884903	3.97E-79
ENSG00000166091	*CMTM5*	0.702710339	9.28E-79
ENSG00000224924	*LINC00320*	0.700743347	3.82E-78
ENSG00000101190	*TCFL5*	0.700470214	4.64E-78
ENSG00000257337	*TNS2-AS1*	0.698590982	1.77E-77
ENSG00000254806	*SYS1-DBNDD2*	0.697646407	3.46E-77
ENSG00000214780	ENSG00000214780	0.694240424	3.77E-76
ENSG00000180929	*GPR62*	0.693984582	4.51E-76
ENSG00000129538	*RNASE1*	0.692082295	1.69E-75
ENSG00000136541	*ERMN*	0.689725658	8.53E-75
ENSG00000065361	*ERBB3*	0.689653545	8.96E-75
ENSG00000105707	*HPN*	0.688457919	2.03E-74
ENSG00000238018	*SPTBN1-AS2*	0.688345268	2.19E-74
ENSG00000204282	*TNRC6C-AS1*	0.688206991	2.40E-74
ENSG00000173786	*CNP*	0.687837051	3.09E-74
ENSG00000183111	*ARHGEF37*	0.684362356	3.22E-73
ENSG00000149633	*KIAA1755*	0.684305794	3.34E-73
ENSG00000198585	*NUDT16*	0.684301786	3.35E-73
ENSG00000139433	*GLTP*	0.682396681	1.19E-72
ENSG00000211584	*SLC48A1*	0.681018859	2.97E-72
ENSG00000117115	*PADI2*	0.680829288	3.37E-72
ENSG00000205696	*ADARB2-AS1*	0.680775618	3.49E-72

ENSG00000102934	*PLLP*	0.680301906	4.77E-72
ENSG00000164627	*KIF6*	0.680203307	5.09E-72
ENSG00000189058	*APOD*	0.678453044	1.61E-71
ENSG00000124251	*TP53TG5*	0.677277433	3.46E-71
ENSG00000168754	*FAM178B*	0.676270331	6.65E-71
ENSG00000066322	*ELOVL1*	0.675324091	1.23E-70
ENSG00000073417	*PDE8A*	0.674800651	1.72E-70
ENSG00000123560	*PLP1*	0.671996282	1.04E-69
ENSG00000162444	*RBP7*	0.671287838	1.63E-69
ENSG00000141540	*TTYH2*	0.670554993	2.59E-69
ENSG00000184343	*SRPK3*	0.670381632	2.89E-69
ENSG00000204655	*MOG*	0.66334314	2.35E-67
ENSG00000224189	*HAGLR*	0.662299953	4.47E-67
ENSG00000251660	ENSG00000251660	0.662105535	5.03E-67
ENSG00000135525	*MAP7*	0.661139193	9.09E-67
ENSG00000179178	*TMEM125*	0.659709996	2.17E-66
ENSG00000122367	*LDB3*	0.657203876	9.89E-66
ENSG00000178722	*C5orf64*	0.653796488	7.59E-65
ENSG00000169562	*GJB1*	0.653507015	9.01E-65
ENSG00000139629	*GALNT6*	0.653486565	9.12E-65
ENSG00000117266	*CDK18*	0.653400413	9.60E-65
ENSG00000013297	*CLDN11*	0.653100719	1.15E-64
ENSG00000109099	*PMP22*	0.652093254	2.08E-64
ENSG00000177105	*RHOG*	0.650863837	4.29E-64
ENSG00000204323	*SMIM5*	0.650522198	5.25E-64
ENSG00000135929	*CYP27A1*	0.647290157	3.46E-63
ENSG00000152689	*RASGRP3*	0.647158887	3.73E-63
ENSG00000154493	*C10orf90*	0.64580976	8.13E-63
ENSG00000035664	*DAPK2*	0.645280249	1.10E-62
ENSG00000169247	*SH3TC2*	0.644870016	1.40E-62
ENSG00000167191	*GPRC5B*	0.641808529	8.03E-62
ENSG00000183196	*CHST6*	0.641363295	1.03E-61
ENSG00000115935	*WIPF1*	0.641330578	1.05E-61
ENSG00000170271	*FAXDC2*	0.639614851	2.78E-61
ENSG00000103089	*FA2H*	0.639362824	3.20E-61
ENSG00000187764	*SEMA4D*	0.638534344	5.10E-61
ENSG00000150867	*PIP4K2A*	0.637224431	1.06E-60
ENSG00000204161	*TMEM273*	0.637204695	1.07E-60
ENSG00000070759	*TESK2*	0.637014494	1.19E-60
ENSG00000080822	*CLDND1*	0.636631363	1.48E-60
ENSG00000148180	*GSN*	0.635780287	2.37E-60
ENSG00000232504	*ST3GAL5-AS1*	0.634758639	4.18E-60
ENSG00000259070	*LINC00639*	0.632983107	1.11E-59
ENSG00000187775	*DNAH17*	0.632746352	1.27E-59

ENSG00000100181	*TPTEP1*	0.632539444	1.42E-59
ENSG00000161896	*IP6K3*	0.632095028	1.81E-59
ENSG00000105695	*MAG*	0.627864287	1.81E-58
ENSG00000167755	*KLK6*	0.627259767	2.50E-58
ENSG00000060237	*WNK1*	0.627005057	2.87E-58
ENSG00000101049	*SGK2*	0.626826961	3.16E-58
ENSG00000091513	*TF*	0.626755499	3.28E-58
ENSG00000205116	*TMEM88B*	0.6239659	1.46E-57
ENSG00000086205	*FOLH1*	0.623590727	1.78E-57
ENSG00000221817	*PPP3CB-AS1*	0.623482913	1.89E-57
ENSG00000126822	*PLEKHG3*	0.622677642	2.89E-57
ENSG00000172508	*CARNS1*	0.619710806	1.38E-56
ENSG00000128283	*CDC42EP1*	0.619663617	1.41E-56
ENSG00000229645	*C14orf139*	0.618456545	2.65E-56
ENSG00000234198	ENSG00000234198	0.616540883	7.18E-56
ENSG00000237943	*PRKCQ-AS1*	0.616250528	8.35E-56
ENSG00000136867	*SLC31A2*	0.614862022	1.71E-55
ENSG00000012124	*CD22*	0.614814813	1.75E-55
ENSG00000198753	*PLXNB3*	0.614268725	2.32E-55
ENSG00000120729	*MYOT*	0.613643078	3.19E-55
ENSG00000135720	*DYNC1LI2*	0.612830105	4.84E-55
ENSG00000164023	*SGMS2*	0.610238129	1.81E-54
ENSG00000080819	*CPOX*	0.606869863	9.86E-54
ENSG00000092529	*CAPN3*	0.605115034	2.37E-53
ENSG00000160223	*ICOSLG*	0.603650679	4.89E-53
ENSG00000049089	*COL9A2*	0.60337677	5.60E-53
ENSG00000124920	*MYRF*	0.602717404	7.75E-53
ENSG00000099282	*TSPAN15*	0.602488821	8.67E-53
ENSG00000054690	*PLEKHH1*	0.602371898	9.19E-53
ENSG00000143418	*CERS2*	0.602348013	9.30E-53
ENSG00000254081	*LINC01299*	0.602294896	9.54E-53
ENSG00000133256	*PDE6B*	0.601972205	1.12E-52
ENSG00000115993	*TRAK2*	0.60134862	1.52E-52
ENSG00000187098	*MITF*	0.600380416	2.44E-52
ENSG00000254129	ENSG00000254129	0.599144692	4.46E-52
ENSG00000258461	ENSG00000258461	0.598994165	4.80E-52
ENSG00000238194	*PHACTR3-AS1*	0.598663338	5.64E-52
ENSG00000179630	*LACC1*	0.59839709	6.41E-52
ENSG00000158715	*SLC45A3*	0.598025293	7.68E-52
ENSG00000171840	*NINJ2*	0.597219451	1.13E-51
ENSG00000146859	*TMEM140*	0.596878461	1.34E-51
ENSG00000235049	*LINC00940*	0.596336582	1.74E-51
ENSG00000186417	*GLDN*	0.594445417	4.31E-51
ENSG00000158859	*ADAMTS4*	0.59408792	5.11E-51

ENSG00000196187	*TMEM63A*	0.593688547	6.19E-51
ENSG00000184007	*PTP4A2*	0.592588459	1.05E-50
ENSG00000119121	*TRPM6*	0.591701057	1.59E-50
ENSG00000162949	*CAPN13*	0.589832868	3.85E-50
ENSG00000176658	*MYO1D*	0.587708363	1.04E-49
ENSG00000104205	*SGK3*	0.587384489	1.22E-49
ENSG00000158352	*SHROOM4*	0.586689832	1.68E-49
ENSG00000151552	*QDPR*	0.586539979	1.80E-49
ENSG00000198865	*CCDC152*	0.586517561	1.82E-49
ENSG00000237133	ENSG00000237133	0.586250534	2.06E-49
ENSG00000137221	*TJAP1*	0.585713484	2.65E-49
ENSG00000242808	*SOX2-OT*	0.582897874	9.72E-49
ENSG00000158865	*SLC5A11*	0.579288298	5.06E-48
ENSG00000235865	*GSN-AS1*	0.578988034	5.80E-48
ENSG00000126803	*HSPA2*	0.578338785	7.79E-48
ENSG00000070214	*SLC44A1*	0.577819406	9.85E-48
ENSG00000141934	*PLPP2*	0.577686618	1.05E-47
ENSG00000244062	ENSG00000244062	0.577066904	1.38E-47
ENSG00000079435	*LIPE*	0.576224923	2.02E-47
ENSG00000006611	*USH1C*	0.57270227	9.75E-47
ENSG00000135362	*PRR5L*	0.572523088	1.06E-46
ENSG00000147459	*DOCK5*	0.572433328	1.10E-46
ENSG00000172548	*NIPAL4*	0.571957972	1.36E-46
ENSG00000231840	*TMEM139-AS1*	0.570637435	2.43E-46
ENSG00000089041	*P2RX7*	0.570327371	2.79E-46
ENSG00000141458	*NPC1*	0.570184728	2.97E-46
ENSG00000241163	*LINC00877*	0.569189419	4.59E-46
ENSG00000173253	*DMRT2*	0.567928576	7.98E-46
ENSG00000166348	*USP54*	0.566553792	1.45E-45
ENSG00000166387	*PPFIBP2*	0.565656802	2.14E-45
ENSG00000119139	*TJP2*	0.565462751	2.33E-45
ENSG00000250722	*SELENOP*	0.564505665	3.53E-45
ENSG00000006831	*ADIPOR2*	0.564413317	3.67E-45
ENSG00000254481	*PTP4A2P2*	0.563244924	6.07E-45
ENSG00000236882	*LINC01554*	0.563011558	6.71E-45
ENSG00000107331	*ABCA2*	0.562573106	8.10E-45
ENSG00000219392	*ZNF602P*	0.561699426	1.18E-44
ENSG00000250198	*LINC02199*	0.56086882	1.68E-44
ENSG00000126860	*EVI2A*	0.560738214	1.78E-44
ENSG00000224002	ENSG00000224002	0.559761654	2.69E-44
ENSG00000255138	*GLTPP1*	0.558729404	4.17E-44
ENSG00000255470	ENSG00000255470	0.558685289	4.25E-44
ENSG00000164124	*TMEM144*	0.557904862	5.91E-44
ENSG00000164303	*ENPP6*	0.557700514	6.44E-44

ENSG00000010282	*HHATL*	0.555955693	1.34E-43
ENSG00000259172	*SNRPA1-DT*	0.555336398	1.74E-43
ENSG00000178826	*TMEM139*	0.554742905	2.23E-43
ENSG00000185896	*LAMP1*	0.553928055	3.13E-43
ENSG00000010310	*GIPR*	0.553878378	3.20E-43
ENSG00000019144	*PHLDB1*	0.553081415	4.45E-43
ENSG00000255701	ENSG00000255701	0.5518475	7.43E-43
ENSG00000075651	*PLD1*	0.549350279	2.08E-42
ENSG00000174607	*UGT8*	0.548718668	2.69E-42
ENSG00000131386	*GALNT15*	0.547483306	4.46E-42
ENSG00000011426	*ANLN*	0.545161656	1.14E-41
ENSG00000237798	ENSG00000237798	0.544163041	1.71E-41
ENSG00000150656	*CNDP1*	0.544068549	1.78E-41
ENSG00000136960	*ENPP2*	0.543686727	2.07E-41
ENSG00000244491	ENSG00000244491	0.542456933	3.40E-41
ENSG00000126653	*NSRP1*	0.541867226	4.30E-41
ENSG00000111254	*AKAP3*	0.540741338	6.74E-41
ENSG00000233838	*DPH3P1*	0.538113215	1.91E-40
ENSG00000119280	*C1orf198*	0.536566875	3.52E-40
ENSG00000235189	*LOC105373335*	0.536067541	4.28E-40
ENSG00000140479	*PCSK6*	0.535658367	5.02E-40
ENSG00000167123	*CERCAM*	0.535090749	6.27E-40
ENSG00000255807	*PTP4A1P2*	0.531528217	2.51E-39
ENSG00000242150	ENSG00000242150	0.531282179	2.75E-39
ENSG00000141338	*ABCA8*	0.527858383	1.03E-38
ENSG00000119321	*FKBP15*	0.527688985	1.09E-38
ENSG00000106397	*PLOD3*	0.525548644	2.47E-38
ENSG00000242337	*INHCAP*	0.524286385	3.98E-38
ENSG00000259351	ENSG00000259351	0.520935331	1.40E-37
ENSG00000222466	ENSG00000222466	0.519192544	2.67E-37
ENSG00000176438	*SYNE3*	0.5139524	1.83E-36
ENSG00000072310	*SREBF1*	0.511166676	5.04E-36
ENSG00000100225	*FBXO7*	0.510854543	5.64E-36
ENSG00000185418	*TARS3*	0.51046378	6.49E-36
ENSG00000170537	*TMC7*	0.506524586	2.66E-35
ENSG00000168297	*PXK*	0.50592166	3.29E-35
ENSG00000237782	*PLLPP1*	0.504928988	4.68E-35
ENSG00000116574	*RHOU*	0.50335927	8.14E-35
ENSG00000130962	*PRRG1*	0.502899103	9.57E-35
ENSG00000149218	*ENDOD1*	0.501997914	1.31E-34
ENSG00000184144	*CNTN2*	0.500842361	1.97E-34
ENSG00000082996	*RNF13*	0.499879112	2.75E-34
ENSG00000258761	ENSG00000258761	0.49937149	3.28E-34
ENSG00000160862	*AZGP1*	0.497653062	5.95E-34

ENSG00000154262	*ABCA6*	0.496547543	8.71E-34
ENSG00000107798	*LIPA*	0.491849501	4.33E-33
ENSG00000225279	ENSG00000225279	0.491358879	5.11E-33
ENSG00000135636	*DYSF*	0.490226539	7.49E-33
ENSG00000103184	*SEC14L5*	0.489948902	8.22E-33
ENSG00000250159	ENSG00000250159	0.489102152	1.09E-32
ENSG00000235903	*CPB2-AS1*	0.489070598	1.10E-32
ENSG00000139292	*LGR5*	0.486525821	2.59E-32
ENSG00000099219	*ERMP1*	0.478610237	3.48E-31
ENSG00000078269	*SYNJ2*	0.478011873	4.23E-31
ENSG00000259604	ENSG00000259604	0.476705448	6.45E-31
ENSG00000132854	*KANK4*	0.475486141	9.55E-31
ENSG00000183230	*CTNNA3*	0.474017571	1.53E-30
ENSG00000164161	*HHIP*	0.471844153	3.05E-30
ENSG00000183549	*ACSM5*	0.470210336	5.11E-30
ENSG00000133104	*SPART*	0.467783085	1.10E-29
ENSG00000253965	ENSG00000253965	0.467206234	1.31E-29
ENSG00000135905	*DOCK10*	0.467200264	1.32E-29
ENSG00000064651	*SLC12A2*	0.46443405	3.11E-29
ENSG00000169242	*EFNA1*	0.463324429	4.38E-29
ENSG00000158296	*SLC13A3*	0.463197909	4.55E-29
ENSG00000117408	*IPO13*	0.461254876	8.27E-29
ENSG00000105889	*STEAP1B*	0.45769063	2.45E-28
ENSG00000143409	*MINDY1*	0.456717164	3.28E-28
ENSG00000189067	*LITAF*	0.456095239	3.96E-28
ENSG00000132321	*IQCA1*	0.450104349	2.36E-27
ENSG00000021826	*CPS1*	0.446813827	6.21E-27
ENSG00000155792	*DEPTOR*	0.446543235	6.72E-27
ENSG00000118322	*ATP10B*	0.445766814	8.42E-27
ENSG00000133313	*CNDP2*	0.445712925	8.56E-27
ENSG00000257761	ENSG00000257761	0.444235035	1.31E-26
ENSG00000005893	*LAMP2*	0.443922303	1.44E-26
ENSG00000165410	*CFL2*	0.438278878	7.24E-26
ENSG00000119537	*KDSR*	0.4359446	1.40E-25
ENSG00000257283	ENSG00000257283	0.435574088	1.56E-25
ENSG00000170775	*GPR37*	0.435235103	1.71E-25
ENSG00000118298	*CA14*	0.428229455	1.20E-24
ENSG00000249121	ENSG00000249121	0.427805087	1.34E-24
ENSG00000186056	*MATN1-AS1*	0.426631999	1.85E-24
ENSG00000186340	*THBS2*	0.426102494	2.14E-24
ENSG00000175785	*PRIMA1*	0.422818415	5.21E-24
ENSG00000138316	*ADAMTS14*	0.422711307	5.37E-24
ENSG00000125703	*ATG4C*	0.422469566	5.73E-24
ENSG00000057704	*TMCC3*	0.420191752	1.05E-23

ENSG00000130150	*MOSPD2*	0.419733226	1.19E-23
ENSG00000112276	*BVES*	0.419342974	1.32E-23
ENSG00000175764	*TTLL11*	0.419010938	1.45E-23
ENSG00000157483	*MYO1E*	0.416045041	3.17E-23
ENSG00000196517	*SLC6A9*	0.410495677	1.35E-22
ENSG00000100596	*SPTLC2*	0.408452878	2.28E-22
ENSG00000178425	*NT5DC1*	0.406235899	4.03E-22
ENSG00000158079	*PTPDC1*	0.404503394	6.26E-22
ENSG00000250615	ENSG00000250615	0.404202565	6.75E-22
ENSG00000232623	*MRAP-AS1*	0.399075935	2.45E-21
ENSG00000178401	*DNAJC22*	0.398454204	2.85E-21
ENSG00000080200	*CRYBG3*	0.397181671	3.91E-21
ENSG00000160179	*ABCG1*	0.392430688	1.25E-20
ENSG00000152910	*CNTNAP4*	0.390732479	1.89E-20
ENSG00000081479	*LRP2*	0.390154221	2.18E-20
ENSG00000179363	*TMEM31*	0.388577639	3.18E-20
ENSG00000242689	*CNTF*	0.386026932	5.85E-20
ENSG00000242010	ENSG00000242010	0.383700428	1.02E-19
ENSG00000220739	ENSG00000220739	0.376728903	5.16E-19
ENSG00000088970	*KIZ*	0.374374953	8.87E-19
ENSG00000249379	ENSG00000249379	0.374103748	9.43E-19
ENSG00000112893	*MAN2A1*	0.372568518	1.34E-18
ENSG00000242068	ENSG00000242068	0.371092809	1.87E-18
ENSG00000224101	*ELMO1-AS1*	0.36875905	3.16E-18
ENSG00000234603	ENSG00000234603	0.365310932	6.81E-18
ENSG00000103381	*CPPED1*	0.363354176	1.05E-17
ENSG00000154642	*C21orf91*	0.362735069	1.20E-17
ENSG00000174137	*FAM53A*	0.359208599	2.59E-17
ENSG00000226496	*LINC00323*	0.358023827	3.35E-17
ENSG00000248706	ENSG00000248706	0.354063049	7.84E-17
ENSG00000151914	*DST*	0.351929862	1.23E-16
ENSG00000158516	*CPA2*	0.349664741	1.98E-16
ENSG00000154864	*PIEZO2*	0.34620573	4.08E-16
ENSG00000039319	*ZFYVE16*	0.340629876	1.28E-15
ENSG00000068650	*ATP11A*	0.33829354	2.05E-15
ENSG00000233878	ENSG00000233878	0.336143772	3.15E-15
ENSG00000123700	*KCNJ2*	0.324751389	2.92E-14
ENSG00000144815	*NXPE3*	0.319522336	7.87E-14
ENSG00000167257	*RNF214*	0.314626652	1.96E-13
ENSG00000164176	*EDIL3*	0.312933806	2.67E-13
ENSG00000245330	*DEPTOR-AS1*	0.311808439	3.28E-13
ENSG00000214961	*NT5DC1P2*	0.311783352	3.29E-13
ENSG00000236326	ENSG00000236326	0.307829183	6.74E-13
ENSG00000186318	*BACE1*	0.306487369	8.57E-13

ENSG00000244676	ENSG00000244676	0.301063315	2.24E-12
ENSG00000184292	*TACSTD2*	0.298878815	3.27E-12
ENSG00000159592	*GPBP1L1*	0.298802693	3.32E-12
ENSG00000189134	*NKAPL*	0.293818709	7.81E-12
ENSG00000113396	*SLC27A6*	0.288187007	2.01E-11
ENSG00000124275	*MTRR*	0.285127805	3.34E-11
ENSG00000121690	*DEPDC7*	0.27529467	1.63E-10
ENSG00000125954	*CHURC1-FNTB*	0.268873816	4.44E-10
ENSG00000246515	ENSG00000246515	0.267467976	5.51E-10
ENSG00000166923	*GREM1*	0.266897729	6.01E-10
ENSG00000198690	*FAN1*	0.255534967	3.27E-09
ENSG00000253921	*ATOX1-AS1*	0.251683966	5.71E-09
ENSG00000257365	*FNTB*	0.226928397	1.64E-07
ENSG00000103978	*TMEM87A*	0.192318402	9.85E-06
ENSG00000237381	*NT5DC1P1*	0.192280641	9.89E-06
ENSG00000226752	*CUTALP*	0.136577607	0.001780423
ENSG00000228604	ENSG00000228604	0.130065939	0.002936354

Supplementary Table 22. Gene set enrichment analysis performed with DAVID online tool of the *PSG8-AS1* high-correlated gene expression module.

Showing FDR < 0.2 results.

TERM	PVALUE	BONFERRONI	BENJAMINI	FDR
GO:0016021~integral component of membrane	1.13E-07	2.78E-05	2.78E-05	2.64E-05
GO:0019911~structural constituent of myelin sheath	6.17E-06	2.57E-03	2.57E-03	2.55E-03
R-HSA-1660661:R-HSA-1660661	1.90E-04	3.43E-02	3.49E-02	3.49E-02
hsa00565:Ether lipid metabolism	4.74E-04	5.80E-02	3.66E-02	3.66E-02
hsa00600:Sphingolipid metabolism	5.81E-04	7.06E-02	3.66E-02	3.66E-02
GO:0008233~peptidase activity	2.05E-04	8.18E-02	4.27E-02	4.23E-02
GO:0005886~plasma membrane	5.97E-04	1.37E-01	5.35E-02	5.09E-02
GO:0005768~endosome	9.92E-04	2.17E-01	5.35E-02	5.09E-02
GO:0043218~compact myelin	1.07E-03	2.31E-01	5.35E-02	5.09E-02
GO:0043209~myelin sheath	1.09E-03	2.35E-01	5.35E-02	5.09E-02
GO:0051015~actin filament binding	4.05E-04	1.55E-01	5.63E-02	5.59E-02
GO:0016324~apical plasma membrane	2.08E-03	4.00E-01	8.02E-02	7.63E-02
GO:0005887~integral component of plasma membrane	2.78E-03	4.95E-01	8.02E-02	7.63E-02
GO:0015629~actin cytoskeleton	2.90E-03	5.10E-01	8.02E-02	7.63E-02
GO:0009986~cell surface	2.93E-03	5.15E-01	8.02E-02	7.63E-02
GO:0055037~recycling endosome	3.79E-03	6.07E-01	8.60E-02	8.18E-02
GO:0070062~extracellular exosome	3.98E-03	6.25E-01	8.60E-02	8.18E-02
GO:0005765~lysosomal membrane	4.19E-03	6.44E-01	8.60E-02	8.18E-02
GO:0016887~ATPase activity	8.19E-04	2.89E-01	8.54E-02	8.47E-02
121.OligoDiff	5.08E-02	1.88E-01	1.02E-01	1.02E-01
98.Dietary_Fat_and_Gene_Transcription	5.08E-02	1.88E-01	1.02E-01	1.02E-01
GO:0030139~endocytic vesicle	6.04E-03	7.75E-01	1.14E-01	1.09E-01
GO:0004180~carboxypeptidase activity	1.40E-03	4.43E-01	1.17E-01	1.16E-01
GO:0009925~basal plasma membrane	7.54E-03	8.45E-01	1.32E-01	1.26E-01
GO:0008360~regulation of cell shape	1.22E-04	1.28E-01	1.36E-01	1.36E-01
GO:0042552~myelination	3.77E-04	3.45E-01	1.42E-01	1.41E-01
GO:0030148~sphingolipid biosynthetic process	4.61E-04	4.03E-01	1.42E-01	1.41E-01
GO:0055085~transmembrane transport	5.05E-04	4.32E-01	1.42E-01	1.41E-01
GO:0005875~microtubule associated complex	1.07E-02	9.29E-01	1.68E-01	1.60E-01
GO:0033270~paranode region of axon	1.10E-02	9.33E-01	1.68E-01	1.60E-01
hsa00340:Histidine metabolism	3.86E-03	3.86E-01	1.62E-01	1.62E-01

Supplementary Table 23. Top 300 *PSG8-AS1* protein binding predictions by catRapiD v2.1.

Proteins are identified by the UNIPROT accession followed by the protein name and ordered by IPS. In most cases, the *PSG8-AS1* nucleotides 704-779 harbor the strongest prediction signal.

PROTEIN_ID	TOP_RNA_FRAGMENT	INTERACTION_PROPENSITY	Z_SCORE
Q9UGP8_SEC63	704-779	120.13	5.65
O14776_TCERG1	704-779	118.29	5.55
Q9H501_ESF1	704-779	117.46	5.5
P54253_ATXN1	704-779	116.71	5.46
O60293_ZFC3H1	704-779	116.52	5.45
Q5T7N2_L1TD1	704-779	115.99	5.42
O43719_HTATSF1	704-779	115.96	5.41
Q9BWU0_SLC4A1AP	704-779	115.68	5.4
Q9UIF8_BAZ2B	704-779	115.63	5.39
Q15022_SUZ12	704-779	115.61	5.39
O43290_SART1	704-779	113.73	5.29
Q92900_UPF1	704-779	113.55	5.28
P54277_PMS1	704-779	113.4	5.27
P49589_CARS1	704-779	112.73	5.23
Q04637_EIF4G1	704-779	112.25	5.2
Q13206_DDX10	704-779	111.54	5.16
Q9BXT4_TDRD1	704-779	111.05	5.13
Q96KR1_ZFR	704-779	110.96	5.13
P46459_NSF	704-779	110.92	5.12
Q6PKG0_LARP1	704-779	109.82	5.06
P35221_CTNNA1	704-779	109.24	5.03
Q9BUQ8_DDX23	704-779	109.08	5.02
O94804_STK10	704-779	108.65	4.99
Q14692_BMS1	704-779	108.64	4.99
Q9HCS7_XAB2	704-779	107.86	4.95
Q92888_ARHGEF1	704-779	107.82	4.95
Q9NRA8_EIF4ENIF1	704-779	107.49	4.93
P49792_RANBP2	704-779	107.07	4.9
P36776_LONP1	704-779	106.94	4.89
Q7KZ85_SUPT6H	704-779	106.91	4.89
O75643_SNRNP200	704-779	106.85	4.89
Q8TEX9_IPO4	704-779	106.68	4.88
Q9HCE1_MOV10	704-779	106	4.84
Q8IY21_DDX60	704-779	104.14	4.73
Q8IX12_CCAR1	704-779	103.28	4.68

Q8WXH0_SYNE2	704-779	102.83	4.66
Q7L014_DDX46	704-779	102.34	4.63
Q86US8_SMG6	704-779	102	4.61
P26639_TARS1	704-779	101.92	4.61
Q08211_DHX9	704-779	101.17	4.56
Q99459_CDC5L	704-779	100.6	4.53
Q96HW7_INTS4	704-779	99.45	4.46
P26640_VARS1	704-779	99.37	4.46
Q14980_NUMA1	704-779	99.28	4.45
P09327_VIL1	704-779	99.1	4.44
O60522_TDRD6	704-779	98.98	4.44
Q9ULW0_TPX2	704-779	98.9	4.43
Q8N1G2_CMTR1	704-779	98.84	4.43
Q9BYX4_IFIH1	704-779	98.74	4.42
Q9Y2W1_THRAP3	704-779	98.68	4.42
Q05682_CALD1	704-779	98.65	4.42
O75534_CSDE1	704-779	98.48	4.41
Q15061_WDR43	704-779	98.39	4.4
Q15393_SF3B3	704-779	98.39	4.4
P16383_GCFC2	704-779	98.3	4.4
O60287_URB1	704-779	98.17	4.39
O75116_ROCK2	704-779	97.95	4.38
Q9P2M7_CGN	704-779	97.81	4.37
Q8WYQ5_DGCR8	704-779	97.61	4.36
Q96T21_SECISBP2	704-779	97.57	4.36
Q7KZF4_SND1	704-779	97.17	4.33
P43243_MATR3	704-779	97.04	4.33
P14625_HSP90B1	704-779	97.01	4.32
Q92834_RPGR	704-779	96.51	4.29
Q9Y5B9_SUPT16H	704-779	96.42	4.29
Q92974_ARHGEF2	704-779	96.31	4.28
P46821_MAP1B	704-779	96.27	4.28
Q96Q15_SMG1	704-779	96.22	4.28
Q86UP2_KTN1	704-779	96.2	4.28
Q00341_HDLBP	704-779	96.19	4.28
P53618_COPB1	704-779	96.14	4.27
Q32MZ4_LRRFIP1	704-779	96.08	4.27
P41252_IARS1	704-779	95.82	4.25
Q6P9B9_INTS5	704-779	95.71	4.25
Q99590_SCAF11	704-779	95.65	4.25
Q92620_DHX38	704-779	95.47	4.24
Q8IYL2_TRMT44	704-779	95.33	4.23
O60763_USO1	704-779	95.27	4.22
Q14152_EIF3A	704-779	95.06	4.21

Q6Y7W6_GIGYF2	704-779	94.89	4.2
Q69YN4_VIRMA	704-779	94.79	4.2
Q13439_GOLGA4	704-779	94.64	4.19
O14787_TNPO2	704-779	94.4	4.17
Q01082_SPTBN1	704-779	94.24	4.17
Q9Y6Y8_SEC23IP	704-779	93.95	4.15
Q14241_ELOA	704-779	93.83	4.14
Q7Z2Z2_EFL1	704-779	93.77	4.14
Q14789_GOLGB1	704-779	93.7	4.13
Q8NDT2_RBM15B	704-779	93.65	4.13
P57678_GEMIN4	704-779	93.57	4.13
P21127_CDK11B	704-779	93.52	4.12
Q9UK61_TASOR	704-779	93.42	4.12
P05023_ATP1A1	704-779	93.12	4.1
O15397_IPO8	704-779	92.83	4.08
O95373_IPO7	704-779	92.76	4.08
Q9H2U1_DHX36	704-779	92.76	4.08
Q92797_SYMPK	704-779	92.67	4.07
O75400_PRPF40A	704-779	92.66	4.07
Q9Y2L1_DIS3	704-779	92.49	4.06
Q9Y678_COPG1	704-779	92.46	4.06
Q9UBB9_TFIP11	704-779	92.39	4.06
Q7L7V1_DHX32	704-779	92.3	4.05
Q14697_GANAB	704-779	92.25	4.05
Q9UIG0_BAZ1B	704-779	92.22	4.05
Q8NCM8_DYNC2H1	704-779	92.17	4.05
Q8WUM4_PDCD6IP	704-779	92.17	4.05
Q9UJC3_HOOK1	704-779	92.14	4.04
O43592_XPOT	704-779	92	4.04
Q6ZU64_CFAP65	704-779	91.96	4.03
P24928_POLR2A	704-779	91.87	4.03
P09874_PARP1	704-779	91.49	4.01
Q7Z406_MYH14	704-779	91.37	4
Q6IMN6_CAPRIN2	704-779	91.17	3.99
P78316_NOP14	704-779	91.15	3.99
Q92614_MYO18A	704-779	90.9	3.97
Q9NRR4_DROSHA	704-779	90.8	3.97
P42704_LRPPRC	704-779	90.72	3.96
P55196_AFDN	704-779	90.59	3.96
Q05707_COL14A1	704-779	90.24	3.94
Q9Y3T9_NOC2L	704-779	90.2	3.93
O43707_ACTN4	704-779	90.04	3.92
Q9HCG8_CWC22	704-779	90.01	3.92

O15042_U2SURP	704-779	89.95	3.92
P46087_NOP2	704-779	89.89	3.91
Q05193_DNM1	704-779	89.7	3.9
Q58A45_PAN3	704-779	89.63	3.9
Q9C0B0_UNK	704-779	89.6	3.9
Q86XP3_DDX42	704-779	89.51	3.89
Q14151_SAFB2	704-779	89.44	3.89
Q9C0C9_UBE2O	704-779	89.26	3.88
Q92616_GCN1	704-779	89.24	3.88
Q9BZJ0_CRNKL1	704-779	89.16	3.87
O00566_MPHOSPH10	704-779	89.12	3.87
Q8N163_CCAR2	704-779	89	3.86
Q5T5J6_SWT1	704-779	88.98	3.86
P07814_EPRS1	704-779	88.97	3.86
Q14690_PDCD11	704-779	88.88	3.86
Q9C0E2_XPO4	704-779	88.81	3.85
Q86VP6_CAND1	704-779	88.8	3.85
Q9BX66_SORBS1	704-779	88.6	3.84
Q9UII4_HERC5	704-779	88.56	3.84
Q9P2E3_ZNFX1	704-779	88.54	3.84
Q86UR5_RIMS1	704-779	88.49	3.83
Q9Y520_PRRC2C	704-779	88.38	3.83
Q13144_EIF2B5	704-779	88.34	3.83
Q9NZJ5_EIF2AK3	704-779	88.27	3.82
O00410_IPO5	704-779	88.14	3.81
Q14974_KPNB1	704-779	88.02	3.81
Q9H583_HEATR1	704-779	87.96	3.8
Q9UI17_DMGDH	704-779	87.92	3.8
P54886_ALDH18A1	704-779	87.88	3.8
Q7Z5L2_R3HCC1L	704-779	87.69	3.79
P30414_NKTR	704-779	87.66	3.79
P55072_VCP	704-779	87.45	3.77
P17480_UBTF	704-779	87.43	3.77
Q9H0D6_XRN2	704-779	87.42	3.77
P78344_EIF4G2	704-779	87.39	3.77
Q5JTZ9_AARS2	704-779	87.37	3.77
Q8WUQ7_CACTIN	704-779	87.31	3.77
Q15020_SART3	704-779	87.21	3.76
Q7Z3Z4_PIWIL4	704-779	87.12	3.75
Q9NQC3_RTN4	704-779	87.02	3.75
Q9UKV3_ACIN1	704-779	87	3.75
Q969S9_GFM2	704-779	86.89	3.74
O75460_ERN1	704-779	86.88	3.74

Q8IXT5_RBM12B	704-779	86.84	3.74
Q96EK7_FAM120B	704-779	86.81	3.74
O43491_EPB41L2	704-779	86.75	3.73
Q7Z333_SETX	704-779	86.73	3.73
P78527_PRKDC	704-779	86.66	3.73
P78332_RBM6	704-779	86.65	3.73
Q9NR97_TLR8	704-779	86.6	3.73
Q15459_SF3A1	704-779	86.57	3.72
Q96J94_PIWIL1	704-779	86.51	3.72
Q8TDD1_DDX54	704-779	86.5	3.72
Q9BQ52_ELAC2	704-779	86.32	3.71
Q5QJE6_DNTTIP2	704-779	86.19	3.7
Q13435_SF3B2	704-779	86.09	3.7
P21333_FLNA	704-779	85.9	3.68
O75152_ZC3H11A	704-779	85.89	3.68
Q9H9A5_CNOT10	704-779	85.76	3.68
Q9Y6K5_OAS3	704-779	85.71	3.67
Q9P2E9_RRBP1	704-779	85.65	3.67
P11498_PC	704-779	85.56	3.66
Q9BRZ2_TRIM56	704-779	85.55	3.66
Q03701_CEBPZ	704-779	85.47	3.66
Q8N5C6_SRBD1	704-779	85.41	3.66
Q8TC59_PIWIL2	704-779	85.25	3.65
Q9P2J5_LARS1	704-779	85.25	3.65
Q05823_RNASEL	704-779	85.23	3.65
Q9GZR7_DDX24	704-779	85.18	3.64
Q8NAT2_TDRD5	704-779	85.06	3.64
Q9NYK1_TLR7	704-779	84.98	3.63
Q13395_TARBP1	704-779	84.83	3.62
Q8IX01_SUGP2	704-779	84.81	3.62
P15924_DSP	704-779	84.76	3.62
Q12873_CHD3	704-779	84.71	3.62
O60231_DHX16	704-779	84.66	3.61
Q14004_CDK13	704-779	84.57	3.61
Q8IY81_FTSJ3	704-779	84.34	3.6
Q9H2T7_RANBP17	704-779	84.29	3.59
Q96GQ7_DDX27	704-779	84.12	3.58
Q9Y4C8_RBM19	704-779	84.07	3.58
Q8TCS8_PNPT1	704-779	84.06	3.58
Q9BXT8_RNF17	704-779	83.97	3.57
P38646_HSPA9	704-779	83.9	3.57
Q8NI36_WDR36	704-779	83.86	3.57
Q5T0F9_CC2D1B	704-779	83.83	3.57
P27816_MAP4	704-779	83.71	3.56

Q9NRC6_SPTBN5	704-779	83.58	3.55
Q9UPR6_ZFR2	704-779	83.37	3.54
Q07954_LRP1	704-779	83.32	3.54
Q9H0H0_INTS2	704-779	83.2	3.53
Q5VYS8_TUT7	704-779	83.19	3.53
Q9UPN6_SCAF8	704-779	83.13	3.53
Q9UIA9_XPO7	704-779	83.1	3.52
Q8IYT2_CMTR2	704-779	83.08	3.52
Q9UGU0_TCF20	704-779	83.06	3.52
Q9BQG0_MYBBP1A	704-779	83.05	3.52
Q6NWY9_PRPF40B	704-779	82.92	3.51
Q9BXT6_MOV10L1	704-779	82.88	3.51
Q29RF7_PDS5A	704-779	82.68	3.5
Q09161_NCBP1	704-779	82.64	3.5
Q9HCK5_AGO4	704-779	82.34	3.48
Q9UPT8_ZC3H4	704-779	82.3	3.48
Q96QU8_XPO6	704-779	82.14	3.47
Q12788_TBL3	704-779	82.01	3.46
Q9HBL0_TNS1	704-779	81.97	3.46
Q8N5U6_RNF10	704-779	81.96	3.46
Q92973_TNPO1	704-779	81.92	3.46
Q6ZP01_RBM44	704-779	81.73	3.44
Q8WTT2_NOC3L	704-779	81.72	3.44
Q8N3C0_ASCC3	704-779	81.71	3.44
Q14147_DHX34	704-779	81.68	3.44
Q14764_MVP	704-779	81.63	3.44
Q9UPQ9_TNRC6B	704-779	81.6	3.44
Q9Y608_LRRFIP2	1-76	81.55	3.44
Q92499_DDX1	704-779	81.54	3.43
P47897_QARS1	704-779	81.53	3.43
P04114_APOB	704-779	81.5	3.43
Q9C0B9_ZCCHC2	704-779	81.26	3.42
P52756_RBM5	704-779	81.23	3.42
Q75QN2_INTS8	704-779	81.1	3.41
Q9H9J4_USP42	704-779	81.1	3.41
Q99798_ACO2	704-779	81.06	3.41
Q9NVR5_DNAAF2	704-779	81.02	3.4
Q96RP9_GFM1	704-779	80.86	3.4
Q9BXJ9_NAA15	704-779	80.66	3.38
P58107_EPPK1	704-779	80.58	3.38
Q8N201_INTS1	704-779	80.55	3.38
O94906_PRPF6	704-779	80.49	3.37
Q7L8L6_FASTKD5	704-779	80.36	3.37

P55060_CSE1L	704-779	80.32	3.36
Q9BVJ6_UTP14A	704-779	80.24	3.36
Q99973_TEP1	704-779	80.22	3.36
Q5TAP6_UTP14C	704-779	80.21	3.36
O43896_KIF1C	704-779	80.08	3.35
Q00839_HNRNPU	704-779	80.05	3.35
Q9UI26_IPO11	704-779	80.04	3.35
P98175_RBM10	704-779	79.94	3.34
Q9NX05_FAM120C	704-779	79.77	3.33
Q9UKK3_PARP4	704-779	79.71	3.33
P11171_EPB41	704-779	79.69	3.33
Q92667_AKAP1	704-779	79.48	3.32
O43823_AKAP8	704-779	79.46	3.31
O15381_NVL	704-779	79.41	3.31
Q9NYF8_BCLAF1	704-779	79.41	3.31
P13010_XRCC5	704-779	79.39	3.31
Q9NZB2_FAM120A	704-779	79.37	3.31
Q16891_IMMT	704-779	79.35	3.31
O15455_TLR3	704-779	79.32	3.31
P21399_ACO1	704-779	79.15	3.3
Q8NDG6_TDRD9	704-779	79.12	3.29
Q5H9U9_DDX60L	704-779	79.05	3.29
P35579_MYH9	704-779	78.84	3.28
Q8IZH2_XRN1	704-779	78.78	3.28
Q96RS0_TGS1	704-779	78.76	3.27
Q76FK4_NOL8	704-779	78.69	3.27
P13639_EEF2	704-779	78.63	3.27
P59045_NLRP11	704-779	78.59	3.27
Q01974_ROR2	704-779	78.56	3.26
Q86Y13_DZIP3	704-779	78.5	3.26
Q15424_SAFB	704-779	78.47	3.26
P56192_MARS1	704-779	78.44	3.26
Q9NR48_ASH1L	704-779	78.42	3.25
Q9H2Y7_ZNF106	704-779	78.41	3.25
P29375_KDM5A	704-779	78.39	3.25
Q8NHU6_TDRD7	704-779	78.29	3.25
O60518_RANBP6	704-779	78.1	3.24
Q5XG87_TENT4A	704-779	78.07	3.23
Q7Z6Z7_HUWE1	704-779	77.98	3.23
P19338_NCL	704-779	77.65	3.21
Q8NBJ9_SIDT2	704-779	77.65	3.21
Q9UBK2_PPARGC1A	704-779	77.63	3.21
Q9UNX4_WDR3	704-779	77.6	3.21

A2RTX5_TARS3	704-779	77.58	3.21
Q53GS7_GLE1	704-779	77.49	3.2
O43432_EIF4G3	704-779	77.26	3.19
Q9H9P5_UNKL	704-779	76.98	3.17

Supplementary Table 24. Gene set enrichment analysis performed with DAVID online tool of the predicted *PSG8-AS1*-binding proteome.

Showing FDR < 0.2 results.

TERM	PVALUE	BONFERRONI	BENJAMINI	FDR
GO:0006886~intracellular protein transport	1.09E-05	8.56E-03	8.59E-03	8.59E-03
KW-0333~Golgi apparatus	5.05E-04	1.41E-02	1.47E-02	1.26E-02
hsa04530:Tight junction	1.55E-04	1.38E-02	1.39E-02	1.39E-02
GO:0005794~Golgi apparatus	5.16E-05	1.43E-02	1.44E-02	1.44E-02
KW-0206~Cytoskeleton	1.47E-03	4.04E-02	2.13E-02	1.84E-02
KW-0653~Protein transport	4.46E-04	2.43E-02	2.45E-02	2.45E-02
KW-0965~Cell junction	7.33E-03	1.86E-01	7.09E-02	6.11E-02
KW-1003~Cell membrane	1.30E-02	3.07E-01	8.08E-02	6.96E-02
KW-0472~Membrane	1.39E-02	3.25E-01	8.08E-02	6.96E-02
KW-0796~Tight junction	2.01E-02	4.34E-01	9.71E-02	8.37E-02
GO:0000139~Golgi membrane	7.62E-04	1.91E-01	1.06E-01	1.06E-01
KW-0493~Microtubule	3.41E-02	6.21E-01	1.41E-01	1.22E-01
R-HSA-948021~Transport to the Golgi and subsequent modification	8.51E-04	3.02E-01	1.80E-01	1.80E-01
R-HSA-199977~ER to Golgi Anterograde Transport	8.51E-04	3.02E-01	1.80E-01	1.80E-01

References

Abdelaal, T., Michielsen, L., Cats, D., Hoogduin, D., Mei, H., Reinders, M.J.T., Mahfouz, A., 2019. A comparison of automatic cell identification methods for single-cell RNA sequencing data. Genome Biol. 20, 194. https://doi.org/10.1186/s13059-019-1795-z

Abel, K.M., Wicks, S., Susser, E.S., Dalman, C., Pedersen, M.G., Mortensen, P.B., Webb, R.T., 2010. Birth weight, schizophrenia, and adult mental disorder: is risk confined to the smallest babies? Arch. Gen. Psychiatry 67, 923–930. https://doi.org/10.1001/archgenpsychiatry.2010.100

Abrajano, J.J., Qureshi, I.A., Gokhan, S., Zheng, D., Bergman, A., Mehler, M.F., 2009. Differential Deployment of REST and CoREST Promotes Glial Subtype Specification and Oligodendrocyte Lineage Maturation. PLoS ONE 4, e7665. https://doi.org/10.1371/journal.pone.0007665

Abyzov, A., Urban, A.E., Snyder, M., Gerstein, M., 2011. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 21, 974–984. https://doi.org/10.1101/gr.114876.110

Agid, O., Shapira, B., Zislin, J., Ritsner, M., Hanin, B., Murad, H., Troudart, T., Bloch, M., Heresco-Levy, U., Lerer, B., 1999. Environment and vulnerability to major psychiatric illness: a case control study of early parental loss in major depression, bipolar disorder and schizophrenia. Mol. Psychiatry 4, 163–172. https://doi.org/10.1038/sj.mp.4000473

Aguirre, M., Rivas, M.A., Priest, J., 2019. Phenome-wide Burden of Copy-Number Variation in the UK Biobank. Am. J. Hum. Genet. 105, 373–383. https://doi.org/10.1016/j.ajhg.2019.07.001

Akbarian, S., Liu, C., Knowles, J.A., Vaccarino, F.M., Farnham, P.J., Crawford, G.E., Jaffe, A.E., Pinto, D., Dracheva, S., Geschwind, D.H., Mill, J., Nairn, A.C., Abyzov, A., Pochareddy, S., Prabhakar, S., Weissman, S., Sullivan, P.F., State, M.W., Weng, Z., Peters, M.A., White, K.P., Gerstein, M.B., Amiri, A., Armoskus, C., Ashley-Koch, A.E., Bae, T., Beckel-Mitchener, A., Berman, B.P., Coetzee, G.A., Coppola, G., Francoeur, N., Fromer, M., Gao, R., Grennan, K., Herstein, J., Kavanagh, D.H., Ivanov, N.A., Jiang, Y., Kitchen, R.R., Kozlenkov, A., Kundakovic, M., Li, M., Li, Z., Liu, S., Mangravite, L.M., Mattei, E., Markenscoff-Papadimitriou, E., Navarro, F.C.P., North, N., Omberg, L., Panchision, D., Parikshak, N., Poschmann, J., Price, A.J., Purcaro, M., Reddy, T.E., Roussos, P., Schreiner, S., Scuderi, S., Sebra, R., Shibata, M., Shieh, A.W., Skarica, M., Sun, W., Swarup, V., Thomas, A., Tsuji, J., van Bakel, H., Wang, D., Wang, Y., Wang, K., Werling, D.M., Willsey, A.J., Witt, H., Won, H., Wong, C.C.Y., Wray, G.A., Wu, E.Y., Xu, X., Yao, L., Senthil, G., Lehner, T., Sklar, P., Sestan, N., 2015. The PsychENCODE project. Nat. Neurosci. 18, 1707–1712. https://doi.org/10.1038/nn.4156

AL-Asadi, A.M., Klein, B., Meyer, D., 2015. Multiple Comorbidities of 21 Psychological Disorders and Relationships With Psychosocial Variables: A Study of the Online Assessment and Diagnostic System Within a Web-Based Population. J. Med. Internet Res. 17, e4143. https://doi.org/10.2196/jmir.4143

Allen, N.J., Lyons, D.A., 2018. Glia as architects of central nervous system formation and function. Science 362, 181–185. https://doi.org/10.1126/science.aat0473

Amos-Landgraf, J.M., Ji, Y., Gottlieb, W., Depinet, T., Wandstrat, A.E., Cassidy, S.B., Driscoll, D.J., Rogan, P.K., Schwartz, S., Nicholls, R.D., 1999. Chromosome breakage in the Prader-Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints. Am. J. Hum. Genet. 65, 370–386. https://doi.org/10.1086/302510

An, N.A., Zhang, J., Mo, F., Luan, X., Tian, L., Shen, Q.S., Li, X., Li, C., Zhou, F., Zhang, B., Ji, M., Qi, J., Zhou, W.-Z., Ding, W., Chen, J.-Y., Yu, J., Zhang, L., Shu, S., Hu, B., Li, C.-Y., 2023. De novo genes with an lncRNA origin encode unique human brain developmental functionality. Nat. Ecol. Evol. 7, 264–278. https://doi.org/10.1038/s41559-022-01925-6

Anderson, D.M., Anderson, K.M., Chang, C.-L., Makarewich, C.A., Nelson, B.R., McAnally, J.R., Kasaragod, P., Shelton, J.M., Liou, J., Bassel-Duby, R., Olson, E.N., 2015. A Micropeptide Encoded by a Putative Long Non-coding RNA Regulates Muscle Performance. Cell 160, 595–606. https://doi.org/10.1016/j.cell.2015.01.009

Antonacci, F., Dennis, M.Y., Huddleston, J., Sudmant, P.H., Steinberg, K.M., Rosenfeld, J.A., Miroballo, M., Graves, T.A., Vives, L., Malig, M., Denman, L., Raja, A., Stuart, A., Tang, J., Munson, B., Shaffer, L.G., Amemiya, C.T., Wilson, R.K., Eichler, E.E., 2014. Palindromic GOLGA8 core duplicons promote chromosome 15q13.3 microdeletion and evolutionary instability. Nat. Genet. 46, 1293–1302. https://doi.org/10.1038/ng.3120

Arlt, M.F., Wilson, T.E., Glover, T.W., 2012. Replication Stress and Mechanisms of CNV Formation. Curr. Opin. Genet. Dev. 22, 204–210. https://doi.org/10.1016/j.gde.2012.01.009

Armaos, A., Colantoni, A., Proietti, G., Rupert, J., Tartaglia, G.G., 2021. catRAPID omics v2.0: going deeper and wider in the prediction of protein–RNA interactions. Nucleic Acids Res. 49, W72–W79. https://doi.org/10.1093/nar/gkab393

Arnold, L.L., Doherty, T.M., Flor, A.W., Simon, J.A., Chou, J.Y., Chan, W.-Y., Mansfield, B.C., 1999. Pregnancy-Specific Glycoprotein Gene Expression in Recurrent Aborters: A Potential Correlation to Interleukin-10 Expression. Am. J. Reprod. Immunol. 41, 174–182. https://doi.org/10.1111/j.1600-0897.1999.tb00530.x

Arseneault, L., Cannon, M., Fisher, H.L., Polanczyk, G., Moffitt, T.E., Caspi, A., 2011. Childhood trauma and children's emerging psychotic symptoms: A genetically sensitive longitudinal cohort study. Am. J. Psychiatry 168, 65–72. https://doi.org/10.1176/appi.ajp.2010.10040567

Asou, H., Hamada, K., Miyazaki, T., Sakota, T., Hayashi, K., Takeda, Y., Marret, S., Delpech, B., Itoh, K., Uyemura, K., 1995. CNS myelinogenesis in vitro: time course and pattern of rat oligodendrocyte development. J. Neurosci. Res. 40, 519–534. https://doi.org/10.1002/jnr.490400411

Aubert, M., Lebe, R., Oktaviana, A.A., Tang, M., Burhan, B., Hamrullah, Jusdi, A., Abdullah, Hakim, B., Zhao, J., Geria, I.M., Sulistyarto, P.H., Sardi, R., Brumm, A., 2019. Earliest hunting scene in prehistoric art. Nature 576, 442–445. https://doi.org/10.1038/s41586-019-1806-y

Baba, M., Yokoyama, K., Seiriki, K., Naka, Y., Matsumura, K., Kondo, M., Yamamoto, K., Hayashida, M., Kasai, A., Ago, Y., Nagayasu, K., Hayata-Takano, A., Takahashi, A., Yamaguchi, S., Mori, D., Ozaki, N., Yamamoto, T., Takuma, K., Hashimoto, R., Hashimoto, H., Nakazawa, T., 2019. Psychiatric-disorder-related behavioral phenotypes and cortical hyperactivity in a mouse model of 3q29 deletion syndrome. Neuropsychopharmacology 44, 2125–2135. https://doi.org/10.1038/s41386-019-0441-5

Bailey, J.A., Eichler, E.E., 2006. Primate segmental duplications: crucibles of evolution, diversity and disease. Nat. Rev. Genet. 7, 552–564. https://doi.org/10.1038/nrg1895

Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J., Eichler, E.E., 2001. Segmental Duplications: Organization and Impact Within the Current Human Genome Project Assembly. Genome Res. 11, 1005–1017. https://doi.org/10.1101/gr.187101

Ballesteros, A., Mentink-Kane, M.M., Warren, J., Kaplan, G.G., Dveksler, G.S., 2015. Induction and Activation of Latent Transforming Growth Factor-β1 Are Carried out by Two Distinct Domains of Pregnancy-specific Glycoprotein 1 (PSG1)*. J. Biol. Chem. 290, 4422–4431. https://doi.org/10.1074/jbc.M114.597518

Barešić, A., Nash, A.J., Dahoun, T., Howes, O., Lenhard, B., 2020. Understanding the genetics of neuropsychiatric disorders: the potential role of genomic regulatory blocks. Mol. Psychiatry 25, 6–18. https://doi.org/10.1038/s41380-019-0518-x

Baron, W., Ozgen, H., Klunder, B., de Jonge, J.C., Nomden, A., Plat, A., Trifilieff, E., de Vries, H., Hoekstra, D., 2015. The Major Myelin-Resident Protein PLP Is Transported to Myelin Membranes via a Transcytotic Mechanism: Involvement of Sulfatide. Mol. Cell. Biol. 35, 288–302. https://doi.org/10.1128/MCB.00848-14

Bartel, D.P., 2018. Metazoan MicroRNAs. Cell 173, 20–51. https://doi.org/10.1016/j.cell.2018.03.006

Bartonicek, N., Clark, M.B., Quek, X.C., Torpy, J.R., Pritchard, A.L., Maag, J.L.V., Gloss, B.S., Crawford, J., Taft, R.J., Hayward, N.K., Montgomery, G.W., Mattick, J.S., Mercer, T.R., Dinger, M.E., 2017. Intergenic disease-associated regions are abundant in novel transcripts. Genome Biol. 18, 241. https://doi.org/10.1186/s13059-017-1363-3

Bassett, A.S., Chow, E.W.C., 2008. Schizophrenia and 22q11.2 Deletion Syndrome. Curr. Psychiatry Rep. 10, 148–157.

Baudoin, N.C., Bloomfield, M., 2021. Karyotype Aberrations in Action: The Evolution of Cancer Genomes and the Tumor Microenvironment. Genes 12, 558. https://doi.org/10.3390/genes12040558

Beaulieu, J.-M., Sotnikova, T.D., Marion, S., Lefkowitz, R.J., Gainetdinov, R.R., Caron, M.G., 2005. An Akt/β-Arrestin 2/PP2A Signaling Complex Mediates Dopaminergic Neurotransmission and Behavior. Cell 122, 261–273. https://doi.org/10.1016/j.cell.2005.05.012

Berg, J., Sorensen, S.A., Ting, J.T., Miller, J.A., Chartrand, T., Buchin, A., Bakken, T.E., Budzillo, A., Dee, N., Ding, S.-L., Gouwens, N.W., Hodge, R.D., Kalmbach, B., Lee, C., Lee, B.R., Alfiler, L., Baker, K., Barkan, E., Beller, A.,

Berry, K., Bertagnolli, D., Bickley, K., Bomben, J., Braun, T., Brouner, K., Casper, T., Chong, P., Crichton, K., Dalley, R., de Frates, R., Desta, T., Lee, S.D., D'Orazi, F., Dotson, N., Egdorf, T., Enstrom, R., Farrell, C., Feng, D., Fong, O., Furdan, S., Galakhova, A.A., Gamlin, C., Gary, A., Glandon, A., Goldy, J., Gorham, M., Goriounova, N.A., Gratiy, S., Graybuck, L., Gu, H., Hadley, K., Hansen, N., Heistek, T.S., Henry, A.M., Heyer, D.B., Hill, D., Hill, C., Hupp, M., Jarsky, T., Kebede, S., Keene, L., Kim, L., Kim, M.-H., Kroll, M., Latimer, C., Levi, B.P., Link, K.E., Mallory, M., Mann, R., Marshall, D., Maxwell, M., McGraw, M., McMillen, D., Melief, E., Mertens, E.J., Mezei, L., Mihut, N., Mok, S., Molnar, G., Mukora, A., Ng, Lindsay, Ngo, K., Nicovich, P.R., Nyhus, J., Olah, G., Oldre, A., Omstead, V., Ozsvar, A., Park, D., Peng, H., Pham, T., Pom, C.A., Potekhina, L., Rajanbabu, R., Ransford, S., Reid, D., Rimorin, C., Ruiz, A., Sandman, D., Sulc, J., Sunkin, S.M., Szafer, A., Szemenyei, V., Thomsen, E.R., Tieu, M., Torkelson, A., Trinh, J., Tung, H., Wakeman, W., Waleboer, F., Ward, K., Wilbers, R., Williams, G., Yao, Z., Yoon, J.-G., Anastassiou, C., Arkhipov, A., Barzo, P., Bernard, A., Cobbs, C., de Witt Hamer, P.C., Ellenbogen, R.G., Esposito, L., Ferreira, M., Gwinn, R.P., Hawrylycz, M.J., Hof, P.R., Idema, S., Jones, A.R., Keene, C.D., Ko, A.L., Murphy, G.J., Ng, Lydia, Ojemann, J.G., Patel, A.P., Phillips, J.W., Silbergeld, D.L., Smith, K., Tasic, B., Yuste, R., Segev, I., de Kock, C.P.J., Mansvelder, H.D., Tamas, G., Zeng, H., Koch, C., Lein, E.S., 2021. Human neocortical expansion involves glutamatergic neuron diversification. Nature 598, 151–158. https://doi.org/10.1038/s41586-021-03813-8

Bergen, S.E., Ploner, A., Howrigan, D., O'Donovan, M.C., Smoller, J.W., Sullivan, P.F., Sebat, J., Neale, B., Kendler, K.S., 2019. Joint Contributions of Rare Copy Number Variants and Common SNPs to Risk for Schizophrenia. Am. J. Psychiatry 176, 29–35. https://doi.org/10.1176/appi.ajp.2018.17040467

Berto, S., Mendizabal, I., Usui, N., Toriumi, K., Chatterjee, P., Douglas, C., Tamminga, C.A., Preuss, T.M., Yi, S.V., Konopka, G., 2019. Accelerated evolution of oligodendrocytes in the human brain. Proc. Natl. Acad. Sci. 116, 24334–24342. https://doi.org/10.1073/pnas.1907982116

Besnard, J., Ruda, G.F., Setola, V., Abecassis, K., Rodriguiz, R.M., Huang, X.-P., Norval, S., Sassano, M.F., Shin, A.I., Webster, L.A., Simeons, F.R.C., Stojanovski, L., Prat, A., Seidah, N.G., Constam, D.B., Bickerton, G.R., Read, K.D., Wetsel, W.C., Gilbert, I.H., Roth, B.L., Hopkins, A.L., 2012. Automated design of ligands to polypharmacological profiles. Nature 492, 10.1038/nature11691. https://doi.org/10.1038/nature11691

Bettegowda, C., Agrawal, N., Jiao, Y., Sausen, M., Wood, L.D., Hruban, R.H., Rodriguez, F.J., Cahill, D.P., McLendon, R., Riggins, G., Velculescu, V.E., Oba-Shinjo, S.M., Nagahashi Marie, S.K., Vogelstein, B., Bigner, D., Yan, H., Papadopoulos, N., Kinzler, K.W., 2011. Mutations in CIC and FUBP1 Contribute to Human Oligodendroglioma. Science 333, 1453–1455. https://doi.org/10.1126/science.1210557

Bhattacharyya, U., Bhatia, T., Deshpande, S.N., Thelma, B.K., 2022. Genetic variations in evolutionary accelerated regions disrupt cognition in

schizophrenia. Psychiatry Res. 314, 114586. https://doi.org/10.1016/j.psychres.2022.114586
Bigdeli, T.B., Genovese, G., Georgakopoulos, P., Meyers, J.L., Peterson, R.E., Iyegbe, C.O., Medeiros, H., Valderrama, J., Achtyes, E.D., Kotov, R., Stahl, E.A., Abbott, C., Azevedo, M.H., Belliveau, R.A., Bevilacqua, E., Bromet, E.J., Byerley, W., Carvalho, C.B., Chapman, S.B., DeLisi, L.E., Dumont, A.L., O'Dushlaine, C., Evgrafov, O.V., Fochtmann, L.J., Gage, D., Kennedy, J.L., Kinkead, B., Macedo, A., Moran, J.L., Morley, C.P., Dewan, M.J., Nemesh, J., Perkins, D.O., Purcell, S.M., Rakofsky, J.J., Scolnick, E.M., Sklar, B.M., Sklar, P., Smoller, J.W., Sullivan, P.F., Macciardi, F., Marder, S.R., Gur, R.C., Gur, R.E., Braff, D.L., Nicolini, H., Escamilla, M.A., Vawter, M.P., Sobell, J.L., Malaspina, D., Lehrer, D.S., Buckley, P.F., Rapaport, M.H., Knowles, J.A., Fanous, A.H., Pato, M.T., McCarroll, S.A., Pato, C.N., 2020. Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. Mol. Psychiatry 25, 2455–2467. https://doi.org/10.1038/s41380-019-0517-y
Blankley, R.T., Fisher, C., Westwood, M., North, R., Baker, P.N., Walker, M.J., Williamson, A., Whetton, A.D., Lin, W., McCowan, L., Roberts, C.T., Cooper, G.J.S., Unwin, R.D., Myers, J.E., 2013. A Label-free Selected Reaction Monitoring Workflow Identifies a Subset of Pregnancy Specific Glycoproteins as Potential Predictive Markers of Early-onset Pre-eclampsia *. Mol. Cell. Proteomics 12, 3148–3159. https://doi.org/10.1074/mcp.M112.026872
Boeva, V., Zinovyev, A., Bleakley, K., Vert, J.-P., Janoueix-Lerosey, I., Delattre, O., Barillot, E., 2011. Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. Bioinforma. Oxf. Engl. 27, 268–269. https://doi.org/10.1093/bioinformatics/btq635
Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., Villanueva, A., Esteller, M., Guil, S., 2015. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. Proc. Natl. Acad. Sci. U. S. A. 112, 5785–5790. https://doi.org/10.1073/pnas.1421197112
Bradl, M., Lassmann, H., 2010. Oligodendrocytes: biology and pathology. Acta Neuropathol. (Berl.) 119, 37–53. https://doi.org/10.1007/s00401-009-0601-5
Brandler, W.M., Antaki, D., Gujral, M., Noor, A., Rosanio, G., Chapman, T.R., Barrera, D.J., Lin, G.N., Malhotra, D., Watts, A.C., Wong, L.C., Estabillo, J.A., Gadomski, T.E., Hong, O., Fajardo, K.V.F., Bhandari, A., Owen, R., Baughn, M., Yuan, J., Solomon, T., Moyzis, A.G., Maile, M.S., Sanders, S.J., Reiner, G.E., Vaux, K.K., Strom, C.M., Zhang, K., Muotri, A.R., Akshoomoff, N., Leal, S.M., Pierce, K., Courchesne, E., Iakoucheva, L.M., Corsello, C., Sebat, J., 2016. Frequency and Complexity of De Novo Structural Mutation in Autism. Am. J. Hum. Genet. 98, 667–679. https://doi.org/10.1016/j.ajhg.2016.02.018
Bristow, G.C., Thomson, D.M., Openshaw, R.L., Mitchell, E.J., Pratt, J.A., Dawson, N., Morris, B.J., 2020. 16p11 Duplication Disrupts Hippocampal-Orbitofrontal-Amygdala Connectivity, Revealing a Neural Circuit

Endophenotype for Schizophrenia. Cell Rep. 31, 107536. https://doi.org/10.1016/j.celrep.2020.107536
Bryois, J., Calini, D., Macnair, W., Foo, L., Urich, E., Ortmann, W., Iglesias, V.A., Selvaraj, S., Nutma, E., Marzin, M., Amor, S., Williams, A., Castelo-Branco, G., Menon, V., De Jager, P., Malhotra, D., 2022. Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. Nat. Neurosci. 25, 1104–1112. https://doi.org/10.1038/s41593-022-01128-z
Burgess, T., Brown, N.J., Stark, Z., Bruno, D.L., Oertel, R., Chong, B., Calabro, V., Kornberg, A., Sanderson, C., Kelly, J., Howell, K.B., Savarirayan, R., Hinds, R., Greenway, A., Slater, H.R., White, S.M., 2014. Characterization of core clinical phenotypes associated with recurrent proximal 15q25.2 microdeletions. Am. J. Med. Genet. A. 164A, 77–86. https://doi.org/10.1002/ajmg.a.36203
Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., Rinn, J.L., 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 25, 1915–1927. https://doi.org/10.1101/gad.17446611
Cai, H., Yu, Y., Ni, X., Li, C., Hu, Y., Wang, J., Chen, F., Xi, S., Chen, Z., 2020. LncRNA LINC00998 inhibits the malignant glioma phenotype via the CBX3-mediated c-Met/Akt/mTOR axis. Cell Death Dis. 11, 1–17. https://doi.org/10.1038/s41419-020-03247-6
Cai, M., Que, Y., Chen, X., Chen, Y., Liang, B., Huang, H., Xu, L., Lin, N., 2022. 16p13.11 microdeletion/microduplication in fetuses: investigation of associated ultrasound phenotypes, genetic anomalies, and pregnancy outcome follow-up. BMC Pregnancy Childbirth 22, 913. https://doi.org/10.1186/s12884-022-05267-w
Calo, E., Flynn, R.A., Martin, L., Spitale, R.C., Chang, H.Y., Wysocka, J., 2015. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature 518, 249–253. https://doi.org/10.1038/nature13923
Campbell, C.D., Sampas, N., Tsalenko, A., Sudmant, P.H., Kidd, J.M., Malig, M., Vu, T.H., Vives, L., Tsang, P., Bruhn, L., Eichler, E.E., 2011. Population-genetic properties of differentiated human copy-number polymorphisms. Am. J. Hum. Genet. 88, 317–332. https://doi.org/10.1016/j.ajhg.2011.02.004
Cannon, M., Jones, P.B., Murray, R.M., 2002. Obstetric complications and schizophrenia: historical and meta-analytic review. Am. J. Psychiatry 159, 1080–1092. https://doi.org/10.1176/appi.ajp.159.7.1080
Cantsilieris, S., Sunkin, S.M., Johnson, M.E., Anaclerio, F., Huddleston, J., Baker, C., Dougherty, M.L., Underwood, J.G., Sulovari, A., Hsieh, P., Mao, Y., Catacchio, C.R., Malig, M., Welch, A.E., Sorensen, M., Munson, K.M., Jiang, W., Girirajan, S., Ventura, M., Lamb, B.T., Conlon, R.A., Eichler, E.E., 2020. An evolutionary driver of interspersed segmental duplications in primates. Genome Biol. 21, 202. https://doi.org/10.1186/s13059-020-02074-4
Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V.B., Brenner, S.E., Batalov, S., Forrest, A.R.R., Zavolan, M., Davis, M.J., Wilming, L.G., Aidinis, V., Allen, J.E., Ambesi-Impiombato, A., Apweiler,

R., Aturaliya, R.N., Bailey, T.L., Bansal, M., Baxter, L., Beisel, K.W., Bersano, T., Bono, H., Chalk, A.M., Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R., Crowe, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C.F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T.R., Gojobori, T., Green, R.E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T.K., Hirokawa, N., Hill, D., Huminiecki, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S.P.T., Kruger, A., Kummerfeld, S.K., Kurochkin, I.V., Lareau, L.F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Morris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakauchi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K.C., Pavan, W.J., Pavesi, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.F., Ring, B.Z., Ringwald, M., Rost, B., Ruan, Y., Salzberg, S.L., Sandelin, A., Schneider, C., Schönbach, C., Sekiguchi, K., Semple, C. a. M., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, E., Sugiura, K., Sultana, R., Takenaka, Y., Taki, K., Tammoja, K., Tan, S.L., Tang, S., Taylor, M.S., Tegner, J., Teichmann, S.A., Ueda, H.R., van Nimwegen, E., Verardo, R., Wei, C.L., Yagi, K., Yamanishi, H., Zabarovsky, E., Zhu, S., Zimmer, A., Hide, W., Bult, C., Grimmond, S.M., Teasdale, R.D., Liu, E.T., Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J.S., Hume, D.A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Plessy, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watahiki, A., Okamura-Oho, Y., Suzuki, H., Kawai, J., Hayashizaki, Y., FANTOM Consortium, RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group), 2005. The transcriptional landscape of the mammalian genome. Science 309, 1559–1563. https://doi.org/10.1126/science.1112014

Carson, J.H., Worboys, K., Ainger, K., Barbarese, E., 1997. Translocation of myelin basic protein mRNA in oligodendrocytes requires microtubules and kinesin. Cell Motil. Cytoskeleton 38, 318–328. https://doi.org/10.1002/(SICI)1097-0169(1997)38:4<318::AID-CM2>3.0.CO;2-#

Caseras, X., Kirov, G., Kendall, K.M., Rees, E., Legge, S.E., Bracher-Smith, M., Escott-Price, V., Murphy, K., 2021. Effects of genomic copy number variants penetrant for schizophrenia on cortical thickness and surface area in healthy individuals: analysis of the UK Biobank. Br. J. Psychiatry 218, 104–111. https://doi.org/10.1192/bjp.2020.139

Cengiz, M., Karacetin, G., Topal, M., Yuksel, M.E., Eseroglu, T., Akdeniz, G.B., Demir, T., Erkiran, M., Dirican, A., Bayoglu, B., 2023. COMT rs4680 and DRD2 rs6275 variants and their association with YMRS scores in children

with early-onset bipolar disorder. Eur. J. Psychiatry 37, 8–14. https://doi.org/10.1016/j.ejpsy.2022.07.003
Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A.P., Sander, C., Schultz, N., 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2, 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095
Chang, C.L., Chang, C.Y., Lee, D.X., Cheng, P.J., 2016. Characterization of Human Pregnancy Specific Glycoprotein (PSG) Gene Copy Number Variations in Pre-eclampsia Patients. Adv. Exp. Med. Biol. 924, 63–65. https://doi.org/10.1007/978-3-319-42044-8_12
Chang, H., Li, L., Li, M., Xiao, X., 2017. Rare and common variants at 16p11.2 are associated with schizophrenia. Schizophr. Res. 184, 105–108. https://doi.org/10.1016/j.schres.2016.11.031
Charlson, F.J., Ferrari, A.J., Santomauro, D.F., Diminic, S., Stockings, E., Scott, J.G., McGrath, J.J., Whiteford, H.A., 2018. Global Epidemiology and Burden of Schizophrenia: Findings From the Global Burden of Disease Study 2016. Schizophr. Bull. 44, 1195–1203. https://doi.org/10.1093/schbul/sby058
Chen, J., Calhoun, V.D., Perrone-Bizzozero, N.I., Pearlson, G.D., Sui, J., Du, Y., Liu, J., 2016. A pilot study on commonality and specificity of copy number variants in schizophrenia and bipolar disorder. Transl. Psychiatry 6, e824–e824. https://doi.org/10.1038/tp.2016.96
Chen, S., The gnomAD consortium, 2022. A genome-wide mutational constraint map quantified from variation in 76,156 human genomes | bioRxiv [WWW Document]. URL https://www.biorxiv.org/content/10.1101/2022.03.20.485034v2.full (accessed 6.7.23).
Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., Saunders, C.T., 2016. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32, 1220–1222. https://doi.org/10.1093/bioinformatics/btv710
Chiaromonte, F., Yap, V.B., Miller, W., 2002. Scoring pairwise genomic sequence alignments. Pac. Symp. Biocomput. Pac. Symp. Biocomput. 115–126. https://doi.org/10.1142/9789812799623_0012
Chirita Emandi, A., Dobrescu, A.I., Doros, G., Hyon, C., Miclea, D., Popoiu, C., Puiu, M., Arghirescu, S., 2019. A Novel 3q29 Deletion in Association With Developmental Delay and Heart Malformation—Case Report With Literature Review. Front. Pediatr. 7.
Chu, C., Qu, K., Zhong, F., Artandi, S.E., Chang, H.Y., 2011. Genomic maps of lincRNA occupancy reveal principles of RNA-chromatin interactions. Mol. Cell 44, 667–678. https://doi.org/10.1016/j.molcel.2011.08.027
Clayton-Smith, J., Giblin, C., Smith, R.A., Dunn, C., Willatt, L., 2010. Familial 3q29 microdeletion syndrome providing further evidence of involvement of the 3q29 region in bipolar disorder. Clin. Dysmorphol. 19, 128. https://doi.org/10.1097/MCD.0b013e32833a1e3c
Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C., Ragoussis, J., 2007. QuantiSNP: an Objective Bayes

Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. 35, 2013–2025. https://doi.org/10.1093/nar/gkm076

Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., Abdel-Hamid, H., Bader, P., McCracken, E., Niyazov, D., Leppig, K., Thiese, H., Hummel, M., Alexander, N., Gorski, J., Kussmann, J., Shashi, V., Johnson, K., Rehder, C., Ballif, B.C., Shaffer, L.G., Eichler, E.E., 2011. A copy number variation morbidity map of developmental delay. Nat. Genet. 43, 838–846. https://doi.org/10.1038/ng.909

Corces, M.R., Shcherbina, A., Kundu, S., Gloudemans, M.J., Frésard, L., Granja, J.M., Louie, B.H., Eulalio, T., Shams, S., Bagdatli, S.T., Mumbach, M.R., Liu, B., Montine, K.S., Greenleaf, W.J., Kundaje, A., Montgomery, S.B., Chang, H.Y., Montine, T.J., 2020. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. Nat. Genet. 52, 1158–1168. https://doi.org/10.1038/s41588-020-00721-x

Coutelier, M., Holtgrewe, M., Jäger, M., Flöttman, R., Mensah, M.A., Spielmann, M., Krawitz, P., Horn, D., Beule, D., Mundlos, S., 2022. Combining callers improves the detection of copy number variants from whole-genome sequencing. Eur. J. Hum. Genet. 30, 178–186. https://doi.org/10.1038/s41431-021-00983-x

Craddock, N., Hurles, M.E., Cardin, N., Pearson, R.D., Plagnol, V., Robson, S., Vukcevic, D., Barnes, C., Conrad, D.F., Giannoulatou, E., Holmes, C., Marchini, J.L., Stirrups, K., Tobin, M.D., Wain, L.V., Yau, C., Aerts, J., Ahmad, T., Daniel Andrews, T., Arbury, H., Attwood, A., Auton, A., Ball, S.G., Balmforth, A.J., Barrett, J.C., Barroso, I., Barton, A., Bennett, A.J., Bhaskar, S., Blaszczyk, K., Bowes, J., Brand, O.J., Braund, P.S., Bredin, F., Breen, G., Brown, M.J., Bruce, I.N., Bull, J., Burren, O.S., Burton, J., Byrnes, J., Caesar, S., Clee, C.M., Coffey, A.J., Connell, J.M.C., Cooper, J.D., Dominiczak, A.F., Downes, K., Drummond, H.E., Dudakia, D., Dunham, A., Ebbs, B., Eccles, D., Edkins, S., Edwards, C., Elliot, A., Emery, P., Evans, D.M., Evans, G., Eyre, S., Farmer, A., Nicol Ferrier, I., Feuk, L., Fitzgerald, T., Flynn, E., Forbes, A., Forty, L., Franklyn, J.A., Freathy, R.M., Gibbs, P., Gilbert, P., Gokumen, O., Gordon-Smith, K., Gray, E., Green, E., Groves, C.J., Grozeva, D., Gwilliam, R., Hall, A., Hammond, N., Hardy, M., Harrison, P., Hassanali, N., Hebaishi, H., Hines, S., Hinks, A., Hitman, G.A., Hocking, L., Howard, E., Howard, P., Howson, J.M.M., Hughes, D., Hunt, S., Isaacs, J.D., Jain, M., Jewell, D.P., Johnson, T., Jolley, J.D., Jones, I.R., Jones, L.A., Kirov, G., Langford, C.F., Lango-Allen, H., Mark Lathrop, G., Lee, J., Lee, K.L., Lees, C., Lewis, K., Lindgren, C.M., Maisuria-Armer, M., Maller, J., Mansfield, J., Martin, P., Massey, D.C.O., McArdle, W.L., McGuffin, P., McLay, K.E., Mentzer, A., Mimmack, M.L., Morgan, A.E., Morris, A.P., Mowat, C., Myers, S., Newman, W., Nimmo, E.R., O'Donovan, M.C., Onipinla, A., Onyiah, I., Ovington, N.R., Owen, M.J., Palin, K., Parnell, K., Pernet, D., Perry, J.R.B., Phillips, A., Pinto, D., Prescott, N.J., Prokopenko, I., Quail, M.A., Rafelt, S., Rayner, N.W., Redon, R., Reid, D.M., Renwick, A., Ring, S.M., Robertson, N., Russell, E., St Clair, D., Sambrook, J.G.,

Sanderson, J.D., Schuilenburg, H., Scott, C.E., Scott, R., Seal, S., Shaw-Hawkins, S., Shields, B.M., Simmonds, M.J., Smyth, D.J., Somaskantharajah, E., Spanova, K., Steer, S., Stephens, J., Stevens, H.E., Stone, M.A., Su, Z., Symmons, D.P.M., Thompson, J.R., Thomson, W., Travers, M.E., Turnbull, C., Valsesia, A., Walker, M., Walker, N.M., Wallace, C., Warren-Perry, M., Watkins, N.A., Webster, J., Weedon, M.N., Wilson, A.G., Woodburn, M., Wordsworth, B.P., Young, A.H., Zeggini, E., Carter, N.P., Frayling, T.M., Lee, C., McVean, G., Munroe, P.B., Palotie, A., Sawcer, S.J., Scherer, S.W., Strachan, D.P., Tyler-Smith, C., Brown, M.A., Burton, P.R., Caulfield, M.J., Compston, A., Farrall, M., Gough, S.C.L., Hall, A.S., Hattersley, A.T., Hill, A.V.S., Mathew, C.G., Pembrey, M., Satsangi, J., Stratton, M.R., Worthington, J., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.I., Ouwehand, W.H., Parkes, M., Rahman, N., Todd, J.A., Samani, N.J., Donnelly, P., The Wellcome Trust Case Control Consortium, 2010. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 464, 713–720. https://doi.org/10.1038/nature08979

Dalman, C., Thomas, H.V., David, A.S., Gentz, J., Lewis, G., Allebeck, P., 2001. Signs of asphyxia at birth and risk of schizophrenia. Population-based case-control study. Br. J. Psychiatry J. Ment. Sci. 179, 403–408. https://doi.org/10.1192/bjp.179.5.403

Dasouki, M.J., Lushington, G.H., Hovanes, K., Casey, J., Gorre, M., 2011. The 3q29 microdeletion syndrome: Report of three new unrelated patients and in silico "RNA binding" analysis of the 3q29 region. Am. J. Med. Genet. A. 155, 1654–1660. https://doi.org/10.1002/ajmg.a.34080

de Faria, O., Pivonkova, H., Varga, B., Timmler, S., Evans, K.A., Káradóttir, R.T., 2021. Periods of synchronized myelin changes shape brain function and plasticity. Nat. Neurosci. 24, 1508–1521. https://doi.org/10.1038/s41593-021-00917-2

De Juan Romero, C., Borrell, V., 2015. Coevolution of radial glial cells and the cerebral cortex. Glia 63, 1303–1319. https://doi.org/10.1002/glia.22827

De Kleijn, K.M.A., Zuure, W.A., Peijnenborg, J., Heuvelmans, J.M., Martens, G.J.M., 2019. Reappraisal of Human HOG and MO3.13 Cell Lines as a Model to Study Oligodendrocyte Functioning. Cells 8, 1096. https://doi.org/10.3390/cells8091096

DeFelipe, J., Alonso-Nanclares, L., Arellano, J.I., 2002. Microstructure of the neocortex: comparative aspects. J. Neurocytol. 31, 299–316. https://doi.org/10.1023/a:1024130211265

Dennis, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., Lempicki, R.A., 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 4, R60. https://doi.org/10.1186/gb-2003-4-9-r60

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J.B., Lipovich, L., Gonzalez, J.M., Thomas, M., Davis, C.A., Shiekhattar, R., Gingeras, T.R., Hubbard, T.J., Notredame, C., Harrow, J., Guigó, R., 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and

expression. Genome Res. 22, 1775–1789. https://doi.org/10.1101/gr.132159.111
di Iulio, J., Bartha, I., Wong, E.H.M., Yu, H.-C., Lavrenko, V., Yang, D., Jung, I., Hicks, M.A., Shah, N., Kirkness, E.F., Fabani, M.M., Biggs, W.H., Ren, B., Venter, J.C., Telenti, A., 2018. The human noncoding genome defined by genetic diversity. Nat. Genet. 50, 333–337. https://doi.org/10.1038/s41588-018-0062-7
Dias, R., Evans, D., Chen, S.-F., Chen, K.-Y., Loguercio, S., Chan, L., Torkamani, A., 2022. Rapid, Reference-Free human genotype imputation with denoising autoencoders. eLife 11, e75600. https://doi.org/10.7554/eLife.75600
Dl, H., Cd, S., M, G.-D., Pb, P., Ja, M., J, Z., Jj, A.-G., P, C.-D., H, C., R, de B., M, L., J, R.-M., Gc, W., Awg, P., 2018. U-Th dating of carbonate crusts reveals Neandertal origin of Iberian cave art. Science 359. https://doi.org/10.1126/science.aap7778
Dobson, R., Giovannoni, G., 2019. Multiple sclerosis - a review. Eur. J. Neurol. 26, 27–40. https://doi.org/10.1111/ene.13819
Doelken, S.C., Seeger, K., Hundsdoerfer, P., Weber-Ferro, W., Klopocki, E., Graul-Neumann, L., 2013. Proximal and distal 15q25.2 microdeletions-genotype-phenotype delineation of two neurodevelopmental susceptibility loci. Am. J. Med. Genet. A. 161A, 218–224. https://doi.org/10.1002/ajmg.a.35695
Dollfus, S., Lyne, J., 2017. Negative symptoms: History of the concept and their position in diagnosis of schizophrenia. Schizophr. Res., Special Section: Negative Symptoms 186, 3–7. https://doi.org/10.1016/j.schres.2016.06.024
Donahue, C.J., Glasser, M.F., Preuss, T.M., Rilling, J.K., Van Essen, D.C., 2018. Quantitative assessment of prefrontal cortex in humans relative to nonhuman primates. Proc. Natl. Acad. Sci. 115, E5183–E5192. https://doi.org/10.1073/pnas.1721653115
Dong, X., Chen, K., Cuevas-Diaz Duran, R., You, Y., Sloan, S.A., Zhang, Y., Zong, S., Cao, Q., Barres, B.A., Wu, J.Q., 2015. Comprehensive Identification of Long Non-coding RNAs in Purified Cell Types from the Brain Reveals Functional LncRNA in OPC Fate Determination. PLoS Genet. 11, e1005669. https://doi.org/10.1371/journal.pgen.1005669
Dor, T., Cinnamon, Y., Raymond, L., Shaag, A., Bouslam, N., Bouhouche, A., Gaussen, M., Meyer, V., Durr, A., Brice, A., Benomar, A., Stevanin, G., Schuelke, M., Edvardson, S., 2014. KIF1C mutations in two families with hereditary spastic paraparesis and cerebellar dysfunction. J. Med. Genet. 51, 137–142. https://doi.org/10.1136/jmedgenet-2013-102012
Dornburg, A., Mallik, R., Wang, Z., Bernal, M.A., Thompson, B., Bruford, E.A., Nebert, D.W., Vasiliou, V., Yohe, L.R., Yoder, J.A., Townsend, J.P., 2022. Placing human gene families into their evolutionary context. Hum. Genomics 16, 56. https://doi.org/10.1186/s40246-022-00429-5
du Manoir, S., Speicher, M.R., Joos, S., Schröck, E., Popp, S., Döhner, H., Kovacs, G., Robert-Nicoud, M., Lichter, P., Cremer, T., 1993. Detection of complete and partial chromosome gains and losses by comparative genomic in situ hybridization. Hum. Genet. 90, 590–610. https://doi.org/10.1007/BF00202476
Ducray, F., Idbaih, A., de Reyniès, A., Bièche, I., Thillet, J., Mokhtari, K., Lair, S., Marie, Y., Paris, S., Vidaud, M., Hoang-Xuan, K., Delattre, O., Delattre, J.-Y.,

Sanson, M., 2008. Anaplastic oligodendrogliomas with 1p19q codeletion have a proneural gene expression profile. Mol. Cancer 7, 41. https://doi.org/10.1186/1476-4598-7-41

Eaton, W.W., Martins, S.S., Nestadt, G., Bienvenu, O.J., Clarke, D., Alexandre, P., 2008. The Burden of Mental Disorders. Epidemiol. Rev. 30, 1–14. https://doi.org/10.1093/epirev/mxn011

Ebbert, M.T.W., Farrugia, S.L., Sens, J.P., Jansen-West, K., Gendron, T.F., Prudencio, M., McLaughlin, I.J., Bowman, B., Seetin, M., DeJesus-Hernandez, M., Jackson, J., Brown, P.H., Dickson, D.W., van Blitterswijk, M., Rademakers, R., Petrucelli, L., Fryer, J.D., 2018. Long-read sequencing across the C9orf72 'GGGGCC' repeat expansion: implications for clinical use and genetic discovery efforts in human disease. Mol. Neurodegener. 13, 46. https://doi.org/10.1186/s13024-018-0274-4

Ebbert, M.T.W., Jensen, T.D., Jansen-West, K., Sens, J.P., Reddy, J.S., Ridge, P.G., Kauwe, J.S.K., Belzil, V., Pregent, L., Carrasquillo, M.M., Keene, D., Larson, E., Crane, P., Asmann, Y.W., Ertekin-Taner, N., Younkin, S.G., Ross, O.A., Rademakers, R., Petrucelli, L., Fryer, J.D., 2019. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. Genome Biol. 20, 97. https://doi.org/10.1186/s13059-019-1707-2

Eckel-Passow, J.E., Lachance, D.H., Molinaro, A.M., Walsh, K.M., Decker, P.A., Sicotte, H., Pekmezci, M., Rice, T., Kosel, M.L., Smirnov, I.V., Sarkar, G., Caron, A.A., Kollmeyer, T.M., Praska, C.E., Chada, A.R., Halder, C., Hansen, H.M., McCoy, L.S., Bracci, P.M., Marshall, R., Zheng, S., Reis, G.F., Pico, A.R., O'Neill, B.P., Buckner, J.C., Giannini, C., Huse, J.T., Perry, A., Tihan, T., Berger, M.S., Chang, S.M., Prados, M.D., Wiemels, J., Wiencke, J.K., Wrensch, M.R., Jenkins, R.B., 2015. Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. N. Engl. J. Med. 372, 2499–2508. https://doi.org/10.1056/NEJMoa1407279

Elbaz, B., Aaker, J.D., Isaac, S., Kolarzyk, A., Brugarolas, P., Eden, A., Popko, B., 2018. Phosphorylation State of ZFP24 Controls Oligodendrocyte Differentiation. Cell Rep. 23, 2254–2263. https://doi.org/10.1016/j.celrep.2018.04.089

El-Hattab, A.W., Smolarek, T.A., Walker, M.E., Schorry, E.K., Immken, L.L., Patel, G., Abbott, M.-A., Lanpher, B.C., Ou, Z., Kang, S.-H.L., Patel, A., Scaglia, F., Lupski, J.R., Cheung, S.W., Stankiewicz, P., 2009. Redefined genomic architecture in 15q24 directed by patient deletion/duplication breakpoint mapping. Hum. Genet. 126, 589–602. https://doi.org/10.1007/s00439-009-0706-x

Elston, G.N., Benavides-Piccione, R., DeFelipe, J., 2001. The Pyramidal Cell in Cognition: A Comparative Study in Human and Monkey. J. Neurosci. 21, RC163. https://doi.org/10.1523/JNEUROSCI.21-17-j0002.2001

Erady, C., Amin, K., Onilogbo, T.O.A.E., Tomasik, J., Jukes-Jones, R., Umrania, Y., Bahn, S., Prabakaran, S., 2022. Novel open reading frames in human accelerated regions and transposable elements reveal new leads to understand schizophrenia and bipolar disorder. Mol. Psychiatry 27, 1455–1468. https://doi.org/10.1038/s41380-021-01405-6

Faghihi, M.A., Modarresi, F., Khalil, A.M., Wood, D.E., Sahagan, B.G., Morgan, T.E., Finch, C.E., St. Laurent III, G., Kenny, P.J., Wahlestedt, C., 2008. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase. Nat. Med. 14, 723–730. https://doi.org/10.1038/nm1784

Feldmann, A., Amphornrat, J., Schönherr, M., Winterstein, C., Möbius, W., Ruhwedel, T., Danglot, L., Nave, K.-A., Galli, T., Bruns, D., Trotter, J., Krämer-Albers, E.-M., 2011. Transport of the Major Myelin Proteolipid Protein Is Directed by VAMP3 and VAMP7. J. Neurosci. 31, 5659–5672. https://doi.org/10.1523/JNEUROSCI.6638-10.2011

Fessel, J., 2022. Abnormal oligodendrocyte function in schizophrenia explains the long latent interval in some patients. Transl. Psychiatry 12, 1–8. https://doi.org/10.1038/s41398-022-01879-0

Fields, R.D., Araque, A., Johansen-Berg, H., Lim, S.-S., Lynch, G., Nave, K.-A., Nedergaard, M., Perez, R., Sejnowski, T., Wake, H., 2014. Glial biology in learning and cognition. Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry 20, 426–431. https://doi.org/10.1177/1073858413504465

Fischer, J., Fernández Ortuño, E., Marsoner, F., Artioli, A., Peters, J., Namba, T., Eugster Oegema, C., Huttner, W.B., Ladewig, J., Heide, M., 2022. Human-specific ARHGAP11B ensures human-like basal progenitor levels in hominid cerebral organoids. EMBO Rep. 23, e54728. https://doi.org/10.15252/embr.202254728

Fisher, H.L., Schreier, A., Zammit, S., Maughan, B., Munafò, M.R., Lewis, G., Wolke, D., 2013. Pathways between childhood victimization and psychosis-like symptoms in the ALSPAC birth cohort. Schizophr. Bull. 39, 1045–1055. https://doi.org/10.1093/schbul/sbs088

Florio, M., Albert, M., Taverna, E., Namba, T., Brandl, H., Lewitus, E., Haffner, C., Sykes, A., Wong, F.K., Peters, J., Guhr, E., Klemroth, S., Prüfer, K., Kelso, J., Naumann, R., Nüsslein, I., Dahl, A., Lachmann, R., Pääbo, S., Huttner, W.B., 2015. Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. Science 347, 1465–1470. https://doi.org/10.1126/science.aaa1975

Florio, M., Heide, M., Pinson, A., Brandl, H., Albert, M., Winkler, S., Wimberger, P., Huttner, W.B., Hiller, M., 2018. Evolution and cell-type specificity of human-specific genes preferentially expressed in progenitors of fetal neocortex. eLife 7, e32332. https://doi.org/10.7554/eLife.32332

Florio, M., Namba, T., Pääbo, S., Hiller, M., Huttner, W.B., 2016. A single splice site mutation in human-specific ARHGAP11B causes basal progenitor amplification. Sci. Adv. 2, e1601941. https://doi.org/10.1126/sciadv.1601941

Fowler, A., 2022. DECoN: A Detection and Visualization Tool for Exonic Copy Number Variants, in: Ng, C., Piscuoglio, S. (Eds.), Variant Calling: Methods and Protocols, Methods in Molecular Biology. Springer US, New York, NY, pp. 77–88. https://doi.org/10.1007/978-1-0716-2293-3_6

Galovic, M., Ferreira-Atuesta, C., Abraira, L., Döhler, N., Sinka, L., Brigo, F., Bentes, C., Zelano, J., Koepp, M.J., 2021. Seizures and Epilepsy After Stroke: Epidemiology, Biomarkers and Management. Drugs Aging 38, 285–299. https://doi.org/10.1007/s40266-021-00837-7

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., Schultz, N., 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. 6, pl1. https://doi.org/10.1126/scisignal.2004088

GBD 2019 Diseases and Injuries Collaborators, 2020. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Lond. Engl. 396, 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9

Gehring, N.H., Roignant, J.-Y., 2021. Anything but Ordinary – Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. Trends Genet. 37, 355–372. https://doi.org/10.1016/j.tig.2020.10.008

Gekas, J., van den Berg, D.-G., Durand, A., Vallée, M., Wildschut, H.I.J., Bujold, E., Forest, J.-C., Rousseau, F., Reinharz, D., 2011. Rapid testing versus karyotyping in Down's syndrome screening: cost-effectiveness and detection of clinically significant chromosome abnormalities. Eur. J. Hum. Genet. 19, 3–9. https://doi.org/10.1038/ejhg.2010.138

Gettler, K., Levantovsky, R., Moscati, A., Giri, M., Wu, Y., Hsu, N.-Y., Chuang, L.-S., Sazonovs, A., Venkateswaran, S., Korie, U., Chasteau, C., Duerr, R.H., Silverberg, M.S., Snapper, S.B., Daly, M.J., McGovern, D.P., Brant, S.R., Kugathasan, S., Anderson, C.A., Itan, Y., Cho, J.H., 2021. Common and rare variant prediction and penetrance of IBD in a large, multi-ethnic, health system-based biobank cohort. Gastroenterology 160, 1546–1557. https://doi.org/10.1053/j.gastro.2020.12.034

Gianfrancesco, O., Bubb, V.J., Quinn, J.P., 2019. Treating the "E" in "G × E": Trauma-Informed Approaches and Psychological Therapy Interventions in Psychosis. Front. Psychiatry 10, 9. https://doi.org/10.3389/fpsyt.2019.00009

Giegé, R., 2008. Toward a more complete view of tRNA biology. Nat. Struct. Mol. Biol. 15, 1007–1014. https://doi.org/10.1038/nsmb.1498

Gill, B.J., Pisapia, D.J., Malone, H.R., Goldstein, H., Lei, L., Sonabend, A., Yun, J., Samanamud, J., Sims, J.S., Banu, M., Dovas, A., Teich, A.F., Sheth, S.A., McKhann, G.M., Sisti, M.B., Bruce, J.N., Sims, P.A., Canoll, P., 2014. MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma. Proc. Natl. Acad. Sci. U. S. A. 111, 12550–12555. https://doi.org/10.1073/pnas.1405839111

Girirajan, S., Brkanac, Z., Coe, B.P., Baker, C., Vives, L., Vu, T.H., Shafer, N., Bernier, R., Ferrero, G.B., Silengo, M., Warren, S.T., Moreno, C.S., Fichera, M., Romano, C., Raskind, W.H., Eichler, E.E., 2011. Relative burden of large CNVs on a range of neurodevelopmental phenotypes. PLoS Genet. 7, e1002334. https://doi.org/10.1371/journal.pgen.1002334

Girirajan, S., Rosenfeld, J.A., Coe, B.P., Parikh, S., Friedman, N., Goldstein, A., Filipink, R.A., McConnell, J.S., Angle, B., Meschino, W.S., Nezarati, M.M., Asamoah, A., Jackson, K.E., Gowans, G.C., Martin, J.A., Carmany, E.P., Stockton, D.W., Schnur, R.E., Penney, L.S., Martin, D.M., Raskin, S., Leppig, K., Thiese, H., Smith, R., Aberg, E., Niyazov, D.M., Escobar, L.F., El-Khechen, D., Johnson, K.D., Lebel, R.R., Siefkas, K., Ball, S., Shur, N., McGuire, M., Brasington, C.K., Spence, J.E., Martin, L.S., Clericuzio, C., Ballif, B.C., Shaffer, L.G., Eichler, E.E., 2012. Phenotypic Heterogeneity of

Genomic Disorders and Rare Copy-Number Variants. N. Engl. J. Med. 367, 1321–1331. https://doi.org/10.1056/NEJMoa1200395
Giussani, P., Prinetti, A., Tringali, C., 2021. The role of Sphingolipids in myelination and myelin stability and their involvement in childhood and adult demyelinating disorders. J. Neurochem. 156, 403–414. https://doi.org/10.1111/jnc.15133
Gladitz, J., Klink, B., Seifert, M., 2018. Network-based analysis of oligodendrogliomas predicts novel cancer gene candidates within the region of the 1p/19q co-deletion. Acta Neuropathol. Commun. 6, 49. https://doi.org/10.1186/s40478-018-0544-y
Glasser, M.F., Goyal, M.S., Preuss, T.M., Raichle, M.E., Van Essen, D.C., 2014. Trends and properties of human cerebral cortex: Correlations with cortical myelin content. NeuroImage, In-vivo Brodmann Mapping of the Human Brain 93, 165–175. https://doi.org/10.1016/j.neuroimage.2013.03.060
Gonen-Yaacovi, G., de Souza, L.C., Levy, R., Urbanski, M., Josse, G., Volle, E., 2013. Rostral and caudal prefrontal contribution to creativity: a meta-analysis of functional imaging data. Front. Hum. Neurosci. 7, 465. https://doi.org/10.3389/fnhum.2013.00465
Gong, C., Maquat, L.E., 2011. lncRNAs transactivate Staufen1-mediated mRNA decay by duplexing with 3'UTRs via Alu elements. Nature 470, 284. https://doi.org/10.1038/nature09701
Gormley, M., Ona, K., Kapidzic, M., Garrido-Gomez, T., Zdravkovic, T., Fisher, S.J., 2017. Preeclampsia: novel insights from global RNA profiling of trophoblast subpopulations. Am. J. Obstet. Gynecol. 217, 200.e1-200.e17. https://doi.org/10.1016/j.ajog.2017.03.017
Grotheer, M., Rosenke, M., Wu, H., Kular, H., Querdasi, F.R., Natu, V.S., Yeatman, J.D., Grill-Spector, K., 2022. White matter myelination during early infancy is linked to spatial gradients and myelin content at birth. Nat. Commun. 13, 997. https://doi.org/10.1038/s41467-022-28326-4
Grotzinger, A.D., Mallard, T.T., Akingbuwa, W.A., Ip, H.F., Adams, M.J., Lewis, C.M., McIntosh, A.M., Grove, J., Dalsgaard, S., Lesch, K.-P., Strom, N., Meier, S.M., Mattheisen, M., Børglum, A.D., Mors, O., Breen, G., iPSYCH, Tourette Syndrome and Obsessive Compulsive Disorder Working Group of the Psychiatric Genetics Consortium, Bipolar Disorder Working Group of the Psychiatric Genetics Consortium, Major Depressive Disorder Working Group of the Psychiatric Genetics Consortium, Schizophrenia Working Group of the Psychiatric Genetics Consortium, Lee, P.H., Kendler, K.S., Smoller, J.W., Tucker-Drob, E.M., Nivard, M.G., 2022. Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. Nat. Genet. 54, 548–559. https://doi.org/10.1038/s41588-022-01057-4
Grove, W.M., Andreasen, N.C., 1982. Simultaneous tests of many hypotheses in exploratory research. J. Nerv. Ment. Dis. 170, 3–8. https://doi.org/10.1097/00005053-198201000-00002
Gulsuner, S., Stein, D.J., Susser, E.S., Sibeko, G., Pretorius, A., Walsh, T., Majara, L., Mndini, M.M., Mqulwana, S.G., Ntola, O.A., Casadei, S., Ngqengelele, L.L., Korchina, V., van der Merwe, C., Malan, M., Fader, K.M., Feng, M.,

Willoughby, E., Muzny, D., Baldinger, A., Andrews, H.F., Gur, R.C., Gibbs, R.A., Zingela, Z., Nagdee, M., Ramesar, R.S., King, M.-C., McClellan, J.M., 2020. Genetics of schizophrenia in the South African Xhosa. Science 367, 569–573. https://doi.org/10.1126/science.aay8833

Gussow, A.B., Copeland, B.R., Dhindsa, R.S., Wang, Q., Petrovski, S., Majoros, W.H., Allen, A.S., Goldstein, D.B., 2017. Orion: Detecting regions of the human non-coding genome that are intolerant to variation using population genetics. PLOS ONE 12, e0181604. https://doi.org/10.1371/journal.pone.0181604

Gymrek, M., Willems, T., Guilmatre, A., Zeng, H., Markus, B., Georgiev, S., Daly, M.J., Price, A.L., Pritchard, J.K., Sharp, A.J., Erlich, Y., 2016. Abundant contribution of short tandem repeats to gene expression variation in humans. Nat. Genet. 48, 22–29. https://doi.org/10.1038/ng.3461

Hafner, A., Bulyk, M.L., Jambhekar, A., Lahav, G., 2019. The multiple mechanisms that regulate p53 activity and cell fate. Nat. Rev. Mol. Cell Biol. 20, 199–210. https://doi.org/10.1038/s41580-019-0110-x

Hahn, S.-J., Kim, S., Choi, Y.S., Lee, J., Kang, J., 2022. Prediction of type 2 diabetes using genome-wide polygenic risk score and metabolic profiles: A machine learning analysis of population-based 10-year prospective cohort study. eBioMedicine 86. https://doi.org/10.1016/j.ebiom.2022.104383

Haig, D., 2008. Placental Growth Hormone-Related Proteins and Prolactin-Related Proteins. Placenta, Placenta: Platform for Life 29, 36–41. https://doi.org/10.1016/j.placenta.2007.09.010

Haig, D., 1993. Genetic Conflicts in Human Pregnancy. Q. Rev. Biol. 68, 495–532. https://doi.org/10.1086/418300

Hall, L.L., Lawrence, J.B., 2010. XIST RNA and Architecture of the Inactive X Chromosome Implications for the Repeat Genome. Cold Spring Harb. Symp. Quant. Biol. 75, 345–356. https://doi.org/10.1101/sqb.2010.75.030

Halldorsson, B.V., Eggertsson, H.P., Moore, K.H.S., Hauswedell, H., Eiriksson, O., Ulfarsson, M.O., Palsson, G., Hardarson, M.T., Oddsson, A., Jensson, B.O., Kristmundsdottir, S., Sigurpalsdottir, B.D., Stefansson, O.A., Beyter, D., Holley, G., Tragante, V., Gylfason, A., Olason, P.I., Zink, F., Asgeirsdottir, M., Sverrisson, S.T., Sigurdsson, B., Gudjonsson, S.A., Sigurdsson, G.T., Halldorsson, G.H., Sveinbjornsson, G., Norland, K., Styrkarsdottir, U., Magnusdottir, D.N., Snorradottir, S., Kristinsson, K., Sobech, E., Jonsson, Helgi, Geirsson, A.J., Olafsson, I., Jonsson, P., Pedersen, O.B., Erikstrup, C., Brunak, S., Ostrowski, S.R., Thorleifsson, G., Jonsson, F., Melsted, P., Jonsdottir, I., Rafnar, T., Holm, H., Stefansson, H., Saemundsdottir, J., Gudbjartsson, D.F., Magnusson, O.T., Masson, G., Thorsteinsdottir, U., Helgason, A., Jonsson, Hakon, Sulem, P., Stefansson, K., 2022. The sequences of 150,119 genomes in the UK Biobank. Nature 607, 732–740. https://doi.org/10.1038/s41586-022-04965-x

Halldorsson, B.V., Palsson, G., Stefansson, O.A., Jonsson, H., Hardarson, M.T., Eggertsson, H.P., Gunnarsson, B., Oddsson, A., Halldorsson, G.H., Zink, F., Gudjonsson, S.A., Frigge, M.L., Thorleifsson, G., Sigurdsson, A., Stacey, S.N., Sulem, P., Masson, G., Helgason, A., Gudbjartsson, D.F., Thorsteinsdottir, U., Stefansson, K., 2019. Characterizing mutagenic effects of

recombination through a sequence-level genetic map. Science 363, eaau1043. https://doi.org/10.1126/science.aau1043
Hamid, A.A., Pettibone, J.R., Mabrouk, O.S., Hetrick, V.L., Schmidt, R., Vander Weele, C.M., Kennedy, R.T., Aragona, B.J., Berke, J.D., 2016. Mesolimbic dopamine signals the value of work. Nat. Neurosci. 19, 117–126. https://doi.org/10.1038/nn.4173
Han, L., Zhao, X., Benton, M.L., Perumal, T., Collins, R.L., Hoffman, G.E., Johnson, J.S., Sloofman, L., Wang, H.Z., Stone, M.R., CommonMind Consortium, Brennand, K.J., Brand, H., Sieberts, S.K., Marenco, S., Peters, M.A., Lipska, B.K., Roussos, P., Capra, J.A., Talkowski, M., Ruderfer, D.M., 2020. Functional annotation of rare structural variation in the human brain. Nat. Commun. 11, 2990. https://doi.org/10.1038/s41467-020-16736-1
Harbuz, R., Zouari, R., Pierre, V., Ben Khelifa, M., Kharouf, M., Coutton, C., Merdassi, G., Abada, F., Escoffier, J., Nikas, Y., Vialard, F., Koscinski, I., Triki, C., Sermondade, N., Schweitzer, T., Zhioua, A., Zhioua, F., Latrous, H., Halouani, L., Ouafi, M., Makni, M., Jouk, P.-S., Sèle, B., Hennebicq, S., Satre, V., Viville, S., Arnoult, C., Lunardi, J., Ray, P.F., 2011. A Recurrent Deletion of DPY19L2 Causes Infertility in Man by Blocking Sperm Head Elongation and Acrosome Formation. Am. J. Hum. Genet. 88, 351–361. https://doi.org/10.1016/j.ajhg.2011.02.007
Hattori, E., Liu, C., Badner, J.A., Bonner, T.I., Christian, S.L., Maheshwari, M., Detera-Wadleigh, S.D., Gibbs, R.A., Gershon, E.S., 2003. Polymorphisms at the G72/G30 Gene Locus, on 13q33, Are Associated with Bipolar Disorder in Two Independent Pedigree Series. Am. J. Hum. Genet. 72, 1131–1140.
Havdahl, A., Niarchou, M., Starnawska, A., Uddin, M., van der Merwe, C., Warrier, V., n.d. Genetic contributions to autism spectrum disorder. Psychol. Med. 51, 2260–2273. https://doi.org/10.1017/S0033291721000192
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., Chakravarty, M.M., Chapin, M., Chong, J., Dalley, R.A., Daly, B.D., Dang, C., Datta, S., Dee, N., Dolbeare, T.A., Faber, V., Feng, D., Fowler, D.R., Goldy, J., Gregor, B.W., Haradon, Z., Haynor, D.R., Hohmann, J.G., Horvath, S., Howard, R.E., Jeromin, A., Jochim, J.M., Kinnunen, M., Lau, C., Lazarz, E.T., Lee, C., Lemon, T.A., Li, L., Li, Y., Morris, J.A., Overly, C.C., Parker, P.D., Parry, S.E., Reding, M., Royall, J.J., Schulkin, J., Sequeira, P.A., Slaughterbeck, C.R., Smith, S.C., Sodt, A.J., Sunkin, S.M., Swanson, B.E., Vawter, M.P., Williams, D., Wohnoutka, P., Zielke, H.R., Geschwind, D.H., Hof, P.R., Smith, S.M., Koch, C., Grant, S.G.N., Jones, A.R., 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 489, 391–399. https://doi.org/10.1038/nature11405
He, D., Wang, J., Lu, Y., Deng, Y., Zhao, C., Xu, L., Chen, Y., Hu, Y.-C., Zhou, W., Lu, Q.R., 2017. lncRNA Functional Networks in Oligodendrocytes Reveal Stage-Specific Myelination Control by an lncOL1/Suz12 Complex in the CNS. Neuron 93, 362–378. https://doi.org/10.1016/j.neuron.2016.11.044
Heide, M., Haffner, C., Murayama, A., Kurotaki, Y., Shinohara, H., Okano, H., Sasaki, E., Huttner, W.B., 2020. Human-specific ARHGAP11B increases size

and folding of primate neocortex in the fetal marmoset. Science 369, 546–550. https://doi.org/10.1126/science.abb2401
Hennessey, B.A., Amabile, T.M., 2010. Creativity. Annu. Rev. Psychol. 61, 569–598. https://doi.org/10.1146/annurev.psych.093008.100416
Herbert, P.S., 1959. Creativity and mental illness. Psychiatr. Q. 33, 534–547. https://doi.org/10.1007/BF01563109
Hertz, J.B., Schultz-Larsen, P., 1983. Human placental lactogen, pregnancy-specific beta-1-glycoprotein and alpha-fetoprotein in serum in threatened abortion. Int. J. Gynecol. Obstet. 21, 111–117. https://doi.org/10.1016/0020-7292(83)90047-4
Higashi, K., Medic, G., Littlewood, K.J., Diez, T., Granström, O., De Hert, M., 2013. Medication adherence in schizophrenia: factors influencing adherence and consequences of nonadherence, a systematic literature review. Ther. Adv. Psychopharmacol. 3, 200–218. https://doi.org/10.1177/2045125312474019
Hillmer, A., Chawar, C., Sanger, S., D'Elia, A., Butt, M., Kapoor, R., Kapczinski, F., Thabane, L., Samaan, Z., 2021. Genetic basis of cannabis use: a systematic review. BMC Med. Genomics 14, 203. https://doi.org/10.1186/s12920-021-01035-5
Hjorthøj, C., Stürup, A.E., McGrath, J.J., Nordentoft, M., 2017. Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. Lancet Psychiatry 4, 295–301. https://doi.org/10.1016/S2215-0366(17)30078-0
Höftberger, R., Lassmann, H., 2018. Inflammatory demyelinating diseases of the central nervous system. Handb. Clin. Neurol. 145, 263–283. https://doi.org/10.1016/B978-0-12-802395-2.00019-5
Houston, A., Williams, J.M., Rovis, T.L., Shanley, D.K., O'Riordan, R.T., Kiely, P.A., Ball, M., Barry, O.P., Kelly, J., Fanning, A., MacSharry, J., Mandelboim, O., Singer, B.B., Jonjic, S., Moore, T., 2016. Pregnancy-specific glycoprotein expression in normal gastrointestinal tract and in tumors detected with novel monoclonal antibodies. mAbs 8, 491–500. https://doi.org/10.1080/19420862.2015.1134410
Howes, O.D., Kapur, S., 2009. The dopamine hypothesis of schizophrenia: version III--the final common pathway. Schizophr. Bull. 35, 549–562. https://doi.org/10.1093/schbul/sbp006
Howes, O.D., Onwordi, E.C., 2023. The synaptic hypothesis of schizophrenia version III: a master mechanism. Mol. Psychiatry. https://doi.org/10.1038/s41380-023-02043-w
Howrigan, D.P., Rose, S.A., Samocha, K.E., Fromer, M., Cerrato, F., Chen, W.J., Churchhouse, C., Chambert, K., Chandler, S.D., Daly, M.J., Dumont, A., Genovese, G., Hwu, H.-G., Laird, N., Kosmicki, J.A., Moran, J.L., Roe, C., Singh, T., Wang, S.-H., Faraone, S.V., Glatt, S.J., McCarroll, S.A., Tsuang, M., Neale, B.M., 2020. Exome sequencing in schizophrenia-affected parent–offspring trios reveals risk conferred by protein-coding de novo mutations. Nat. Neurosci. 23, 185–193. https://doi.org/10.1038/s41593-019-0564-3
Huang, Z., Du, Y., Wen, J., Lu, B., Zhao, Y., 2022. snoRNAs: functions and mechanisms in biological processes, and roles in tumor pathophysiology. Cell Death Discov. 8, 1–10. https://doi.org/10.1038/s41420-022-01056-8

Hujoel, M.L.A., Sherman, M.A., Barton, A.R., Mukamel, R.E., Sankaran, V.G., Terao, C., Loh, P.-R., 2022. Influences of rare copy-number variation on human complex traits. Cell 185, 4233-4248.e27. https://doi.org/10.1016/j.cell.2022.09.028

Hwang, R., Tiwari, A.K., Zai, C.C., Felsky, D., Remington, E., Wallace, T., Tong, R.P., Souza, R.P., Oh, G., Potkin, S.G., Lieberman, J.A., Meltzer, H.Y., Kennedy, J.L., 2012. Dopamine D4 and D5 receptor gene variant effects on clozapine response in schizophrenia: Replication and exploration. Prog. Neuropsychopharmacol. Biol. Psychiatry 37, 62–75. https://doi.org/10.1016/j.pnpbp.2011.11.018

Hyun, K., Jeon, J., Park, K., Kim, J., 2017. Writing, erasing and reading histone lysine methylations. Exp. Mol. Med. 49, e324–e324. https://doi.org/10.1038/emm.2017.11

Ianevski, A., Giri, A.K., Aittokallio, T., 2022. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. Nat. Commun. 13, 1246. https://doi.org/10.1038/s41467-022-28803-w

Ingason, A., Kirov, G., Giegling, I., Hansen, T., Isles, A.R., Jakobsen, K.D., Kristinsson, K.T., le Roux, L., Gustafsson, O., Craddock, N., Möller, H.-J., McQuillin, A., Muglia, P., Cichon, S., Rietschel, M., Ophoff, R.A., Djurovic, S., Andreassen, O.A., Pietiläinen, O.P.H., Peltonen, L., Dempster, E., Collier, D.A., St. Clair, D., Rasmussen, H.B., Glenthøj, B.Y., Kiemeney, L.A., Franke, B., Tosato, S., Bonetto, C., Saemundsen, E., Hreidarsson, S.J., Nöthen, M.M., Gurling, H., O'Donovan, M.C., Owen, M.J., Sigurdsson, E., Petursson, H., Stefansson, H., Rujescu, D., Stefansson, K., Werge, T., 2011a. Maternally Derived Microduplications at 15q11-q13: Implication of Imprinted Genes in Psychotic Illness. Am. J. Psychiatry 168, 408–417. https://doi.org/10.1176/appi.ajp.2010.09111660

Ingason, A., Rujescu, D., St Clair, M., 2011b. Copy number variations of chromosome 16p13.1 region associated with schizophrenia | Molecular Psychiatry [WWW Document]. URL https://www.nature.com/articles/mp2009101 (accessed 6.11.23).

Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J.S., Jackson, S.E., Wills, M.R., Weissman, J.S., 2014. Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. Cell Rep. 8, 1365–1379. https://doi.org/10.1016/j.celrep.2014.07.045

Išerić, H., Alkan, C., Hach, F., Numanagić, I., 2022. Fast characterization of segmental duplication structure in multiple genome assemblies. Algorithms Mol. Biol. AMB 17, 4. https://doi.org/10.1186/s13015-022-00210-2

Ishizuka, K., Yoshida, T., Kawabata, T., Imai, A., Mori, H., Kimura, H., Inada, T., Okahisa, Y., Egawa, J., Usami, M., Kushima, I., Morikawa, M., Okada, T., Ikeda, M., Branko, A., Mori, D., Someya, T., Iwata, N., Ozaki, N., 2020. Functional characterization of rare NRXN1 variants identified in autism spectrum disorders and schizophrenia. J. Neurodev. Disord. 12, 25. https://doi.org/10.1186/s11689-020-09325-2

Jääskeläinen, E., Juola, P., Hirvonen, N., McGrath, J.J., Saha, S., Isohanni, M., Veijola, J., Miettunen, J., 2013. A Systematic Review and Meta-Analysis of

Recovery in Schizophrenia. Schizophr. Bull. 39, 1296–1306. https://doi.org/10.1093/schbul/sbs130
Jacobs, P.A., Strong, J.A., 1959. A case of human intersexuality having a possible XXY sex-determining mechanism. Nature 183, 302–303. https://doi.org/10.1038/183302a0
Jain, A.K., Xi, Y., McCarthy, R., Allton, K., Akdemir, K.C., Patel, L.R., Aronow, B., Lin, C., Li, W., Yang, L., Barton, M.C., 2016. LncPRESS1 is a p53-regulated lncRNA that safeguards pluripotency by disrupting SIRT6 mediated de-acetylation of histone H3K56. Mol. Cell 64, 967–981. https://doi.org/10.1016/j.molcel.2016.10.039
Jamison, K.R., 1989. Mood Disorders and Patterns of Creativity in British Writers and Artists. Psychiatry 52, 125–134. https://doi.org/10.1080/00332747.1989.11024436
Jaubert, J., Verheyden, S., Genty, D., Soulier, M., Cheng, H., Blamart, D., Burlet, C., Camus, H., Delaby, S., Deldicque, D., Edwards, R.L., Ferrier, C., Lacrampe-Cuyaubère, F., Lévêque, F., Maksud, F., Mora, P., Muth, X., Régnier, É., Rouzaud, J.-N., Santos, F., 2016. Early Neanderthal constructions deep in Bruniquel Cave in southwestern France. Nature 534, 111–114. https://doi.org/10.1038/nature18291
Jeong, H., Mendizabal, I., Berto, S., Chatterjee, P., Layman, T., Usui, N., Toriumi, K., Douglas, C., Singh, D., Huh, I., Preuss, T.M., Konopka, G., Yi, S.V., 2021. Evolution of DNA methylation in the human brain. Nat. Commun. 12, 2021. https://doi.org/10.1038/s41467-021-21917-7
Jiang, Y., Oldridge, D.A., Diskin, S.J., Zhang, N.R., 2015. CODEX: a normalization and copy number variation detection method for whole exome sequencing. Nucleic Acids Res. 43, e39. https://doi.org/10.1093/nar/gku1363
Jiang, Z., Tang, H., Ventura, M., Cardone, M.F., Marques-Bonet, T., She, X., Pevzner, P.A., Eichler, E.E., 2007. Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. Nat. Genet. 39, 1361–1368. https://doi.org/10.1038/ng.2007.9
Jiménez, A.J., Domínguez-Pinos, M.-D., Guerra, M.M., Fernández-Llebrez, P., Pérez-Fígares, J.-M., 2014. Structure and function of the ependymal barrier and diseases associated with ependyma disruption. Tissue Barriers 2, e28426. https://doi.org/10.4161/tisb.28426
Jo, B.-S., Choi, S.S., 2015. Introns: The Functional Benefits of Introns in Genomes. Genomics Inform. 13, 112–118. https://doi.org/10.5808/GI.2015.13.4.112
Johansson, L.F., van Dijk, F., de Boer, E.N., van Dijk-Bos, K.K., Jongbloed, J.D.H., van der Hout, A.H., Westers, H., Sinke, R.J., Swertz, M.A., Sijmons, R.H., Sikkema-Raddatz, B., 2016. CoNVaDING: Single Exon Variation Detection in Targeted NGS Data. Hum. Mutat. 37, 457–464. https://doi.org/10.1002/humu.22969
Jonas, S., Izaurralde, E., 2015. Towards a molecular understanding of microRNA-mediated gene silencing. Nat. Rev. Genet. 16, 421–433. https://doi.org/10.1038/nrg3965
Joordens, J.C.A., d'Errico, F., Wesselingh, F.P., Munro, S., de Vos, J., Wallinga, J., Ankjærgaard, C., Reimann, T., Wijbrans, J.R., Kuiper, K.F., Mücher, H.J., Coqueugniot, H., Prié, V., Joosten, I., van Os, B., Schulp, A.S., Panuel, M.,

van der Haas, V., Lustenhouwer, W., Reijmer, J.J.G., Roebroeks, W., 2015. Homo erectus at Trinil on Java used shells for tool production and engraving. Nature 518, 228–231. https://doi.org/10.1038/nature13962
Jostins, L., Barrett, J.C., 2011. Genetic risk prediction in complex disease. Hum. Mol. Genet. 20, R182–R188. https://doi.org/10.1093/hmg/ddr378
Juda, A., 1949. The relationship between highest mental capacity and psychic abnormalities. Am. J. Psychiatry 106, 296–307. https://doi.org/10.1176/ajp.106.4.296
Kachar, B., Behar, T., Dubois-Dalcq, M., 1986. Cell shape and motility of oligodendrocytes cultured without neurons. Cell Tissue Res. 244, 27–38. https://doi.org/10.1007/BF00218378
Kallioniemi, A., Visakorpi, T., Karhu, R., Pinkel, D., Kallioniemi, O.P., 1996. Gene Copy Number Analysis by Fluorescence in Situ Hybridization and Comparative Genomic Hybridization. Methods San Diego Calif 9, 113–121. https://doi.org/10.1006/meth.1996.0015
Kalsner, L., Chamberlain, S.J., 2015. Prader-Willi, Angelman, and 15q11-q13 duplication syndromes. Pediatr. Clin. North Am. 62, 587–606. https://doi.org/10.1016/j.pcl.2015.03.004
Kammerer, R., Ballesteros, A., Bonsor, D., Warren, J., Williams, J.M., Moore, T., Dveksler, G., 2020. Equine pregnancy-specific glycoprotein CEACAM49 secreted by endometrial cup cells activates TGFB. Reproduction 160, 685–694. https://doi.org/10.1530/REP-20-0277
Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., Tammana, H., Gingeras, T.R., 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316, 1484–1488. https://doi.org/10.1126/science.1138341
Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Neale, B.M., Daly, M.J., MacArthur, D.G., 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443. https://doi.org/10.1038/s41586-020-2308-7
Karijolich, J., Yu, Y.-T., 2010. Spliceosomal snRNA modifications and their function. RNA Biol. 7, 192–204.

Karlsson, J.L., 1970. Genetic association of giftedness and creativity with schizophrenia. Hereditas 66, 177–181. https://doi.org/10.1111/j.1601-5223.1970.tb02343.x

Kato, M., Kawaguchi, T., Ishikawa, S., Umeda, T., Nakamichi, R., Shapero, M.H., Jones, K.W., Nakamura, Y., Aburatani, H., Tsunoda, T., 2010. Population-genetic nature of copy number variations in the human genome. Hum. Mol. Genet. 19, 761–773. https://doi.org/10.1093/hmg/ddp541

Katsel, P., Roussos, P., Fam, P., Khan, S., Tan, W., Hirose, T., Nakagawa, S., Pletnikov, M.V., Haroutunian, V., 2019. The expression of long noncoding RNA NEAT1 is reduced in schizophrenia and modulates oligodendrocytes transcription. Npj Schizophr. 5, 1–9. https://doi.org/10.1038/s41537-019-0071-2

Katsushima, K., Natsume, A., Ohka, F., Shinjo, K., Hatanaka, A., Ichimura, N., Sato, S., Takahashi, S., Kimura, H., Totoki, Y., Shibata, T., Naito, M., Kim, H.J., Miyata, K., Kataoka, K., Kondo, Y., 2016. Targeting the Notch-regulated non-coding RNA TUG1 for glioma treatment. Nat. Commun. 7, 13616. https://doi.org/10.1038/ncomms13616

Kawano, K., Ebisawa, M., Hase, K., Fukuda, S., Hijikata, A., Kawano, S., Date, Y., Tsuneda, S., Itoh, K., Ohno, H., 2007. Psg18 Is Specifically Expressed in Follicle-associated Epithelium. Cell Struct. Funct. 32, 115–126. https://doi.org/10.1247/csf.07014

Kegeles, L.S., Abi-Dargham, A., Frankle, W.G., Gil, R., Cooper, T.B., Slifstein, M., Hwang, D.-R., Huang, Y., Haber, S.N., Laruelle, M., 2010. Increased Synaptic Dopamine Function in Associative Regions of the Striatum in Schizophrenia. Arch. Gen. Psychiatry 67, 231–239. https://doi.org/10.1001/archgenpsychiatry.2010.10

Kellis, M., Wold, B., Snyder, M.P., Bernstein, B.E., Kundaje, A., Marinov, G.K., Ward, L.D., Birney, E., Crawford, G.E., Dekker, J., Dunham, I., Elnitski, L.L., Farnham, P.J., Feingold, E.A., Gerstein, M., Giddings, M.C., Gilbert, D.M., Gingeras, T.R., Green, E.D., Guigo, R., Hubbard, T., Kent, J., Lieb, J.D., Myers, R.M., Pazin, M.J., Ren, B., Stamatoyannopoulos, J.A., Weng, Z., White, K.P., Hardison, R.C., 2014. Defining functional DNA elements in the human genome. Proc. Natl. Acad. Sci. 111, 6131–6138. https://doi.org/10.1073/pnas.1318948111

Kelly, J.J.P., Blough, M.D., Stechishin, O.D.M., Chan, J.A.W., Beauchamp, D., Perizzolo, M., Demetrick, D.J., Steele, L., Auer, R.N., Hader, W.J., Westgate, M., Parney, I.F., Jenkins, R., Cairncross, J.G., Weiss, S., 2010. Oligodendroglioma cell lines containing t(1;19)(q10;p10). Neuro-Oncol. 12, 745–755. https://doi.org/10.1093/neuonc/noq031

Kelly, T.M., Daley, D.C., 2013. Integrated Treatment of Substance Use and Psychiatric Disorders. Soc. Work Public Health 28, 388–406. https://doi.org/10.1080/19371918.2013.774673

Kent, W.J., 2002. BLAT—The BLAST-Like Alignment Tool. Genome Res. 12, 656–664. https://doi.org/10.1101/gr.229202

Kerr, D.J., Yang, L., 2021. Personalising cancer medicine with prognostic markers. eBioMedicine 72, 103577. https://doi.org/10.1016/j.ebiom.2021.103577

Kessler, R.C., Chiu, W.T., Demler, O., Merikangas, K.R., Walters, E.E., 2005. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch. Gen. Psychiatry 62, 617–627. https://doi.org/10.1001/archpsyc.62.6.617

Kim, J.H., Kim, K.-H., Kim, H.J., Lee, J., Myung, S.C., 2015. Expression of Beta-Defensin 131 Promotes an Innate Immune Response in Human Prostate Epithelial Cells. PLOS ONE 10, e0144776. https://doi.org/10.1371/journal.pone.0144776

Kim, P.M., Lam, H.Y.K., Urban, A.E., Korbel, J.O., Affourtit, J., Grubert, F., Chen, X., Weissman, S., Snyder, M., Gerstein, M.B., 2008. Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. Genome Res. 18, 1865. https://doi.org/10.1101/gr.081422.108

Kirov, G., Rees, E., Walters, J.T.R., Escott-Price, V., Georgieva, L., Richards, A.L., Chambert, K.D., Davies, G., Legge, S.E., Moran, J.L., McCarroll, S.A., O'Donovan, M.C., Owen, M.J., 2014. The Penetrance of Copy Number Variations for Schizophrenia and Developmental Delay. Biol. Psychiatry 75, 378–385. https://doi.org/10.1016/j.biopsych.2013.07.022

Kirov, G., Rujescu, D., Ingason, A., Collier, D.A., O'Donovan, M.C., Owen, M.J., 2009. Neurexin 1 (NRXN1) Deletions in Schizophrenia. Schizophr. Bull. 35, 851–854. https://doi.org/10.1093/schbul/sbp079

Knowles, J.K., Batra, A., Xu, H., Monje, M., 2022. Adaptive and maladaptive myelination in health and disease. Nat. Rev. Neurol. 18, 735–746. https://doi.org/10.1038/s41582-022-00737-3

Kõks, S., Pfaff, A.L., Singleton, L.M., Bubb, V.J., Quinn, J.P., 2022. Non-reference genome transposable elements (TEs) have a significant impact on the progression of the Parkinson's disease. Exp. Biol. Med. 247, 1680–1690. https://doi.org/10.1177/15353702221117147

Konte, B., Walters, J.T.R., Rujescu, D., Legge, S.E., Pardiñas, A.F., Cohen, D., Pirmohamed, M., Tiihonen, J., Hartmann, A.M., Bogers, J.P., van der Weide, J., van der Weide, K., Putkonen, A., Repo-Tiihonen, E., Hallikainen, T., Silva, E., Ingimarsson, O., Sigurdsson, E., Kennedy, J.L., Sullivan, P.F., Rietschel, M., Breen, G., Stefansson, H., Stefansson, K., Collier, D.A., O'Donovan, M.C., Giegling, I., 2021. HLA-DQB1 6672G>C (rs113332494) is associated with clozapine-induced neutropenia and agranulocytosis in individuals of European ancestry. Transl. Psychiatry 11, 214. https://doi.org/10.1038/s41398-021-01322-w

Kotlicka-Antczak, M., Pawełczyk, A., Rabe-Jabłońska, J., Smigielski, J., Pawełczyk, T., 2014. Obstetrical complications and Apgar score in subjects at risk of psychosis. J. Psychiatr. Res. 48, 79–85. https://doi.org/10.1016/j.jpsychires.2013.10.004

Kozel, B.A., Barak, B., Kim, C.A., Mervis, C.B., Osborne, L.R., Porter, M., Pober, B.R., 2021. Williams syndrome. Nat. Rev. Dis. Primer 7, 42. https://doi.org/10.1038/s41572-021-00276-z

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton,

M.L., Wu, Y.-C., Pfenning, A.R., Wang, X., Claussnitzer, M., Yaping Liu, Coarfa, C., Alan Harris, R., Shoresh, N., Epstein, C.B., Gjoneska, E., Leung, D., Xie, W., David Hawkins, R., Lister, R., Hong, C., Gascard, P., Mungall, A.J., Moore, R., Chuah, E., Tam, A., Canfield, T.K., Scott Hansen, R., Kaul, R., Sabo, P.J., Bansal, M.S., Carles, A., Dixon, J.R., Farh, K.-H., Feizi, S., Karlic, R., Kim, A.-R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T.R., Neph, S.J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R.C., Siebenthall, K.T., Sinnott-Armstrong, N.A., Stevens, M., Thurman, R.E., Wu, J., Zhang, B., Zhou, X., Beaudet, A.E., Boyer, L.A., De Jager, P.L., Farnham, P.J., Fisher, S.J., Haussler, D., Jones, S.J.M., Li, W., Marra, M.A., McManus, M.T., Sunyaev, S., Thomson, J.A., Tlsty, T.D., Tsai, L.-H., Wang, W., Waterland, R.A., Zhang, M.Q., Chadwick, L.H., Bernstein, B.E., Costello, J.F., Ecker, J.R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J.A., Wang, T., Kellis, M., 2015. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330. https://doi.org/10.1038/nature14248

Lam, M., Chen, C.-Y., Li, Z., Martin, A.R., Bryois, J., Ma, Xixian, 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., Periyasamy, S., Takahashi, A., Xu, Z., Yu, H., Zhu, F., Chen, W.J., Faraone, S., Glatt, S.J., He, L., Hyman, S.E., Hwu, H.-G., McCarroll, S.A., Neale, B.M., Sklar, P., Wildenauer, D.B., Yu, X., Zhang, D., Mowry, B.J., Lee, J., Holmans, P., Xu, S., Sullivan, P.F., Ripke, S., O'Donovan, M.C., Daly, M.J., Qin, S., Sham, P., Iwata, N., Hong, K.S., Schwab, S.G., Yue, W., Tsuang, M., Liu, J., Ma, Xiancang, Kahn, R.S., Shi, Y., Huang, H., 2019. Comparative genetic architectures of schizophrenia in East Asian and European populations. Nat. Genet. 51, 1670–1678. https://doi.org/10.1038/s41588-019-0512-x

Lam, M., Takeo, K., Almeida, R.G., Cooper, M.H., Wu, K., Iyer, M., Kantarci, H., Zuchero, J.B., 2022. CNS myelination requires VAMP2/3-mediated membrane expansion in oligodendrocytes. Nat. Commun. 13, 5583. https://doi.org/10.1038/s41467-022-33200-4

Lamptey, R.N.L., Chaulagain, B., Trivedi, R., Gothwal, A., Layek, B., Singh, J., 2022. A Review of the Common Neurodegenerative Disorders: Current Therapeutic Approaches and the Potential Role of Nanotherapeutics. Int. J. Mol. Sci. 23, 1851. https://doi.org/10.3390/ijms23031851

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, Christina, Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R.,

Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.-F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.-C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G.R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F.A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.-P., Yeh, R.-F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, Christopher, Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Patrinos, A., Morgan, M.J., International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, C. for G.R., The Sanger Centre:, Washington University Genome Sequencing Center, US DOE Joint Genome Institute:, Baylor College of Medicine Human Genome Sequencing Center:, RIKEN Genomic Sciences Center:, Genoscope and CNRS UMR-8030:, Department of Genome Analysis, I. of M.B., GTC Sequencing Center:, Beijing Genomics Institute/Human Genome Center:, Multimegabase Sequencing Center, T.I. for S.B., Stanford Genome Technology Center:, University of Oklahoma's Advanced Center for Genome Technology:, Max Planck Institute for Molecular Genetics:, Cold Spring Harbor Laboratory, L.A.H.G.C., GBF—German Research Centre for Biotechnology:, *Genome Analysis Group (listed in alphabetical order, also includes individuals listed under other headings):, Scientific management: National Human Genome Research Institute, U.N.I. of H., Stanford Human Genome Center:, University

of Washington Genome Center:, Department of Molecular Biology, K.U.S. of M., University of Texas Southwestern Medical Center at Dallas:, Office of Science, U.D. of E., The Wellcome Trust:, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921. https://doi.org/10.1038/35057062

Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559. https://doi.org/10.1186/1471-2105-9-559

Lanktree, M.B., Johansen, C.T., Anand, S.S., Davis, A.D., Miller, R., Yusuf, S., Hegele, R.A., on behalf of the SHARE, and S.-A.I., 2010. Genetic variation in hyaluronan metabolism loci is associated with plasma plasminogen activator inhibitor-1 concentration. Blood 116, 2160–2163. https://doi.org/10.1182/blood-2010-02-269902

Lapointe, S., Perry, A., Butowski, N.A., 2018. Primary brain tumours in adults. The Lancet 392, 432–446. https://doi.org/10.1016/S0140-6736(18)30990-5

Laukoter, S., Pauler, F.M., Beattie, R., Amberg, N., Hansen, A.H., Streicher, C., Penz, T., Bock, C., Hippenmeyer, S., 2020. Cell-Type Specificity of Genomic Imprinting in Cerebral Cortex. Neuron 107, 1160-1179.e9. https://doi.org/10.1016/j.neuron.2020.06.031

Laursen, L.S., Chan, C.W., ffrench-Constant, C., 2009. An Integrin–Contactin Complex Regulates CNS Myelination by Differential Fyn Phosphorylation. J. Neurosci. 29, 9174–9185. https://doi.org/10.1523/JNEUROSCI.5942-08.2009

Lee, H., Thacker, S., Sarn, N., Dutta, R., Eng, C., 2019. Constitutional mislocalization of Pten drives precocious maturation in oligodendrocytes and aberrant myelination in model of autism spectrum disorder. Transl. Psychiatry 9, 1–12. https://doi.org/10.1038/s41398-018-0364-7

Lee, J.A., Carvalho, C.M.B., Lupski, J.R., 2007. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131, 1235–1247. https://doi.org/10.1016/j.cell.2007.11.037

Lee, K.Y., Joo, E.-J., Ji, Y.I., Kim, D.-H., Park, J., Chung, I.-W., Lee, S.I., Joo, Y.H., Ahn, Y.M., Song, J.Y., Kim, Y.S., 2011. Associations between DRDs and schizophrenia in a Korean population: multi-stage association analyses. Exp. Mol. Med. 43, 44–52. https://doi.org/10.3858/emm.2011.43.1.005

Lee, P.H., Anttila, V., Won, H., Feng, Y.-C.A., Rosenthal, J., Zhu, Z., Tucker-Drob, E.M., Nivard, M.G., Grotzinger, A.D., Posthuma, D., Wang, M.M.-J., Yu, D., Stahl, E.A., Walters, R.K., Anney, R.J.L., Duncan, L.E., Ge, T., Adolfsson, R., Banaschewski, T., Belangero, S., Cook, E.H., Coppola, G., Derks, E.M., Hoekstra, P.J., Kaprio, J., Keski-Rahkonen, A., Kirov, G., Kranzler, H.R., Luykx, J.J., Rohde, L.A., Zai, C.C., Agerbo, E., Arranz, M.J., Asherson, P., Bækvad-Hansen, M., Baldursson, G., Bellgrove, M., Belliveau, R.A., Buitelaar, J., Burton, C.L., Bybjerg-Grauholm, J., Casas, M., Cerrato, F., Chambert, K., Churchhouse, C., Cormand, B., Crosbie, J., Dalsgaard, S., Demontis, D., Doyle, A.E., Dumont, A., Elia, J., Grove, J., Gudmundsson, O.O., Haavik, J., Hakonarson, H., Hansen, C.S., Hartman, C.A., Hawi, Z., Hervás, A., Hougaard, D.M., Howrigan, D.P., Huang, H., Kuntsi, J., Langley, K., Lesch, K.-P., Leung, P.W.L., Loo, S.K., Martin, J., Martin, A.R., McGough, J.J., Medland, S.E., Moran, J.L., Mors, O., Mortensen, P.B., Oades,

R.D., Palmer, D.S., Pedersen, C.B., Pedersen, M.G., Peters, T., Poterba, T., Poulsen, J.B., Ramos-Quiroga, J.A., Reif, A., Ribasés, M., Rothenberger, A., Rovira, P., Sánchez-Mora, C., Satterstrom, F.K., Schachar, R., Artigas, M.S., Steinberg, S., Stefansson, H., Turley, P., Walters, G.B., Werge, T., Zayats, T., Arking, D.E., Bettella, F., Buxbaum, J.D., Christensen, J.H., Collins, R.L., Coon, H., De Rubeis, S., Delorme, R., Grice, D.E., Hansen, T.F., Holmans, P.A., Hope, S., Hultman, C.M., Klei, L., Ladd-Acosta, C., Magnusson, P., Nærland, T., Nyegaard, M., Pinto, D., Qvist, P., Rehnström, K., Reichenberg, A., Reichert, J., Roeder, K., Rouleau, G.A., Saemundsen, E., Sanders, S.J., Sandin, S., St Pourcain, B., Stefansson, K., Sutcliffe, J.S., Talkowski, M.E., Weiss, L.A., Willsey, A.J., Agartz, I., Akil, H., Albani, D., Alda, M., Als, T.D., Anjorin, A., Backlund, L., Bass, N., Bauer, M., Baune, B.T., Bellivier, F., Bergen, S.E., Berrettini, W.H., Biernacka, J.M., Blackwood, D.H.R., Bøen, E., Budde, M., Bunney, W., Burmeister, M., Byerley, W., Byrne, E.M., Cichon, S., Clarke, T.-K., Coleman, J.R.I., Craddock, N., Curtis, D., Czerski, P.M., Dale, A.M., Dalkner, N., Dannlowski, U., Degenhardt, F., Di Florio, A., Elvsåshagen, T., Etain, B., Fischer, S.B., Forstner, A.J., Forty, L., Frank, J., Frye, M., Fullerton, J.M., Gade, K., Gaspar, H.A., Gershon, E.S., Gill, M., Goes, F.S., Gordon, S.D., Gordon-Smith, K., Green, M.J., Greenwood, T.A., Grigoroiu-Serbanescu, M., Guzman-Parra, J., Hauser, J., Hautzinger, M., Heilbronner, U., Herms, S., Hoffmann, P., Holland, D., Jamain, S., Jones, I., Jones, L.A., Kandaswamy, R., Kelsoe, J.R., Kennedy, J.L., Joachim, O.K., Kittel-Schneider, S., Kogevinas, M., Koller, A.C., Lavebratt, C., Lewis, C.M., Li, Q.S., Lissowska, J., Loohuis, L.M.O., Lucae, S., Maaser, A., Malt, U.F., Martin, N.G., Martinsson, L., McElroy, S.L., McMahon, F.J., McQuillin, A., Melle, I., Metspalu, A., Millischer, V., Mitchell, P.B., Montgomery, G.W., Morken, G., Morris, D.W., Müller-Myhsok, B., Mullins, N., Myers, R.M., Nievergelt, C.M., Nordentoft, M., Adolfsson, A.N., Nöthen, M.M., Ophoff, R.A., Owen, M.J., Paciga, S.A., Pato, C.N., Pato, M.T., Perlis, R.H., Perry, A., Potash, J.B., Reinbold, C.S., Rietschel, M., Rivera, M., Roberson, M., Schalling, M., Schofield, P.R., Schulze, T.G., Scott, L.J., Serretti, A., Sigurdsson, E., Smeland, O.B., Stordal, E., Streit, F., Strohmaier, J., Thorgeirsson, T.E., Treutlein, J., Turecki, G., Vaaler, A.E., Vieta, E., Vincent, J.B., Wang, Y., Witt, S.H., Zandi, P., Adan, R.A.H., Alfredsson, L., Ando, T., Aschauer, H., Baker, J.H., Bencko, V., Bergen, A.W., Birgegård, A., Perica, V.B., Brandt, H., Burghardt, R., Carlberg, L., Cassina, M., Clementi, M., Courtet, P., Crawford, S., Crow, S., Crowley, J.J., Danner, U.N., Davis, O.S.P., Degortes, D., DeSocio, J.E., Dick, D.M., Dina, C., Docampo, E., Egberts, K., Ehrlich, S., Espeseth, T., Fernández-Aranda, F., Fichter, M.M., Foretova, L., Forzan, M., Gambaro, G., Giegling, I., Gonidakis, F., Gorwood, P., Mayora, M.G., Guo, Y., Halmi, K.A., Hatzikotoulas, K., Hebebrand, J., Helder, S.G., Herpertz-Dahlmann, B., Herzog, W., Hinney, A., Imgart, H., Jiménez-Murcia, S., Johnson, C., Jordan, J., Julià, A., Kaminská, D., Karhunen, L., Karwautz, A., Kas, M.J.H., Kaye, W.H., Kennedy, M.A., Kim, Y.-R., Klareskog, L., Klump, K.L., Knudsen, G.P.S., Landén, M., Le Hellard, S., Levitan, R.D., Li, D., Lichtenstein, P., Maj, M., Marsal, S., McDevitt, S., Mitchell, J., Monteleone, P., Monteleone, A.M., Munn-Chernoff, M.A.,

Nacmias, B., Navratilova, M., O'Toole, J.K., Padyukov, L., Pantel, J., Papezova, H., Rabionet, R., Raevuori, A., Ramoz, N., Reichborn-Kjennerud, T., Ricca, V., Roberts, M., Rujescu, D., Rybakowski, F., Scherag, A., Schmidt, U., Seitz, J., Slachtova, L., Slof-Op't Landt, M.C.T., Slopien, A., Sorbi, S., Southam, L., Strober, M., Tortorella, A., Tozzi, F., Treasure, J., Tziouvas, K., van Elburg, A.A., Wade, T.D., Wagner, G., Walton, E., Watson, H.J., Wichmann, H.-E., Woodside, D.B., Zeggini, E., Zerwas, S., Zipfel, S., Adams, M.J., Andlauer, T.F.M., Berger, K., Binder, E.B., Boomsma, D.I., Castelao, E., Colodro-Conde, L., Direk, N., Docherty, A.R., Domenici, E., Domschke, K., Dunn, E.C., Foo, J.C., de. Geus, E.J.C., Grabe, H.J., Hamilton, S.P., Horn, C., Hottenga, J.-J., Howard, D., Ising, M., Kloiber, S., Levinson, D.F., Lewis, G., Magnusson, P.K.E., Mbarek, H., Middeldorp, C.M., Mostafavi, S., Nyholt, D.R., Penninx, B.WJH., Peterson, R.E., Pistis, G., Porteous, D.J., Preisig, M., Quiroz, J.A., Schaefer, C., Schulte, E.C., Shi, J., Smith, D.J., Thomson, P.A., Tiemeier, H., Uher, R., van der Auwera, S., Weissman, M.M., Alexander, M., Begemann, M., Bramon, E., Buccola, N.G., Cairns, M.J., Campion, D., Carr, V.J., Cloninger, C.R., Cohen, D., Collier, D.A., Corvin, A., DeLisi, L.E., Donohoe, G., Dudbridge, F., Duan, J., Freedman, R., Gejman, P.V., Golimbet, V., Godard, S., Ehrenreich, H., Hartmann, A.M., Henskens, F.A., Ikeda, M., Iwata, N., Jablensky, A.V., Joa, I., Jönsson, E.G., Kelly, B.J., Knight, J., Konte, B., Laurent-Levinson, C., Lee, J., Lencz, T., Lerer, B., Loughland, C.M., Malhotra, A.K., Mallet, J., McDonald, C., Mitjans, M., Mowry, B.J., Murphy, K.C., Murray, R.M., O'Neill, F.A., Oh, S.-Y., Palotie, A., Pantelis, C., Pulver, A.E., Petryshen, T.L., Quested, D.J., Riley, B., Sanders, A.R., Schall, U., Schwab, S.G., Scott, R.J., Sham, P.C., Silverman, J.M., Sim, K., Steixner, A.A., Tooney, P.A., van Os, J., Vawter, M.P., Walsh, D., Weiser, M., Wildenauer, D.B., Williams, N.M., Wormley, B.K., Zhang, F., Androutsos, C., Arnold, P.D., Barr, C.L., Barta, C., Bey, K., Bienvenu, O.J., Black, D.W., Brown, L.W., Budman, C., Cath, D., Cheon, K.-A., Ciullo, V., Coffey, B.J., Cusi, D., Davis, L.K., Denys, D., Depienne, C., Dietrich, A., Eapen, V., Falkai, P., Fernandez, T.V., Garcia-Delgar, B., Geller, D.A., Gilbert, D.L., Grados, M.A., Greenberg, E., Grünblatt, E., Hagstrøm, J., Hanna, G.L., Hartmann, A., Hedderly, T., Heiman, G.A., Heyman, I., Hong, H.J., Huang, A., Huyser, C., Ibanez-Gomez, L., Khramtsova, E.A., Kim, Y.K., Kim, Y.-S., King, R.A., Koh, Y.-J., Konstantinidis, A., Kook, S., Kuperman, S., Leventhal, B.L., Lochner, C., Ludolph, A.G., Madruga-Garrido, M., Malaty, I., Maras, A., McCracken, J.T., Meijer, I.A., Mir, P., Morer, A., Müller-Vahl, K.R., Münchau, A., Murphy, T.L., Naarden, A., Nagy, P., Nestadt, G., Nestadt, P.S., Nicolini, H., Nurmi, E.L., Okun, M.S., Paschou, P., Piras, Fabrizio, Piras, Federica, Pittenger, C., Plessen, K.J., Richter, M.A., Rizzo, R., Robertson, M., Roessner, V., Ruhrmann, S., Samuels, J.F., Sandor, P., Schlögelhofer, M., Shin, E.-Y., Singer, H., Song, D.-H., Song, J., Spalletta, G., Stein, D.J., Stewart, S.E., Storch, E.A., Stranger, B., Stuhrmann, M., Tarnok, Z., Tischfield, J.A., Tübing, J., Visscher, F., Vulink, N., Wagner, M., Walitza, S., Wanderer, S., Woods, M., Worbe, Y., Zai, G., Zinner, S.H., Sullivan, P.F., Franke, B., Daly, M.J., Bulik, C.M., Lewis, C.M., McIntosh, A.M., O'Donovan, M.C., Zheutlin,

A., Andreassen, O.A., Børglum, A.D., Breen, G., Edenberg, H.J., Fanous, A.H., Faraone, S.V., Gelernter, J., Mathews, C.A., Mattheisen, M., Mitchell, K.S., Neale, M.C., Nurnberger, J.I., Ripke, S., Santangelo, S.L., Scharf, J.M., Stein, M.B., Thornton, L.M., Walters, J.T.R., Wray, N.R., Geschwind, D.H., Neale, B.M., Kendler, K.S., Smoller, J.W., 2019. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell 179, 1469-1482.e11. https://doi.org/10.1016/j.cell.2019.11.020

Legge, S.E., Santoro, M.L., Periyasamy, S., Okewole, A., Arsalan, A., Kowalec, K., 2021. Genetic architecture of schizophrenia: a review of major advancements. Psychol. Med. 51, 2168–2177. https://doi.org/10.1017/S0033291720005334

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.-H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J., MacArthur, D.G., 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291. https://doi.org/10.1038/nature19057

Lello, L., Avery, S.G., Tellier, L., Vazquez, A.I., de los Campos, G., Hsu, S.D.H., 2018. Accurate Genomic Prediction of Human Height. Genetics 210, 477–497. https://doi.org/10.1534/genetics.118.301267

Lesurf, R., Cotto, K.C., Wang, G., Griffith, M., Kasaian, K., Jones, S.J.M., Montgomery, S.B., Griffith, O.L., 2016. ORegAnno 3.0: a community-driven resource for curated regulatory annotation. Nucleic Acids Res. 44, D126–D132. https://doi.org/10.1093/nar/gkv1203

Levchenko, A., Gusev, F., Rogaev, E., 2023. The evolutionary origin of psychosis. Front. Psychiatry 14.

Lewis, C.M., Vassos, E., 2020. Polygenic risk scores: from research tools to clinical instruments. Genome Med. 12, 44. https://doi.org/10.1186/s13073-020-00742-5

Lewis, E.M.A., Kaushik, K., Sandoval, L.A., Antony, I., Dietmann, S., Kroll, K.L., 2021. Epigenetic regulation during human cortical development: Seq-ing answers from the brain to the organoid. Neurochem. Int. 147, 105039. https://doi.org/10.1016/j.neuint.2021.105039

Li, F., Lisi, E.C., Wohler, E.S., Hamosh, A., Batista, D.A.S., 2009. 3q29 interstitial microdeletion syndrome: An inherited case associated with cardiac defect and normal cognition. Eur. J. Med. Genet. 52, 349–352. https://doi.org/10.1016/j.ejmg.2009.05.001

Li, P., Snyder, G.L., Vanover, K.E., 2016. Dopamine Targeting Drugs for the Treatment of Schizophrenia: Past, Present and Future. Curr. Top. Med. Chem. 16, 3385–3403. https://doi.org/10.2174/1568026616666160608084834
Li, Y., Guo, B., Yang, R., Xiao, Z., Gao, X., Yu, J., Li, S., Luo, Y., 2018. A novel long noncoding RNA lnc158 promotes the differentiation of mouse neural precursor cells into oligodendrocytes by targeting nuclear factor-IB. NeuroReport 29, 1121. https://doi.org/10.1097/WNR.0000000000001083
Li, Y.R., Glessner, J.T., Coe, B.P., Li, J., Mohebnasab, M., Chang, X., Connolly, J., Kao, C., Wei, Z., Bradfield, J., Kim, C., Hou, C., Khan, M., Mentch, F., Qiu, H., Bakay, M., Cardinale, C., Lemma, M., Abrams, D., Bridglall-Jhingoor, A., Behr, M., Harrison, S., Otieno, G., Thomas, A., Wang, F., Chiavacci, R., Wu, L., Hadley, D., Goldmuntz, E., Elia, J., Maris, J., Grundmeier, R., Devoto, M., Keating, B., March, M., Pellagrino, R., Grant, S.F.A., Sleiman, P.M.A., Li, M., Eichler, E.E., Hakonarson, H., 2020. Rare copy number variants in over 100,000 European ancestry subjects reveal multiple disease associations. Nat. Commun. 11, 255. https://doi.org/10.1038/s41467-019-13624-1
Lipka, J., Kapitein, L.C., Jaworski, J., Hoogenraad, C.C., 2016. Microtubule-binding protein doublecortin-like kinase 1 (DCLK1) guides kinesin-3-mediated cargo transport to dendrites. EMBO J. 35, 302–318. https://doi.org/10.15252/embj.201592929
Lipovich, L., Dachet, F., Cai, J., Bagla, S., Balan, K., Jia, H., Loeb, J.A., 2012. Activity-Dependent Human Brain Coding/Noncoding Gene Regulatory Networks. Genetics 192, 1133–1148. https://doi.org/10.1534/genetics.112.145128
Liu, L., Fan, D., Ding, N., Hu, Y., Cai, G., Wang, L., Xin, L., Xia, Q., Li, X., Xu, S., Xu, J., Yang, X., Zou, Y., Pan, F., 2014. The relationship between DRD2 gene polymorphisms (C957T and C939T) and schizophrenia: A meta-analysis. Neurosci. Lett. 583, 43–48. https://doi.org/10.1016/j.neulet.2014.09.024
Liu, L., Zhang, Yang, Niu, G., Li, Q., Li, Z., Zhu, T., Feng, C., Liu, X., Zhang, Yuansheng, Xu, T., Chen, R., Teng, X., Zhang, R., Zou, D., Ma, L., Zhang, Z., 2022. BrainBase: a curated knowledgebase for brain diseases. Nucleic Acids Res. 50, D1131–D1138. https://doi.org/10.1093/nar/gkab987
Liu, P., Carvalho, C.M.B., Hastings, P.J., Lupski, J.R., 2012. Mechanisms for recurrent and complex human genomic rearrangements. Curr. Opin. Genet. Dev. 22, 211–220. https://doi.org/10.1016/j.gde.2012.02.012
Liu, P., Lacaria, M., Zhang, F., Withers, M., Hastings, P.J., Lupski, J.R., 2011. Frequency of Nonallelic Homologous Recombination Is Correlated with Length of Homology: Evidence that Ectopic Synapsis Precedes Ectopic Crossing-Over. Am. J. Hum. Genet. 89, 580–588. https://doi.org/10.1016/j.ajhg.2011.09.009
Liu, S.J., Nowakowski, T.J., Pollen, A.A., Lui, J.H., Horlbeck, M.A., Attenello, F.J., He, D., Weissman, J.S., Kriegstein, A.R., Diaz, A.A., Lim, D.A., 2016. Single-cell analysis of long non-coding RNAs in the developing human neocortex. Genome Biol. 17, 67. https://doi.org/10.1186/s13059-016-0932-1
Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., Fleming, J.,

Siminoff, L., Traino, H., Mosavel, M., Barker, L., Jewell, S., Rohrer, D., Maxim, D., Filkins, D., Harbach, P., Cortadillo, E., Berghuis, B., Turner, L., Hudson, E., Feenstra, K., Sobin, L., Robb, J., Branton, P., Korzeniewski, G., Shive, C., Tabor, D., Qi, L., Groch, K., Nampally, S., Buia, S., Zimmerman, A., Smith, A., Burges, R., Robinson, K., Valentino, K., Bradbury, D., Cosentino, M., Diaz-Mayoral, N., Kennedy, M., Engel, T., Williams, P., Erickson, K., Ardlie, K., Winckler, W., Getz, G., DeLuca, D., MacArthur, D., Kellis, M., Thomson, A., Young, T., Gelfand, E., Donovan, M., Meng, Y., Grant, G., Mash, D., Marcus, Y., Basile, M., Liu, J., Zhu, J., Tu, Z., Cox, N.J., Nicolae, D.L., Gamazon, E.R., Im, H.K., Konkashbaev, A., Pritchard, J., Stevens, M., Flutre, T., Wen, X., Dermitzakis, E.T., Lappalainen, T., Guigo, R., Monlong, J., Sammeth, M., Koller, D., Battle, A., Mostafavi, S., McCarthy, M., Rivas, M., Maller, J., Rusyn, I., Nobel, A., Wright, F., Shabalin, A., Feolo, M., Sharopova, N., Sturcke, A., Paschal, J., Anderson, J.M., Wilder, E.L., Derr, L.K., Green, E.D., Struewing, J.P., Temple, G., Volpi, S., Boyer, J.T., Thomson, E.J., Guyer, M.S., Ng, C., Abdallah, A., Colantuoni, D., Insel, T.R., Koester, S.E., Little, A.R., Bender, P.K., Lehner, T., Yao, Y., Compton, C.C., Vaught, J.B., Sawyer, S., Lockhart, N.C., Demchok, J., Moore, H.F., 2013. The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585. https://doi.org/10.1038/ng.2653

Lopez-Caraballo, L., Martorell-Marugan, J., Carmona-Sáez, P., Gonzalez-Munoz, E., 2020. iPS-Derived Early Oligodendrocyte Progenitor Cells from SPMS Patients Reveal Deficient In Vitro Cell Migration Stimulation. Cells 9, 1803. https://doi.org/10.3390/cells9081803

Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., Ellison, D.W., 2016. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol. (Berl.) 131, 803–820. https://doi.org/10.1007/s00401-016-1545-1

Lowther, C., Speevak, M., Armour, C.M., Goh, E.S., Graham, G.E., Li, C., Zeesman, S., Nowaczyk, M.J.M., Schultz, L.-A., Morra, A., Nicolson, R., Bikangaga, P., Samdup, D., Zaazou, M., Boyd, K., Jung, J.H., Siu, V., Rajguru, M., Goobie, S., Tarnopolsky, M.A., Prasad, C., Dick, P.T., Hussain, A.S., Walinga, M., Reijenga, R.G., Gazzellone, M., Lionel, A.C., Marshall, C.R., Scherer, S.W., Stavropoulos, D.J., McCready, E., Bassett, A.S., 2017. Molecular characterization of NRXN1 deletions from 19,263 clinical microarray cases identifies exons important for neurodevelopmental disease expression. Genet. Med. 19, 53–61. https://doi.org/10.1038/gim.2016.54

Ludwig, A.M., 1992. Creative Achievement and Psychopathology: Comparison among Professions. Am. J. Psychother. 46, 330–354. https://doi.org/10.1176/appi.psychotherapy.1992.46.3.330

Lv, L., Zhang, Y., Zhao, Yujia, Wei, Q., Zhao, Ye, Yi, Q., 2021. Effects of 1p/19q Codeletion on Immune Phenotype in Low Grade Glioma. Front. Cell. Neurosci. 15.

Lyngdoh, T., Bochud, M., Glaus, J., Castelao, E., Waeber, G., Vollenweider, P., Preisig, M., 2013. Associations of Serum Uric Acid and SLC2A9 Variant with

Depressive and Anxiety Disorders: A Population-Based Study. PLoS ONE 8, e76336. https://doi.org/10.1371/journal.pone.0076336
Ma, L., Chung, W.K., 2014. Quantitative Analysis of Copy Number Variants Based on Real-Time LightCycler PCR. Curr. Protoc. Hum. Genet. Editor. Board Jonathan Haines Al 80, 7.21.1-7.21.8. https://doi.org/10.1002/0471142905.hg0721s80
Maas, D.A., Eijsink, V.D., Spoelder, M., van Hulten, J.A., De Weerd, P., Homberg, J.R., Vallès, A., Nait-Oumesmar, B., Martens, G.J.M., 2020. Interneuron hypomyelination is associated with cognitive inflexibility in a rat model of schizophrenia. Nat. Commun. 11, 2329. https://doi.org/10.1038/s41467-020-16218-4
MacDonald, D.J., Scott, J.M., Gemmell, R.S., Mack, D.S., 1983. A prospective study of three biochemical fetoplacental tests: Serum human placental lactogen, pregnancy-specific β1-glycoprotein, and urinary estrogens, and their relationship to placental insufficiency. Am. J. Obstet. Gynecol. 147, 430–436. https://doi.org/10.1016/S0002-9378(16)32239-6
Maggiolini, F.A.M., Cantsilieris, S., D'Addabbo, P., Manganelli, M., Coe, B.P., Dumont, B.L., Sanders, A.D., Pang, A.W.C., Vollger, M.R., Palumbo, O., Palumbo, P., Accadia, M., Carella, M., Eichler, E.E., Antonacci, F., 2019. Genomic inversions and GOLGA core duplicons underlie disease instability at the 15q25 locus. PLOS Genet. 15, e1008075. https://doi.org/10.1371/journal.pgen.1008075
Mahmoud, A.D., Ballantyne, M.D., Miscianinov, V., Pinel, K., Hung, J., Scanlon, J.P., Iyinikkel, J., Kaczynski, J., Tavares, A.S., Bradshaw, A.C., Mills, N.L., Newby, D.E., Caporali, A., Gould, G.W., George, S.J., Ulitsky, I., Sluimer, J.C., Rodor, J., Baker, A.H., 2019. The Human-Specific and Smooth Muscle Cell-Enriched LncRNA SMILR Promotes Proliferation by Regulating Mitotic CENPF mRNA and Drives Cell-Cycle Progression Which Can Be Targeted to Limit Vascular Remodeling. Circ. Res. 125, 535–551. https://doi.org/10.1161/CIRCRESAHA.119.314876
Malhotra, D., Sebat, J., 2012. CNVs: Harbingers of a Rare Variant Revolution in Psychiatric Genetics. Cell 148, 1223–1241. https://doi.org/10.1016/j.cell.2012.02.039
Malnic, B., Godfrey, P.A., Buck, L.B., 2004. The human olfactory receptor gene family. Proc. Natl. Acad. Sci. 101, 2584–2589. https://doi.org/10.1073/pnas.0307882100
Mangan, R.J., Alsina, F.C., Mosti, F., Sotelo-Fonseca, J.E., Snellings, D.A., Au, E.H., Carvalho, J., Sathyan, L., Johnson, G.D., Reddy, T.E., Silver, D.L., Lowe, C.B., 2022. Adaptive sequence divergence forged new neurodevelopmental enhancers in humans. Cell 185, 4587-4603.e23. https://doi.org/10.1016/j.cell.2022.10.016
Manolio, T.A., Rodriguez, L.L., Brooks, L., Abecasis, G., Ballinger, D., Daly, M., Donnelly, P., Faraone, S.V., Frazer, K., Gabriel, S., Gejman, P., Guttmacher, A., Harris, E.L., Insel, T., Kelsoe, J.R., Lander, E., McCowin, N., Mailman, M.D., Nabel, E., Ostell, J., Pugh, E., Sherry, S., Sullivan, P.F., Thompson, J.F., Warram, J., Wholley, D., Milos, P.M., Collins, F.S., The GAIN Collaborative Research Group, the Collaborative Association Study of

Psoriasis, the International Multi-Center ADHD Genetics Project, the Molecular Genetics of Schizophrenia Collaboration, the Bipolar Genome Study, the Major Depression Stage 1 Genomewide Association in Population-Based Samples Study, the Genetics of Kidneys in Diabetes (GoKinD) Study, 2007. New models of collaboration in genome-wide association studies: the Genetic Association Information Network. Nat. Genet. 39, 1045–1051. https://doi.org/10.1038/ng2127

Marchionni, E., Méneret, A., Keren, B., Melki, J., Denier, C., Durr, A., Apartis, E., Boespflug-Tanguy, O., Mochel, F., 2019. KIF1C Variants Are Associated with Hypomyelination, Ataxia, Tremor, and Dystonia in Fraternal Twins. Tremor Hyperkinetic Mov. 9, 10.7916/tohm.v0.641. https://doi.org/10.7916/tohm.v0.641

Marshall, C.R., Howrigan, D.P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D.S., Antaki, D., Shetty, A., Holmans, P.A., Pinto, D., Gujral, M., Brandler, W.M., Malhotra, D., Wang, Z., Fajarado, K.V.F., Maile, M.S., Ripke, S., Agartz, I., Albus, M., Alexander, M., Amin, F., Atkins, J., Bacanu, S.A., Belliveau, R.A., Bergen, S.E., Bertalan, M., Bevilacqua, E., Bigdeli, T.B., Black, D.W., Bruggeman, R., Buccola, N.G., Buckner, R.L., Bulik-Sullivan, B., Byerley, W., Cahn, W., Cai, G., Cairns, M.J., Campion, D., Cantor, R.M., Carr, V.J., Carrera, N., Catts, S.V., Chambert, K.D., Cheng, W., Cloninger, C.R., Cohen, D., Cormican, P., Craddock, N., Crespo-Facorro, B., Crowley, J.J., Curtis, D., Davidson, M., Davis, K.L., Degenhardt, F., Del Favero, J., DeLisi, L.E., Dikeos, D., Dinan, T., Djurovic, S., Donohoe, G., Drapeau, E., Duan, J., Dudbridge, F., Eichhammer, P., Eriksson, J., Escott-Price, V., Essioux, L., Fanous, A.H., Farh, K.-H., Farrell, M.S., Frank, J., Franke, L., Freedman, R., Freimer, N.B., Friedman, J.I., Forstner, A.J., Fromer, M., Genovese, G., Georgieva, L., Gershon, E.S., Giegling, I., Giusti-Rodríguez, P., Godard, S., Goldstein, J.I., Gratten, J., de Haan, L., Hamshere, M.L., Hansen, M., Hansen, T., Haroutunian, V., Hartmann, A.M., Henskens, F.A., Herms, S., Hirschhorn, J.N., Hoffmann, P., Hofman, A., Huang, H., Ikeda, M., Joa, I., Kähler, A.K., Kahn, R.S., Kalaydjieva, L., Karjalainen, J., Kavanagh, D., Keller, M.C., Kelly, B.J., Kennedy, J.L., Kim, Y., Knowles, J.A., Konte, B., Laurent, C., Lee, P., Lee, S.H., Legge, S.E., Lerer, B., Levy, D.L., Liang, K.-Y., Lieberman, J., Lönnqvist, J., Loughland, C.M., Magnusson, P.K.E., Maher, B.S., Maier, W., Mallet, J., Mattheisen, M., Mattingsdal, M., McCarley, R.W., McDonald, C., McIntosh, A.M., Meier, S., Meijer, C.J., Melle, I., Mesholam-Gately, R.I., Metspalu, A., Michie, P.T., Milani, L., Milanova, V., Mokrab, Y., Morris, D.W., Müller-Myhsok, B., Murphy, K.C., Murray, R.M., Myin-Germeys, I., Nenadic, I., Nertney, D.A., Nestadt, G., Nicodemus, K.K., Nisenbaum, L., Nordin, A., O'Callaghan, E., O'Dushlaine, C., Oh, S.-Y., Olincy, A., Olsen, L., O'Neill, F.A., Van Os, J., Pantelis, C., Papadimitriou, G.N., Parkhomenko, E., Pato, M.T., Paunio, T., Perkins, D.O., Pers, T.H., Pietiläinen, O., Pimm, J., Pocklington, A.J., Powell, J., Price, A., Pulver, A.E., Purcell, S.M., Quested, D., Rasmussen, H.B., Reichenberg, A., Reimers, M.A., Richards, A.L., Roffman, J.L., Roussos, P., Ruderfer, D.M., Salomaa, V., Sanders, A.R., Savitz, A., Schall, U., Schulze, T.G., Schwab, S.G., Scolnick, E.M., Scott, R.J., Seidman, L.J., Shi, J., Silverman, J.M.,

Smoller, J.W., Söderman, E., Spencer, C.C.A., Stahl, E.A., Strengman, E., Strohmaier, J., Stroup, T.S., Suvisaari, J., Svrakic, D.M., Szatkiewicz, J.P., Thirumalai, S., Tooney, P.A., Veijola, J., Visscher, P.M., Waddington, J., Walsh, D., Webb, B.T., Weiser, M., Wildenauer, D.B., Williams, N.M., Williams, S., Witt, S.H., Wolen, A.R., Wormley, B.K., Wray, N.R., Wu, J.Q., Zai, C.C., Adolfsson, R., Andreassen, O.A., Blackwood, D.H.R., Bramon, E., Buxbaum, J.D., Cichon, S., Collier, D.A., Corvin, A., Daly, M.J., Darvasi, A., Domenici, E., Esko, T., Gejman, P.V., Gill, M., Gurling, H., Hultman, C.M., Iwata, N., Jablensky, A.V., Jönsson, E.G., Kendler, K.S., Kirov, G., Knight, J., Levinson, D.F., Li, Q.S., McCarroll, S.A., McQuillin, A., Moran, J.L., Mowry, B.J., Nöthen, M.M., Ophoff, R.A., Owen, M.J., Palotie, A., Pato, C.N., Petryshen, T.L., Posthuma, D., Rietschel, M., Riley, B.P., Rujescu, D., Sklar, P., St Clair, D., Walters, J.T.R., Werge, T., Sullivan, P.F., O'Donovan, M.C., Scherer, S.W., Neale, B.M., Sebat, J., 2017. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat. Genet. 49, 27–35. https://doi.org/10.1038/ng.3725

Marton, R.M., Miura, Y., Sloan, S.A., Li, Q., Revah, O., Levy, R.J., Huguenard, J.R., Pașca, S.P., 2019. Differentiation and maturation of oligodendrocytes in human three-dimensional neural cultures. Nat. Neurosci. 22, 484–491. https://doi.org/10.1038/s41593-018-0316-9

Mas-Ponte, D., Carlevaro-Fita, J., Palumbo, E., Hermoso Pulido, T., Guigo, R., Johnson, R., 2017. LncATLAS database for subcellular localization of long noncoding RNAs. RNA 23, 1080–1087. https://doi.org/10.1261/rna.060814.117

Mathews, J.C., Nadeem, S., Pouryahya, M., Belkhatir, Z., Deasy, J.O., Levine, A.J., Tannenbaum, A.R., 2020. Functional network analysis reveals an immune tolerance mechanism in cancer. Proc. Natl. Acad. Sci. 117, 16339–16345. https://doi.org/10.1073/pnas.2002179117

Mattick, J.S., 2001. Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep. 2, 986–991. https://doi.org/10.1093/embo-reports/kve230

Mattick, J.S., Amaral, P.P., Carninci, P., Carpenter, S., Chang, H.Y., Chen, L.-L., Chen, R., Dean, C., Dinger, M.E., Fitzgerald, K.A., Gingeras, T.R., Guttman, M., Hirose, T., Huarte, M., Johnson, R., Kanduri, C., Kapranov, P., Lawrence, J.B., Lee, J.T., Mendell, J.T., Mercer, T.R., Moore, K.J., Nakagawa, S., Rinn, J.L., Spector, D.L., Ulitsky, I., Wan, Y., Wilusz, J.E., Wu, M., 2023. Long non-coding RNAs: definitions, functions, challenges and recommendations. Nat. Rev. Mol. Cell Biol. 24, 430–447. https://doi.org/10.1038/s41580-022-00566-8

McCarthy, S., Makarov, V., Kirov, G., Addington, A., McClellan, J., Yoon, S., Perkins, D., Dickel, D.E., Kusenda, M., Krastoshevsky, O., Krause, V., Kumar, R.A., Grozeva, D., Malhotra, D., Walsh, T., Zackai, E.H., Kaplan, P., Ganesh, J., Krantz, I.D., Spinner, N.B., Roccanova, P., Bhandari, A., Pavon, K., Lakshmi, B., Leotta, A., Kendall, J., Lee, Y., Vacic, V., Gary, S., Iakoucheva, L., Crow, T.J., Christian, S.L., Lieberman, J., Stroup, S., Lehtimäki, T., Puura, K., Haldeman-Englert, C., Pearl, J., Goodell, M., Willour, V.L., DeRosse, P., Steele, J., Kassem, L., Wolff, J., Chitkara, N., McMahon, F.J., Malhotra, A.K., Potash, J.B., Schulze, T.G., Nöthen, M.M.,

Cichon, S., Rietschel, M., Leibenluft, E., Kustanovich, V., Lajonchere, C.M., Sutcliffe, J.S., Skuse, D., Gill, M., Gallagher, L., Mendell, N.R., Craddock, N., Owen, M.J., O'Donovan, M.C., Shaikh, T.H., Susser, E., DeLisi, L.E., Sullivan, P.F., Deutsch, C.K., Rapoport, J., Levy, D.L., King, M.-C., Sebat, J., 2009. Microduplications of 16p11.2 are Associated with Schizophrenia. Nat. Genet. 41, 1223–1227. https://doi.org/10.1038/ng.474

McDonald-McGinn, D.M., Sullivan, K.E., Marino, B., Philip, N., Swillen, A., Vorstman, J.A.S., Zackai, E.H., Emanuel, B.S., Vermeesch, J.R., Morrow, B.E., Scambler, P.J., Bassett, A.S., 2015. 22q11.2 deletion syndrome. Nat. Rev. Dis. Primer 1, 15071. https://doi.org/10.1038/nrdp.2015.71

McGrath, J., Saha, S., Chant, D., Welham, J., 2008. Schizophrenia: A Concise Overview of Incidence, Prevalence, and Mortality. Epidemiol. Rev. 30, 67–76. https://doi.org/10.1093/epirev/mxn001

McGrath, J.J., Hearle, J., Jenner, L., Plant, K., Drummond, A., Barkla, J.M., 1999. The fertility and fecundity of patients with psychoses. Acta Psychiatr. Scand. 99, 441–446. https://doi.org/10.1111/j.1600-0447.1999.tb00990.x

McKenzie, A.T., Wang, M., Hauberg, M.E., Fullard, J.F., Kozlenkov, A., Keenan, A., Hurd, Y.L., Dracheva, S., Casaccia, P., Roussos, P., Zhang, B., 2018. Brain Cell Type Specific Gene Expression and Co-expression Network Architectures. Sci. Rep. 8, 8868. https://doi.org/10.1038/s41598-018-27293-5

Meltzer, H.Y., 1997. Treatment-resistant schizophrenia--the role of clozapine. Curr. Med. Res. Opin. 14, 1–20. https://doi.org/10.1185/03007999709113338

Mercer, T.R., Qureshi, I.A., Gokhan, S., Dinger, M.E., Li, G., Mattick, J.S., Mehler, M.F., 2010. Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation. BMC Neurosci. 11, 14. https://doi.org/10.1186/1471-2202-11-14

Mighdoll, M.I., Tao, R., Kleinman, J.E., Hyde, T.M., 2015. Myelin, myelin-related disorders, and psychosis. Schizophr. Res. 161, 85–93. https://doi.org/10.1016/j.schres.2014.09.040

Miller, D.J., Duka, T., Stimpson, C.D., Schapiro, S.J., Baze, W.B., McArthur, M.J., Fobbs, A.J., Sousa, A.M.M., Šestan, N., Wildman, D.E., Lipovich, L., Kuzawa, C.W., Hof, P.R., Sherwood, C.C., 2012. Prolonged myelination in human neocortical evolution. Proc. Natl. Acad. Sci. 109, 16480–16485. https://doi.org/10.1073/pnas.1117943109

Miller, J.A., Ding, S.-L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Aiona, K., Arnold, J.M., Bennet, C., Bertagnolli, D., Brouner, K., Butler, S., Caldejon, S., Carey, A., Cuhaciyan, C., Dalley, R.A., Dee, N., Dolbeare, T.A., Facer, B.A.C., Feng, D., Fliss, T.P., Gee, G., Goldy, J., Gourley, L., Gregor, B.W., Gu, G., Howard, R.E., Jochim, J.M., Kuan, C.L., Lau, C., Lee, C.-K., Lee, F., Lemon, T.A., Lesnar, P., McMurray, B., Mastan, N., Mosqueda, N.F., Naluai-Cecchini, T., Ngo, N.-K., Nyhus, J., Oldre, A., Olson, E., Parente, J., Parker, P.D., Parry, S.E., Player, A.S., Pletikos, M., Reding, M., Royall, J.J., Roll, K., Sandman, D., Sarreal, M., Shapouri, S., Shapovalova, N.V., Shen, E.H., Sjoquist, N., Slaughterbeck, C.R., Smith, M., Sodt, A.J., Williams, D., Zöllei, L., Fischl, B., Gerstein, M.B., Geschwind, D.H., Glass, I.A., Hawrylycz, M.J., Hevner, R.F., Huang, H., Jones, A.R., Knowles, J.A., Levitt, P., Phillips, J.W., Sestan, N., Wohnoutka, P., Dang, C.,

Bernard, A., Hohmann, J.G., Lein, E.S., 2014. Transcriptional Landscape of the Prenatal Human Brain. Nature 508, 199–206. https://doi.org/10.1038/nature13185

Mills, J.D., Kavanagh, T., Kim, W.S., Chen, B.J., Waters, P.D., Halliday, G.M., Janitz, M., 2015. High expression of long intervening non-coding RNA OLMALINC in the human cortical white matter is associated with regulation of oligodendrocyte maturation. Mol. Brain 8, 2. https://doi.org/10.1186/s13041-014-0091-9

Missale, C., Nash, S.R., Robinson, S.W., Jaber, M., Caron, M.G., 1998. Dopamine Receptors: From Structure to Function. Physiol. Rev. 78, 189–225. https://doi.org/10.1152/physrev.1998.78.1.189

Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., Kaul, R., Halow, J., Van Nostrand, E.L., Freese, P., Gorkin, D.U., Shen, Y., He, Y., Mackiewicz, M., Pauli-Behn, F., Williams, B.A., Mortazavi, A., Keller, C.A., Zhang, X.-O., Elhajjajy, S.I., Huey, J., Dickel, D.E., Snetkova, V., Wei, X., Wang, X., Rivera-Mulia, J.C., Rozowsky, J., Zhang, Jing, Chhetri, S.B., Zhang, Jialing, Victorsen, A., White, K.P., Visel, A., Yeo, G.W., Burge, C.B., Lécuyer, E., Gilbert, D.M., Dekker, J., Rinn, J., Mendenhall, E.M., Ecker, J.R., Kellis, M., Klein, R.J., Noble, W.S., Kundaje, A., Guigó, R., Farnham, P.J., Cherry, J.M., Myers, R.M., Ren, B., Graveley, B.R., Gerstein, M.B., Pennacchio, L.A., Snyder, M.P., Bernstein, B.E., Wold, B., Hardison, R.C., Gingeras, T.R., Stamatoyannopoulos, J.A., Weng, Z., 2020. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583, 699–710. https://doi.org/10.1038/s41586-020-2493-4

Moore, T., 2012. Review: Parent-offspring conflict and the control of placental function. Placenta, Placenta –Predicting Future Health 33, S33–S36. https://doi.org/10.1016/j.placenta.2011.11.016

Moore, T., Dveksler, G.S., 2014. Pregnancy-specific glycoproteins: complex gene families regulating maternal-fetal interactions. Int. J. Dev. Biol. 58, 273–280. https://doi.org/10.1387/ijdb.130329gd

Moore, T., Williams, J.M., Becerra-Rodriguez, M.A., Dunne, M., Kammerer, R., Dveksler, G., 2022. Pregnancy-specific glycoproteins: evolution, expression, functions and disease associations. Reproduction 163, R11–R23. https://doi.org/10.1530/REP-21-0390

Moreno-Cabrera, J.M., del Valle, J., Castellanos, E., Feliubadaló, L., Pineda, M., Brunet, J., Serra, E., Capellà, G., Lázaro, C., Gel, B., 2020. Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. Eur. J. Hum. Genet. 28, 1645–1655. https://doi.org/10.1038/s41431-020-0675-z

Morgan, C., Kirkbride, J., Hutchinson, G., Craig, T., Morgan, K., Dazzan, P., Boydell, J., Doody, G.A., Jones, P.B., Murray, R.M., Leff, J., Fearon, P., 2008. Cumulative social disadvantage, ethnicity and first-episode psychosis: a case-control study. Psychol. Med. 38, 1701–1715. https://doi.org/10.1017/S0033291708004534

Morgan, C., Kirkbride, J., Leff, J., Craig, T., Hutchinson, G., McKenzie, K., Morgan, K., Dazzan, P., Doody, G.A., Jones, P., Murray, R., Fearon, P., 2007. Parental

separation, loss and psychosis in different ethnic groups: a case-control study. Psychol. Med. 37, 495–503. https://doi.org/10.1017/S0033291706009330
Mulle, J.G., Dodd, A.F., McGrath, J.A., Wolyniec, P.S., Mitchell, A.A., Shetty, A.C., Sobreira, N.L., Valle, D., Rudd, M.K., Satten, G., Cutler, D.J., Pulver, A.E., Warren, S.T., 2010. Microdeletions of 3q29 Confer High Risk for Schizophrenia. Am. J. Hum. Genet. 87, 229–236. https://doi.org/10.1016/j.ajhg.2010.07.013
Mulle, J.G., Pulver, A.E., McGrath, J.M., Wolyniec, P., Dodd, A.F., Cutler, D.J., Sebat, J., Malhotra, D., Nestadt, G., Conrad, D.F., Hurles, M., Barnes, C.P., Ikeda, M., Iwata, N., Levinson, D.F., Gejman, P.V., Sanders, A.R., Duan, J., Mitchell, A.A., Peter, I., Sklar, P., O'Dushlaine, C.T., Grozeva, D., O'Donovan, M.C., Owen, M.J., Hultman, C.M., Kähler, A.K., Sullivan, P.F., Kirov, G., Warren, S.T., 2014. Reciprocal duplication of the Williams-Beuren syndrome deletion on chromosome 7q11.23 is associated with schizophrenia. Biol. Psychiatry 75, 371–377. https://doi.org/10.1016/j.biopsych.2013.05.040
Nancy C. Andreasen, 1987. Creativity and mental illness: prevalence rates in writers and their first-degree relatives. Am. J. Psychiatry 144, 1288–1292. https://doi.org/10.1176/ajp.144.10.1288
Nanni, L., Ceri, S., Logie, C., 2020. Spatial patterns of CTCF sites define the anatomy of TADs and their boundaries. Genome Biol. 21, 197. https://doi.org/10.1186/s13059-020-02108-x
Necsulea, A., Kaessmann, H., 2014. Evolutionary dynamics of coding and non-coding transcriptomes. Nat. Rev. Genet. 15, 734–748. https://doi.org/10.1038/nrg3802
Nelson, B.R., Makarewich, C.A., Anderson, D.M., Winders, B.R., Troupes, C.D., Wu, F., Reese, A.L., McAnally, J.R., Chen, X., Kavalali, E.T., Cannon, S.C., Houser, S.R., Bassel-Duby, R., Olson, E.N., 2016. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science 351, 271. https://doi.org/10.1126/science.aad4076
Niarchou, M., Chawner, S.J.R.A., Doherty, J.L., Maillard, A.M., Jacquemont, S., Chung, W.K., Green-Snyder, L., Bernier, R.A., Goin-Kochel, R.P., Hanson, E., Linden, D.E.J., Linden, S.C., Raymond, F.L., Skuse, D., Hall, J., Owen, M.J., Bree, M.B.M. van den, 2019. Psychiatric disorders in children with 16p11.2 deletion and duplication. Transl. Psychiatry 9, 1–8. https://doi.org/10.1038/s41398-018-0339-8
Nuttle, X., Giannuzzi, G., Duyzend, M.H., Schraiber, J.G., Narvaiza, I., Sudmant, P.H., Penn, O., Chiatante, G., Malig, M., Huddleston, J., Benner, C., Camponeschi, F., Ciofi-Baffoni, S., Stessman, H.A.F., Marchetto, M.C.N., Denman, L., Harshman, L., Baker, C., Raja, A., Penewit, K., Janke, N., Tang, W.J., Ventura, M., Banci, L., Antonacci, F., Akey, J.M., Amemiya, C.T., Gage, F.H., Reymond, A., Eichler, E.E., 2016. Emergence of a Homo sapiens-specific gene family and chromosome 16p11.2 CNV susceptibility. Nature 536, 205–209. https://doi.org/10.1038/nature19075
Oldridge, D.A., Banerjee, S., Setlur, S.R., Sboner, A., Demichelis, F., 2010. Optimizing copy number variation analysis using genome-wide short sequence oligonucleotide arrays. Nucleic Acids Res. 38, 3275–3286. https://doi.org/10.1093/nar/gkq073

Ostroverkhova, D., Przytycka, T.M., Panchenko, A.R., 2023. Cancer driver mutations: predictions and reality. Trends Mol. Med. 0. https://doi.org/10.1016/j.molmed.2023.03.007
Oteyza, A.C., Battaloğlu, E., Ocek, L., Lindig, T., Reichbauer, J., Rebelo, A.P., Gonzalez, M.A., Zorlu, Y., Ozes, B., Timmann, D., Bender, B., Woehlke, G., Züchner, S., Schöls, L., Schüle, R., 2014. Motor protein mutations cause a new form of hereditary spastic paraplegia. Neurology 82, 2007–2016. https://doi.org/10.1212/WNL.0000000000000479
Ouyang, M., Dubois, J., Yu, Q., Mukherjee, P., Huang, H., 2019. Delineation of early brain development from fetuses to infants with diffusion MRI and beyond. NeuroImage 185, 836–850. https://doi.org/10.1016/j.neuroimage.2018.04.017
Owen, M.J., O'Donovan, M.C., Thapar, A., Craddock, N., 2011. Neurodevelopmental hypothesis of schizophrenia. Br. J. Psychiatry J. Ment. Sci. 198, 173–175. https://doi.org/10.1192/bjp.bp.110.084384
Owen, M.J., Sawa, A., Mortensen, P.B., 2016. Schizophrenia. The Lancet 388, 86–97. https://doi.org/10.1016/S0140-6736(15)01121-6
Owen, M.J., Williams, N.M., 2021. Explaining the missing heritability of psychiatric disorders. World Psychiatry 20, 294–295. https://doi.org/10.1002/wps.20870
Ozata, D.M., Gainetdinov, I., Zoch, A., O'Carroll, D., Zamore, P.D., 2019. PIWI-interacting RNAs: small RNAs with big functions. Nat. Rev. Genet. 20, 89–108. https://doi.org/10.1038/s41576-018-0073-3
Pantoja, I.E.M., Smith, M.D., Rajbhandari, L., Cheng, L., Gao, Y., Mahairaki, V., Venkatesan, A., Calabresi, P.A., Fitzgerald, K.C., Whartenby, K.A., 2020. iPSCs from people with MS can differentiate into oligodendrocytes in a homeostatic but not an inflammatory milieu. PLOS ONE 15, e0233980. https://doi.org/10.1371/journal.pone.0233980
Paolilo, R.B., Deiva, K., Neuteboom, R., Rostásy, K., Lim, M., 2020. Acute Disseminated Encephalomyelitis: Current Perspectives. Children 7, 210. https://doi.org/10.3390/children7110210
Paralkar, V.R., Mishra, T., Luan, J., Yao, Y., Kossenkov, A.V., Anderson, S.M., Dunagin, M., Pimkin, M., Gore, M., Sun, D., Konuthula, N., Raj, A., An, X., Mohandas, N., Bodine, D.M., Hardison, R.C., Weiss, M.J., 2014. Lineage and species-specific long noncoding RNAs during erythro-megakaryocytic development. Blood 123, 1927–1937. https://doi.org/10.1182/blood-2013-12-544494
Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., Han, J., Hubbard, L., Lynham, A., Mantripragada, K., Rees, E., MacCabe, J.H., McCarroll, S.A., Baune, B.T., Breen, G., Byrne, E.M., Dannlowski, U., Eley, T.C., Hayward, C., Martin, N.G., McIntosh, A.M., Plomin, R., Porteous, D.J., Wray, N.R., Caballero, A., Geschwind, D.H., Huckins, L.M., Ruderfer, D.M., Santiago, E., Sklar, P., Stahl, E.A., Won, H., Agerbo, E., Als, T.D., Andreassen, O.A., Bækvad-Hansen, M., Mortensen, P.B., Pedersen, C.B., Børglum, A.D., Bybjerg-Grauholm, J., Djurovic, S., Durmishi, N., Pedersen, M.G., Golimbet, V., Grove, J., Hougaard, D.M., Mattheisen, M., Molden, E., Mors, O., Nordentoft, M., Pejovic-Milovancevic, M., Sigurdsson, E., Silagadze, T., Hansen, C.S., Stefansson, K., Stefansson, H., Steinberg, S., Tosato, S., Werge,

T., Collier, D.A., Rujescu, D., Kirov, G., Owen, M.J., O'Donovan, M.C., Walters, J.T.R., 2018. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat. Genet. 50, 381–389. https://doi.org/10.1038/s41588-018-0059-2

Parks, M.M., Lawrence, C.E., Raphael, B.J., 2015. Detecting non-allelic homologous recombination from high-throughput sequencing data. Genome Biol. 16, 72. https://doi.org/10.1186/s13059-015-0633-1

Patel, K.R., Cherian, J., Gohil, K., Atkinson, D., 2014. Schizophrenia: Overview and Treatment Options. Pharm. Ther. 39, 638–645.

Peccarelli, M., Kebaara, B.W., 2014. Regulation of Natural mRNAs by the Nonsense-Mediated mRNA Decay Pathway. Eukaryot. Cell 13, 1126–1135. https://doi.org/10.1128/EC.00090-14

Pelvig, D.P., Pakkenberg, H., Stark, A.K., Pakkenberg, B., 2008. Neocortical glial cell numbers in human brains. Neurobiol. Aging 29, 1754–1762. https://doi.org/10.1016/j.neurobiolaging.2007.04.013

Periyasamy, S., John, S., Padmavati, R., Rajendren, P., Thirunavukkarasu, P., Gratten, J., Vinkhuyzen, A., McRae, A., Holliday, E.G., Nyholt, D.R., Nancarrow, D., Bakshi, A., Hemani, G., Nertney, D., Smith, H., Filippich, C., Patel, K., Fowdar, J., McLean, D., Tirupati, S., Nagasundaram, A., Gundugurti, P.R., Selvaraj, K., Jegadeesan, J., Jorde, L.B., Wray, N.R., Brown, M.A., Suetani, R., Giacomotto, J., Thara, R., Mowry, B.J., 2019. Association of Schizophrenia Risk With Disordered Niacin Metabolism in an Indian Genome-wide Association Study. JAMA Psychiatry 76, 1026–1034. https://doi.org/10.1001/jamapsychiatry.2019.1335

Pervez, M.T., Hasnain, M.J. ul, Abbas, S.H., Moustafa, M.F., Aslam, N., Shah, S.S.M., 2022. A Comprehensive Review of Performance of Next-Generation Sequencing Platforms. BioMed Res. Int. 2022, 3457806. https://doi.org/10.1155/2022/3457806

Petersen, C.C.H., 2007. The functional organization of the barrel cortex. Neuron 56, 339–355. https://doi.org/10.1016/j.neuron.2007.09.017

Peterson, R.E., Kuchenbaecker, K., Walters, R.K., Chen, C.-Y., Popejoy, A.B., Periyasamy, S., Lam, M., Iyegbe, C., Strawbridge, R.J., Brick, L., Carey, C., Martin, A., Meyers, J.L., Su, J., Chen, J., Edwards, A.C., Kalungi, A., Koen, N., Majara, L., Schwarz, E., Smoller, J., Stahl, E.A., Sullivan, P., Vassos, E., Mowry, B., Prieto, M., Cuellar-Barboza, A., Bigdeli, T.B., Edenberg, H.J., Huang, H., Duncan, L.E., 2019. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. Cell 179, 589–603. https://doi.org/10.1016/j.cell.2019.08.051

Phipps, E.A., Thadhani, R., Benzing, T., Karumanchi, S.A., 2019. Pre-eclampsia: pathogenesis, novel diagnostics and therapies. Nat. Rev. Nephrol. 15, 275–289. https://doi.org/10.1038/s41581-019-0119-6

Plana-Ripoll, O., Pedersen, C.B., Holtz, Y., Benros, M.E., Dalsgaard, S., de Jonge, P., Fan, C.C., Degenhardt, L., Ganna, A., Greve, A.N., Gunn, J., Iburg, K.M., Kessing, L.V., Lee, B.K., Lim, C.C.W., Mors, O., Nordentoft, M., Prior, A., Roest, A.M., Saha, S., Schork, A., Scott, J.G., Scott, K.M., Stedman, T., Sørensen, H.J., Werge, T., Whiteford, H.A., Laursen, T.M., Agerbo, E.,

Kessler, R.C., Mortensen, P.B., McGrath, J.J., 2019. Exploring Comorbidity Within Mental Disorders Among a Danish National Population. JAMA Psychiatry 76, 259–270. https://doi.org/10.1001/jamapsychiatry.2018.3658
Pocklington, A.J., Rees, E., Walters, J.T.R., Han, J., Kavanagh, D.H., Chambert, K.D., Holmans, P., Moran, J.L., McCarroll, S.A., Kirov, G., O'Donovan, M.C., Owen, M.J., 2015. Novel Findings from CNVs Implicate Inhibitory and Excitatory Signaling Complexes in Schizophrenia. Neuron 86, 1203–1214. https://doi.org/10.1016/j.neuron.2015.04.022
Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J., Pandolfi, P.P., 2010. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465, 1033–1038. https://doi.org/10.1038/nature09144
Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., Siepel, A., 2010. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 20, 110–121. https://doi.org/10.1101/gr.097857.109
Pompili, M., Amador, X.F., Girardi, P., Harkavy-Friedman, J., Harrow, M., Kaplan, K., Krausz, M., Lester, D., Meltzer, H.Y., Modestin, J., Montross, L.P., Bo Mortensen, P., Munk-Jørgensen, P., Nielsen, J., Nordentoft, M., Saarinen, P.I., Zisook, S., Wilson, S.T., Tatarelli, R., 2007. Suicide risk in schizophrenia: learning from the past to change the future. Ann. Gen. Psychiatry 6, 10. https://doi.org/10.1186/1744-859X-6-10
Ponchel, F., Toomes, C., Bransfield, K., Leong, F.T., Douglas, S.H., Field, S.L., Bell, S.M., Combaret, V., Puisieux, A., Mighell, A.J., Robinson, P.A., Inglehearn, C.F., Isaacs, J.D., Markham, A.F., 2003. Real-time PCR based on SYBR-Green I fluorescence: An alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions. BMC Biotechnol. 3, 18. https://doi.org/10.1186/1472-6750-3-18
Post, F., 1994. Creativity and Psychopathology a Study of 291 World-Famous Men. Br. J. Psychiatry 165, 22–34. https://doi.org/10.1192/bjp.165.1.22
Povysil, G., Tzika, A., Vogt, J., Haunschmid, V., Messiaen, L., Zschocke, J., Klambauer, G., Hochreiter, S., Wimmer, K., 2017. panelcn.MOPS: Copy-number detection in targeted NGS panel data for clinical diagnostics. Hum. Mutat. 38, 889–897. https://doi.org/10.1002/humu.23237
Power, R.A., Steinberg, S., Bjornsdottir, G., Rietveld, C.A., Abdellaoui, A., Nivard, M.M., Johannesson, M., Galesloot, T.E., Hottenga, J.J., Willemsen, G., Cesarini, D., Benjamin, D.J., Magnusson, P.K.E., Ullén, F., Tiemeier, H., Hofman, A., van Rooij, F.J.A., Walters, G.B., Sigurdsson, E., Thorgeirsson, T.E., Ingason, A., Helgason, A., Kong, A., Kiemeney, L.A., Koellinger, P., Boomsma, D.I., Gudbjartsson, D., Stefansson, H., Stefansson, K., 2015. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. Nat. Neurosci. 18, 953–955. https://doi.org/10.1038/nn.4040
Prodanov, T., Bansal, V., 2020. Sensitive alignment using paralogous sequence variants improves long-read mapping and variant calling in segmental duplications. Nucleic Acids Res. 48, e114. https://doi.org/10.1093/nar/gkaa829
Pronin, I.N., Holodny, A.I., Petraikin, A.V., 1997. MRI of high-grade glial tumors: correlation between the degree of contrast enhancement and the volume of

surrounding edema. Neuroradiology 39, 348–350. https://doi.org/10.1007/s002340050421

Pujana, M.A., Nadal, M., Gratacòs, M., Peral, B., Csiszar, K., González-Sarmiento, R., Sumoy, L., Estivill, X., 2001. Additional complexity on human chromosome 15q: identification of a set of newly recognized duplicons (LCR15) on 15q11-q13, 15q24, and 15q26. Genome Res. 11, 98–111. https://doi.org/10.1101/gr.155601

Pujana, M.A., Nadal, M., Guitart, M., Armengol, L., Gratacòs, M., Estivill, X., 2002. Human chromosome 15q11-q14 regions of rearrangements contain clusters of LCR15 duplicons. Eur. J. Hum. Genet. 10, 26–35. https://doi.org/10.1038/sj.ejhg.5200760

Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., Sklar, P., Purcell (Leader), S.M., Stone, J.L., Sullivan, P.F., Ruderfer, D.M., McQuillin, A., Morris, D.W., O'Dushlaine, C.T., Corvin, A., Holmans, P.A., O'Donovan, M.C., Sklar, P., Wray, N.R., Macgregor, S., Sklar, P., Sullivan, P.F., O'Donovan, M.C., Visscher, P.M., Gurling, H., Blackwood, D.H.R., Corvin, A., Craddock, N.J., Gill, M., Hultman, C.M., Kirov, G.K., Lichtenstein, P., McQuillin, A., Muir, W.J., O'Donovan, M.C., Owen, M.J., Pato, C.N., Purcell, S.M., Scolnick, E.M., St Clair, D., Stone, J.L., Sullivan, P.F., Sklar (Leader), P., O'Donovan, M.C., Kirov, G.K., Craddock, N.J., Holmans, P.A., Williams, N.M., Georgieva, L., Nikolov, I., Norton, N., Williams, H., Toncheva, D., Milanova, V., Owen, M.J., Hultman, C.M., Lichtenstein, P., Thelander, E.F., Sullivan, P., Morris, D.W., O'Dushlaine, C.T., Kenny, E., Quinn, E.M., Gill, M., Corvin, A., McQuillin, A., Choudhury, K., Datta, S., Pimm, J., Thirumalai, S., Puri, V., Krasucki, R., Lawrence, J., Quested, D., Bass, N., Gurling, H., Crombie, C., Fraser, G., Leh Kuan, S., Walker, N., St Clair, D., Blackwood, D.H.R., Muir, W.J., McGhee, K.A., Pickard, B., Malloy, P., Maclean, A.W., Van Beck, M., Wray, N.R., Macgregor, S., Visscher, P.M., Pato, M.T., Medeiros, H., Middleton, F., Carvalho, C., Morley, C., Fanous, A., Conti, D., Knowles, J.A., Paz Ferreira, C., Macedo, A., Helena Azevedo, M., Pato, C.N., Stone, J.L., Ruderfer, D.M., Kirby, A.N., Ferreira, M.A.R., Daly, M.J., Purcell, S.M., Sklar, P., Purcell, S.M., Stone, J.L., Chambert, K., Ruderfer, D.M., Kuruvilla, F., Gabriel, S.B., Ardlie, K., Moran, J.L., Daly, M.J., Scolnick, E.M., Sklar, P., The International Schizophrenia Consortium, Manuscript preparation, Data analysis, GWAS analysis subgroup, Polygene analyses subgroup, Management committee, Cardiff University, Karolinska Institutet/University of North Carolina at Chapel Hill, Trinity College Dublin, University College London, University of Aberdeen, University of Edinburgh, Queensland Institute of Medical Research, University of Southern California, Massachusetts General Hospital, Stanley Center for Psychiatric Research and Broad Institute of MIT and Harvard, 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460, 748–752. https://doi.org/10.1038/nature08185

Qaiser, F., Yin, Y., Mervis, C.B., Morris, C.A., Klein-Tasman, B.P., Tam, E., Osborne, L.R., Yuen, R.K.C., 2021. Rare and low frequency genomic variants

impacting neuronal functions modify the Dup7q11.23 phenotype. Orphanet J. Rare Dis. 16, 6. https://doi.org/10.1186/s13023-020-01648-6
Qin, X., Chen, J., Zhou, T., 2020. 22q11.2 deletion syndrome and schizophrenia. Acta Biochim. Biophys. Sin. 52, 1181–1190. https://doi.org/10.1093/abbs/gmaa113
Quintero-Rivera, F., Sharifi-Hannauer, P., Martinez-Agosto, J.A., 2010. Autistic and psychiatric findings associated with the 3q29 microdeletion syndrome: Case report and review. Am. J. Med. Genet. A. 152A, 2459–2467. https://doi.org/10.1002/ajmg.a.33573
Rajagopalan, R., Murrell, J.R., Luo, M., Conlin, L.K., 2020. A highly sensitive and specific workflow for detecting rare copy-number variants from exome sequencing data. Genome Med. 12, 14. https://doi.org/10.1186/s13073-020-0712-0
Rapoport, J.L., Giedd, J.N., Gogtay, N., 2012. Neurodevelopmental model of schizophrenia: update 2012. Mol. Psychiatry 17, 1228–1238. https://doi.org/10.1038/mp.2012.23
Rattila, S., Dunk, C.E., Im, M., Grichenko, O., Zhou, Y., Cohen, M., Yanez-Mo, M., Blois, S.M., Yamada, K.M., Erez, O., Gomez-Lopez, N., Lye, S.J., Hinz, B., Romero, R., Dveksler, G., 2019. Interaction of Pregnancy-Specific Glycoprotein 1 With Integrin α5β1 Is a Modulator of Extravillous Trophoblast Functions. Cells 8, 1369. https://doi.org/10.3390/cells8111369
Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., Korbel, J.O., 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinforma. Oxf. Engl. 28, i333–i339. https://doi.org/10.1093/bioinformatics/bts378
Rees, E., Creeth, H.D.J., Hwu, H.-G., Chen, W.J., Tsuang, M., Glatt, S.J., Rey, R., Kirov, G., Walters, J.T.R., Holmans, P., Owen, M.J., O'Donovan, M.C., 2021. Schizophrenia, autism spectrum disorders and developmental disorders share specific disruptive coding mutations. Nat. Commun. 12, 5353. https://doi.org/10.1038/s41467-021-25532-4
Rees, E., Han, J., Morgan, J., Carrera, N., Escott-Price, V., Pocklington, A.J., Duffield, M., Hall, L.S., Legge, S.E., Pardiñas, A.F., Richards, A.L., Roth, J., Lezheiko, T., Kondratyev, N., Kaleda, V., Golimbet, V., Parellada, M., González-Peñas, J., Arango, C., Gawlik, M., Kirov, G., Walters, J.T.R., Holmans, P., O'Donovan, M.C., Owen, M.J., 2020. De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. Nat. Neurosci. 23, 179–184. https://doi.org/10.1038/s41593-019-0565-2
Rees, E., Kirov, G., 2021. Copy number variation and neuropsychiatric illness. Curr. Opin. Genet. Dev., Molecular and genetic basis of disease 68, 57–63. https://doi.org/10.1016/j.gde.2021.02.014
Rees, E., Kirov, G., Sanders, A., Walters, J.T.R., Chambert, K.D., Shi, J., Szatkiewicz, J., O'Dushlaine, C., Richards, A.L., Green, E.K., Jones, I., Davies, G., Legge, S.E., Moran, J.L., Pato, C., Pato, M., Genovese, G., Levinson, D., Duan, J., Moy, W., Göring, H.H.H., Morris, D., Cormican, P., Kendler, K.S., O'Neill, F.A., Riley, B., Gill, M., Corvin, A., Wellcome Trust Case Control Consortium, Craddock, N., Sklar, P., Hultman, C., Sullivan, P.F., Gejman, P.V., McCarroll, S.A., O'Donovan, M.C., Owen, M.J., 2014a.

Evidence that duplications of 22q11.2 protect against schizophrenia. Mol. Psychiatry 19, 37–40. https://doi.org/10.1038/mp.2013.156
Rees, E., Walters, J.T.R., Chambert, K.D., O'Dushlaine, C., Szatkiewicz, J., Richards, A.L., Georgieva, L., Mahoney-Davies, G., Legge, S.E., Moran, J.L., Genovese, G., Levinson, D., Morris, D.W., Cormican, P., Kendler, K.S., O'Neill, F.A., Riley, B., Gill, M., Corvin, A., Wellcome Trust Case Control Consortium, Sklar, P., Hultman, C., Pato, C., Pato, M., Sullivan, P.F., Gejman, P.V., McCarroll, S.A., O'Donovan, M.C., Owen, M.J., Kirov, G., 2014c. CNV analysis in a large schizophrenia sample implicates deletions at 16p12.1 and SLC1A1 and duplications at 1p36.33 and CGNL1. Hum. Mol. Genet. 23, 1669–1676. https://doi.org/10.1093/hmg/ddt540
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., 2014b. Analysis of copy number variations at 15 schizophrenia-associated loci. Br. J. Psychiatry 204, 108–114. https://doi.org/10.1192/bjp.bp.113.131052
Riggs, E.R., Andersen, E.F., Cherry, A.M., Kantarci, S., Kearney, H., Patel, A., Raca, G., Ritter, D.I., South, S.T., Thorland, E.C., Pineda-Alvarez, D., Aradhya, S., Martin, C.L., 2020. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet. Med. 22, 245–257. https://doi.org/10.1038/s41436-019-0686-8
Rilling, J.K., van den Heuvel, M.P., 2018. Comparative Primate Connectomics. Brain. Behav. Evol. 91, 170–179. https://doi.org/10.1159/000488886
Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, H., Helms, J.A., Farnham, P.J., Segal, E., Chang, H.Y., 2007. Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Non-Coding RNAs. Cell 129, 1311–1323. https://doi.org/10.1016/j.cell.2007.05.022
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., Belliveau Jr, R.A., Bene, J., Bergen, S.E., Bevilacqua, E., Bigdeli, T.B., Black, D.W., Bruggeman, R., Buccola, N.G., Buckner, R.L., Byerley, W., Cahn, W., Cai, G., Campion, D., Cantor, R.M., Carr, V.J., Carrera, N., Catts, S.V., Chambert, K.D., Chan, R.C.K., Chen, R.Y.L., Chen, E.Y.H., Cheng, W., Cheung, E.F.C., Ann Chong, S., Robert Cloninger, C., Cohen, D., Cohen, N., Cormican, P., Craddock, N., Crowley, J.J., Curtis, D., Davidson, M., Davis, K.L., Degenhardt, F., Del Favero, J., Demontis, D., Dikeos, D., Dinan, T., Djurovic, S., Donohoe, G., Drapeau, E., Duan, J., Dudbridge, F., Durmishi, N., Eichhammer, P., Eriksson, J., Escott-Price, V., Essioux, L., Fanous, A.H., Farrell, M.S., Frank, J., Franke, L., Freedman, R., Freimer, N.B., Friedl, M., Friedman, J.I., Fromer, M., Genovese, G., Georgieva, L., Giegling, I., Giusti-Rodríguez, P., Godard, S., Goldstein, J.I., Golimbet, V., Gopal, S., Gratten, J., de Haan, L., Hammer, C., Hamshere, M.L., Hansen, M., Hansen, T., Haroutunian, V., Hartmann, A.M., Henskens, F.A., Herms, S., Hirschhorn,

J.N., Hoffmann, P., Hofman, A., Hollegaard, M.V., Hougaard, D.M., Ikeda, M., Joa, I., Julià, A., Kahn, R.S., Kalaydjieva, L., Karachanak-Yankova, S., Karjalainen, J., Kavanagh, D., Keller, M.C., Kennedy, J.L., Khrunin, A., Kim, Y., Klovins, J., Knowles, J.A., Konte, B., Kucinskas, V., Ausrele Kucinskiene, Z., Kuzelova-Ptackova, H., Kähler, A.K., Laurent, C., Lee Chee Keong, J., Hong Lee, S., Legge, S.E., Lerer, B., Li, M., Li, T., Liang, K.-Y., Lieberman, J., Limborska, S., Loughland, C.M., Lubinski, J., Lönnqvist, J., Macek Jr, M., Magnusson, P.K.E., Maher, B.S., Maier, W., Mallet, J., Marsal, S., Mattheisen, M., Mattingsdal, M., McCarley, R.W., McDonald, C., McIntosh, A.M., Meier, S., Meijer, C.J., Melegh, B., Melle, I., Mesholam-Gately, R.I., Metspalu, A., Michie, P.T., Milani, L., Milanova, V., Mokrab, Y., Morris, D.W., Mors, O., Murphy, K.C., Murray, R.M., Myin-Germeys, I., Müller-Myhsok, B., Nelis, M., Nenadic, I., Nertney, D.A., Nestadt, G., Nicodemus, K.K., Nikitina-Zake, L., Nisenbaum, L., Nordin, A., O'Callaghan, E., O'Dushlaine, C., O'Neill, F.A., Oh, S.-Y., Olincy, A., Olsen, L., Van Os, J., Pantelis, C., Papadimitriou, G.N., Papiol, S., Parkhomenko, E., Pato, M.T., Paunio, T., Pejovic-Milovancevic, M., Perkins, D.O., Pietiläinen, O., Pimm, J., Pocklington, A.J., Powell, J., Price, A., Pulver, A.E., Purcell, S.M., Quested, D., Rasmussen, H.B., Reichenberg, A., Reimers, M.A., Richards, A.L., Roffman, J.L., Roussos, P., Ruderfer, D.M., Salomaa, V., Sanders, A.R., Schall, U., Schubert, C.R., Schulze, T.G., Schwab, S.G., Scolnick, E.M., Scott, R.J., Seidman, L.J., Shi, J., Sigurdsson, E., Silagadze, T., Silverman, J.M., Sim, K., Slominsky, P., Smoller, J.W., So, H.-C., Spencer, ChrisC.A., Stahl, E.A., Stefansson, H., Steinberg, S., Stogmann, E., Straub, R.E., Strengman, E., Strohmaier, J., Scott Stroup, T., Subramaniam, M., Suvisaari, J., Svrakic, D.M., Szatkiewicz, J.P., Söderman, E., Thirumalai, S., Toncheva, D., Tosato, S., Veijola, J., Waddington, J., Walsh, D., Wang, D., Wang, Q., Webb, B.T., Weiser, M., Wildenauer, D.B., Williams, N.M., Williams, S., Witt, S.H., Wolen, A.R., Wong, E.H.M., Wormley, B.K., Simon Xi, H., Zai, C.C., Zheng, X., Zimprich, F., Wray, N.R., Stefansson, K., Visscher, P.M., Trust Case-Control Consortium, W., Adolfsson, R., Andreassen, O.A., Blackwood, D.H.R., Bramon, E., Buxbaum, J.D., Børglum, A.D., Cichon, S., Darvasi, A., Domenici, E., Ehrenreich, H., Esko, T., Gejman, P.V., Gill, M., Gurling, H., Hultman, C.M., Iwata, N., Jablensky, A.V., Jönsson, E.G., Kendler, K.S., Kirov, G., Knight, J., Lencz, T., Levinson, D.F., Li, Q.S., Liu, J., Malhotra, A.K., McCarroll, S.A., McQuillin, A., Moran, J.L., Mortensen, P.B., Mowry, B.J., Nöthen, M.M., Ophoff, R.A., Owen, M.J., Palotie, A., Pato, C.N., Petryshen, T.L., Posthuma, D., Rietschel, M., Riley, B.P., Rujescu, D., Sham, P.C., Sklar, P., St Clair, D., Weinberger, D.R., Wendland, J.R., Werge, T., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Psychosis Endophenotypes International Consortium, 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427. https://doi.org/10.1038/nature13595

Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11, R25. https://doi.org/10.1186/gb-2010-11-3-r25

Romero, C., Werme, J., Jansen, P.R., Gelernter, J., Stein, M.B., Levey, D., Polimanti, R., de Leeuw, C., Posthuma, D., Nagel, M., van der Sluis, S., 2022. Exploring the genetic overlap between twelve psychiatric disorders. Nat. Genet. 54, 1795–1802. https://doi.org/10.1038/s41588-022-01245-2

Rubio-Abadal, E., Ochoa, S., Barajas, A., Baños, I., Dolz, M., Sanchez, B., Del Cacho, N., Carlson, J., Huerta-Ramos, E., GENIPE Group, Usall, J., 2015. Birth weight and obstetric complications determine age at onset in first episode of psychosis. J. Psychiatr. Res. 65, 108–114. https://doi.org/10.1016/j.jpsychires.2015.03.018

Sadakierska-Chudy, A., Filip, M., 2015. A Comprehensive View of the Epigenetic Landscape. Part II: Histone Post-translational Modification, Nucleosome Level, and Chromatin Regulation by ncRNAs. Neurotox. Res. 27, 172–197. https://doi.org/10.1007/s12640-014-9508-6

Samonte, R.V., Eichler, E.E., 2002. Segmental duplications and the evolution of the primate genome. Nat. Rev. Genet. 3, 65–72. https://doi.org/10.1038/nrg705

Sanders, A.R., Duan, J., Levinson, D.F., Shi, J., He, D., Hou, C., Burrell, G.J., Rice, J.P., Nertney, D.A., Olincy, A., Rozic, P., Vinogradov, S., Buccola, N.G., Mowry, B.J., Freedman, R., Amin, F., Black, D.W., Silverman, J.M., Byerley, W.F., Crowe, R.R., Cloninger, C.R., Martinez, M., Gejman, P.V., 2008. No Significant Association of 14 Candidate Genes With Schizophrenia in a Large European Ancestry Sample: Implications for Psychiatric Genetics. Am. J. Psychiatry 165, 497–506. https://doi.org/10.1176/appi.ajp.2007.07101573

Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., Mason, C.E., Bilguvar, K., Celestino-Soper, P.B.S., Choi, M., Crawford, E.L., Davis, L., Wright, N.R.D., Dhodapkar, R.M., DiCola, M., DiLullo, N.M., Fernandez, T.V., Fielding-Singh, V., Fishman, D.O., Frahm, S., Garagaloyan, R., Goh, G.S., Kammela, S., Klei, L., Lowe, J.K., Lund, S.C., McGrew, A.D., Meyer, K.A., Moffat, W.J., Murdoch, J.D., O'Roak, B.J., Ober, G.T., Pottenger, R.S., Raubeson, M.J., Song, Y., Wang, Q., Yaspan, B.L., Yu, T.W., Yurkiewicz, I.R., Beaudet, A.L., Cantor, R.M., Curland, M., Grice, D.E., Günel, M., Lifton, R.P., Mane, S.M., Martin, D.M., Shaw, C.A., Sheldon, M., Tischfield, J.A., Walsh, C.A., Morrow, E.M., Ledbetter, D.H., Fombonne, E., Lord, C., Martin, C.L., Brooks, A.I., Sutcliffe, J.S., Cook, E.H., Geschwind, D., Roeder, K., Devlin, B., State, M.W., 2011. Multiple recurrent de novo copy number variations (CNVs), including duplications of the 7q11.23 Williams-Beuren syndrome region, are strongly associated with autism. Neuron 70, 863–885. https://doi.org/10.1016/j.neuron.2011.05.002

Sanyal, S., Van Tol, H.H.M., 1997. Review the role of dopamine D4 receptors in schizophrenia and antipsychotic action. J. Psychiatr. Res. 31, 219–232. https://doi.org/10.1016/S0022-3956(96)00039-8

Sasaki, Y.T.F., Ideue, T., Sano, M., Mituyama, T., Hirose, T., 2009. MENε/β noncoding RNAs are essential for structural integrity of nuclear paraspeckles. Proc. Natl. Acad. Sci. U. S. A. 106, 2525–2530. https://doi.org/10.1073/pnas.0807899106

Schaaf, C.P., Boone, P.M., Sampath, S., Williams, C., Bader, P.I., Mueller, J.M., Shchelochkov, O.A., Brown, C.W., Crawford, H.P., Phalen, J.A., Tartaglia,

N.R., Evans, P., Campbell, W.M., Chun-Hui Tsai, A., Parsley, L., Grayson, S.W., Scheuerle, A., Luzzi, C.D., Thomas, S.K., Eng, P.A., Kang, S.-H.L., Patel, A., Stankiewicz, P., Cheung, S.W., 2012. Phenotypic spectrum and genotype–phenotype correlations of NRXN1 exon deletions. Eur. J. Hum. Genet. 20, 1240–1247. https://doi.org/10.1038/ejhg.2012.95

Schaid, D.J., Chen, W., Larson, N.B., 2018. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat. Rev. Genet. 19, 491–504. https://doi.org/10.1038/s41576-018-0016-z

Scheepers, F.E., de Mul, J., Boer, F., Hoogendijk, W.J., 2018. Psychosis as an Evolutionary Adaptive Mechanism to Changing Environments. Front. Psychiatry 9.

Schmidt, E.R.E., Polleux, F., 2022. Genetic Mechanisms Underlying the Evolution of Connectivity in the Human Cortex. Front. Neural Circuits 15.

Schneider, M., Debbané, M., Bassett, A.S., Chow, E.W.C., Fung, W.L.A., van den Bree, M., Owen, M., Murphy, K.C., Niarchou, M., Kates, W.R., Antshel, K.M., Fremont, W., McDonald-McGinn, D.M., Gur, R.E., Zackai, E.H., Vorstman, J., Duijff, S.N., Klaassen, P.W.J., Swillen, A., Gothelf, D., Green, T., Weizman, A., Van Amelsvoort, T., Evers, L., Boot, E., Shashi, V., Hooper, S.R., Bearden, C.E., Jalbrzikowski, M., Armando, M., Vicari, S., Murphy, D.G., Ousley, O., Campbell, L.E., Simon, T.J., Eliez, S., International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome, 2014. Psychiatric disorders from childhood to adulthood in 22q11.2 deletion syndrome: results from the International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome. Am. J. Psychiatry 171, 627–639. https://doi.org/10.1176/appi.ajp.2013.13070864

Schoeler, T., Petros, N., Forti, M.D., Klamerus, E., Foglia, E., Murray, R., Bhattacharyya, S., 2017. Poor medication adherence and risk of relapse associated with continued cannabis use in patients with first-episode psychosis: a prospective analysis. Lancet Psychiatry 4, 627–633. https://doi.org/10.1016/S2215-0366(17)30233-X

Schoenfelder, S., Fraser, P., 2019. Long-range enhancer–promoter contacts in gene expression control. Nat. Rev. Genet. 20, 437–455. https://doi.org/10.1038/s41576-019-0128-0

Schöpflin, R., Melo, U.S., Moeinzadeh, H., Heller, D., Laupert, V., Hertzberg, J., Holtgrewe, M., Alavi, N., Klever, M.-K., Jungnitsch, J., Comak, E., Türkmen, S., Horn, D., Duffourd, Y., Faivre, L., Callier, P., Sanlaville, D., Zuffardi, O., Tenconi, R., Kurtas, N.E., Giglio, S., Prager, B., Latos-Bielenska, A., Vogel, I., Bugge, M., Tommerup, N., Spielmann, M., Vitobello, A., Kalscheuer, V.M., Vingron, M., Mundlos, S., 2022. Integration of Hi-C with short and long-read genome sequencing reveals the structure of germline rearranged genomes. Nat. Commun. 13, 6470. https://doi.org/10.1038/s41467-022-34053-7

Schwartz, S., Meshorer, E., Ast, G., 2009. Chromatin organization marks exon-intron structure. Nat. Struct. Mol. Biol. 16, 990–995. https://doi.org/10.1038/nsmb.1659

Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., Genovese, G., Rose,

S.A., Handsaker, R.E., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Daly, M.J., Carroll, M.C., Stevens, B., McCarroll, S.A., 2016. Schizophrenia risk from complex variation of complement component 4. Nature 530, 177–183. https://doi.org/10.1038/nature16549

Shaker, M.R., Pietrogrande, G., Martin, S., Lee, J.-H., Sun, W., Wolvetang, E.J., 2021. Rapid and Efficient Generation of Myelinating Human Oligodendrocytes in Organoids. Front. Cell. Neurosci. 15, 631548. https://doi.org/10.3389/fncel.2021.631548

Shanley, D.K., Kiely, P.A., Golla, K., Allen, S., Martin, K., O'Riordan, R.T., Ball, M., Aplin, J.D., Singer, B.B., Caplice, N., Moran, N., Moore, T., 2013. Pregnancy-Specific Glycoproteins Bind Integrin αIIbβ3 and Inhibit the Platelet—Fibrinogen Interaction. PLOS ONE 8, e57491. https://doi.org/10.1371/journal.pone.0057491

Sherry, S.T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., Sirotkin, K., 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29, 308–311.

Siddiqui, N., Roth, D., Toleikis, A., Zwetsloot, A.J., Cross, R.A., Straube, A., 2022. Force generation of KIF1C is impaired by pathogenic mutations. Curr. Biol. 32, 3862-3870.e6. https://doi.org/10.1016/j.cub.2022.07.029

Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., Barchas, J.D., Bass, N., Bigdeli, T.B., Breen, G., Bromet, E.J., Buckley, P.F., Bunney, W.E., Bybjerg-Grauholm, J., Byerley, W.F., Chapman, S.B., Chen, W.J., Churchhouse, C., Craddock, N., Cusick, C.M., DeLisi, L., Dodge, S., Escamilla, M.A., Eskelinen, S., Fanous, A.H., Faraone, S.V., Fiorentino, A., Francioli, L., Gabriel, S.B., Gage, D., Gagliano Taliun, S.A., Ganna, A., Genovese, G., Glahn, D.C., Grove, J., Hall, M.-H., Hämäläinen, E., Heyne, H.O., Holi, M., Hougaard, D.M., Howrigan, D.P., Huang, H., Hwu, H.-G., Kahn, R.S., Kang, H.M., Karczewski, K.J., Kirov, G., Knowles, J.A., Lee, F.S., Lehrer, D.S., Lescai, F., Malaspina, D., Marder, S.R., McCarroll, S.A., McIntosh, A.M., Medeiros, H., Milani, L., Morley, C.P., Morris, D.W., Mortensen, P.B., Myers, R.M., Nordentoft, M., O'Brien, N.L., Olivares, A.M., Ongur, D., Ouwehand, W.H., Palmer, D.S., Paunio, T., Quested, D., Rapaport, M.H., Rees, E., Rollins, B., Satterstrom, F.K., Schatzberg, A., Scolnick, E., Scott, L.J., Sharp, S.I., Sklar, P., Smoller, J.W., Sobell, J.L., Solomonson, M., Stahl, E.A., Stevens, C.R., Suvisaari, J., Tiao, G., Watson, S.J., Watts, N.A., Blackwood, D.H., Børglum, A.D., Cohen, B.M., Corvin, A.P., Esko, T., Freimer, N.B., Glatt, S.J., Hultman, C.M., McQuillin, A., Palotie, A., Pato, C.N., Pato, M.T., Pulver, A.E., St. Clair, D., Tsuang, M.T., Vawter, M.P., Walters, J.T., Werge, T.M., Ophoff, R.A., Sullivan, P.F., Owen, M.J., Boehnke, M., O'Donovan, M.C., Neale, B.M., Daly, M.J., 2022. Rare coding variants in ten genes confer substantial risk for schizophrenia. Nature 604, 509–516. https://doi.org/10.1038/s41586-022-04556-w

Singh, T., Walters, J.T.R., Johnstone, M., Curtis, D., Suvisaari, J., Torniainen, M., Rees, E., Iyegbe, C., Blackwood, D., McIntosh, A.M., Kirov, G., Geschwind, D., Murray, R.M., Di Forti, M., Bramon, E., Gandal, M., Hultman, C.M., Sklar, P., Palotie, A., Sullivan, P.F., O'Donovan, M.C., Owen, M.J., Barrett, J.C., 2017. The contribution of rare variants to risk of schizophrenia in

individuals with and without intellectual disability. Nat. Genet. 49, 1167–1173. https://doi.org/10.1038/ng.3903
Sloan, K.E., Warda, A.S., Sharma, S., Entian, K.-D., Lafontaine, D.L.J., Bohnsack, M.T., 2017. Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. RNA Biol. 14, 1138–1152. https://doi.org/10.1080/15476286.2016.1259781
Smoller, J.W., Andreassen, O.A., Edenberg, H.J., Faraone, S.V., Glatt, S.J., Kendler, K.S., 2019. Psychiatric genetics and the structure of psychopathology. Mol. Psychiatry 24, 409–420. https://doi.org/10.1038/s41380-017-0010-4
Sousa, A.M.M., Meyer, K.A., Santpere, G., Gulden, F.O., Sestan, N., 2017. Evolution of the Human Nervous System Function, Structure, and Development. Cell 170, 226–247. https://doi.org/10.1016/j.cell.2017.06.036
Spielmann, M., Klopocki, E., 2013. CNVs of noncoding cis-regulatory elements in human disease. Curr. Opin. Genet. Dev. 23, 249–256. https://doi.org/10.1016/j.gde.2013.02.013
Stankiewicz, P., Shaw, C.J., Withers, M., Inoue, K., Lupski, J.R., 2004. Serial segmental duplications during primate evolution result in complex human genome architecture. Genome Res. 14, 2209–2220. https://doi.org/10.1101/gr.2746604
Steadman, P.E., Xia, F., Ahmed, M., Mocle, A.J., Penning, A.R.A., Geraghty, A.C., Steenland, H.W., Monje, M., Josselyn, S.A., Frankland, P.W., 2020. Disruption of Oligodendrogenesis Impairs Memory Consolidation in Adult Mice. Neuron 105, 150-164.e6. https://doi.org/10.1016/j.neuron.2019.10.013
Stilo, S.A., Di Forti, M., Mondelli, V., Falcone, A.M., Russo, M., O'Connor, J., Palmer, E., Paparelli, A., Kolliakou, A., Sirianni, M., Taylor, H., Handley, R., Dazzan, P., Pariante, C., Marques, T.R., Zoccali, R., David, A., Murray, R.M., Morgan, C., 2013. Social disadvantage: cause or consequence of impending psychosis? Schizophr. Bull. 39, 1288–1295. https://doi.org/10.1093/schbul/sbs112
Stilo, S.A., Murray, R.M., 2019. Non-Genetic Factors in Schizophrenia. Curr. Psychiatry Rep. 21, 100. https://doi.org/10.1007/s11920-019-1091-3
Stolt, C.C., Schmitt, S., Lommes, P., Sock, E., Wegner, M., 2005. Impact of transcription factor Sox8 on oligodendrocyte specification in the mouse embryonic spinal cord. Dev. Biol. 281, 309–317. https://doi.org/10.1016/j.ydbio.2005.03.010
Stroud, H., Su, S.C., Hrvatin, S., Greben, A.W., Renthal, W., Boxer, L.D., Nagy, M.A., Hochbaum, D.R., Kinde, B., Gabel, H.W., Greenberg, M.E., 2017. Early-life gene expression in neurons modulates lasting epigenetic states. Cell 171, 1151-1164.e16. https://doi.org/10.1016/j.cell.2017.09.047
Suarez, B.K., Duan, J., Sanders, A.R., Hinrichs, A.L., Jin, C.H., Hou, C., Buccola, N.G., Hale, N., Weilbaecher, A.N., Nertney, D.A., Olincy, A., Green, S., Schaffer, A.W., Smith, C.J., Hannah, D.E., Rice, J.P., Cox, N.J., Martinez, M., Mowry, B.J., Amin, F., Silverman, J.M., Black, D.W., Byerley, W.F., Crowe, R.R., Freedman, R., Cloninger, C.R., Levinson, D.F., Gejman, P.V., 2006. Genomewide Linkage Scan of 409 European-Ancestry and African American Families with Schizophrenia: Suggestive Evidence of Linkage at 8p23.3-p21.2

and 11p13.1-q14.1 in the Combined Sample. Am. J. Hum. Genet. 78, 315–333. https://doi.org/10.1086/500272

Sudmant, P.H., Huddleston, J., Catacchio, C.R., Malig, M., Hillier, L.W., Baker, C., Mohajeri, K., Kondova, I., Bontrop, R.E., Persengiev, S., Antonacci, F., Ventura, M., Prado-Martinez, J., Great Ape Genome Project, Marques-Bonet, T., Eichler, E.E., 2013. Evolution and diversity of copy number variation in the great ape lineage. Genome Res. 23, 1373–1382. https://doi.org/10.1101/gr.158543.113

Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Hsi-Yang Fritz, M., Konkel, M.K., Malhotra, A., Stütz, A.M., Shi, X., Paolo Casale, F., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., Malig, M., Chaisson, M.J.P., Walter, K., Meiers, S., Kashin, S., Garrison, E., Auton, A., Lam, H.Y.K., Jasmine Mu, X., Alkan, C., Antaki, D., Bae, T., Cerveira, E., Chines, P., Chong, Z., Clarke, L., Dal, E., Ding, L., Emery, S., Fan, X., Gujral, M., Kahveci, F., Kidd, J.M., Kong, Y., Lameijer, E.-W., McCarthy, S., Flicek, P., Gibbs, R.A., Marth, G., Mason, C.E., Menelaou, A., Muzny, D.M., Nelson, B.J., Noor, A., Parrish, N.F., Pendleton, M., Quitadamo, A., Raeder, B., Schadt, E.E., Romanovitch, M., Schlattl, A., Sebra, R., Shabalin, A.A., Untergasser, A., Walker, J.A., Wang, M., Yu, F., Zhang, C., Zhang, J., Zheng-Bradley, X., Zhou, W., Zichner, T., Sebat, J., Batzer, M.A., McCarroll, S.A., Mills, R.E., Gerstein, M.B., Bashir, A., Stegle, O., Devine, S.E., Lee, C., Eichler, E.E., Korbel, J.O., 2015. An integrated map of structural variation in 2,504 human genomes. Nature 526, 75–81. https://doi.org/10.1038/nature15394

Sullivan, P.F., Daly, M.J., O'Donovan, M., 2012. Genetic Architectures of Psychiatric Disorders: The Emerging Picture and Its Implications. Nat. Rev. Genet. 13, 537–551. https://doi.org/10.1038/nrg3240

Sunahara, R.K., Guan, H.C., O'Dowd, B.F., Seeman, P., Laurier, L.G., Ng, G., George, S.R., Torchia, J., Van Tol, H.H., Niznik, H.B., 1991. Cloning of the gene for a human dopamine D5 receptor with higher affinity for dopamine than D1. Nature 350, 614–619. https://doi.org/10.1038/350614a0

Sybert, V.P., McCauley, E., 2004. Turner's Syndrome. N. Engl. J. Med. 351, 1227–1238. https://doi.org/10.1056/NEJMra030360

Talevich, E., Shain, A.H., Botton, T., Bastian, B.C., 2016. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput. Biol. 12, e1004873. https://doi.org/10.1371/journal.pcbi.1004873

Talkowski, M.E., Kirov, G., Bamne, M., Georgieva, L., Torres, G., Mansour, H., Chowdari, K.V., Milanova, V., Wood, J., McClain, L., Prasad, K., Shirts, B., Zhang, J., O'Donovan, M.C., Owen, M.J., Devlin, B., Nimgaonkar, V.L., 2008. A network of dopaminergic gene variations implicated as risk factors for schizophrenia. Hum. Mol. Genet. 17, 747–758. https://doi.org/10.1093/hmg/ddm347

Talkowski, M.E., Maussion, G., Crapper, L., Rosenfeld, J.A., Blumenthal, I., Hanscom, C., Chiang, C., Lindgren, A., Pereira, S., Ruderfer, D., Diallo, A.B., Lopez, J.P., Turecki, G., Chen, E.S., Gigek, C., Harris, D.J., Lip, V., An, Y., Biagioli, M., MacDonald, M.E., Lin, M., Haggarty, S.J., Sklar, P., Purcell, S.,

Kellis, M., Schwartz, S., Shaffer, L.G., Natowicz, M.R., Shen, Y., Morton, C.C., Gusella, J.F., Ernst, C., 2012. Disruption of a Large Intergenic Noncoding RNA in Subjects with Neurodevelopmental Disabilities. Am. J. Hum. Genet. 91, 1128–1134. https://doi.org/10.1016/j.ajhg.2012.10.016

Tandon, R., Gaebel, W., Barch, D.M., Bustillo, J., Gur, R.E., Heckers, S., Malaspina, D., Owen, M.J., Schultz, S., Tsuang, M., Van Os, J., Carpenter, W., 2013. Definition and description of schizophrenia in the DSM-5. Schizophr. Res. 150, 3–10. https://doi.org/10.1016/j.schres.2013.05.028

Tansey, K.E., Rees, E., Linden, D.E., Ripke, S., Chambert, K.D., Moran, J.L., McCarroll, S.A., Holmans, P., Kirov, G., Walters, J., Owen, M.J., O'Donovan, M.C., 2016. Common alleles contribute to schizophrenia in CNV carriers. Mol. Psychiatry 21, 1085–1089. https://doi.org/10.1038/mp.2015.143

Tapper, W., Collins, A., Gibson, J., Maniatis, N., Ennis, S., Morton, N.E., 2005. A map of the human genome in linkage disequilibrium units. Proc. Natl. Acad. Sci. 102, 11835–11839. https://doi.org/10.1073/pnas.0505262102

Taylor, S.C., Laperriere, G., Germain, H., 2017. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. Sci. Rep. 7, 2409. https://doi.org/10.1038/s41598-017-02217-x

Temur, M., Serpim, G., Tuzluoğlu, S., Taşgöz, F.N., Şahin, E., Üstünyurt, E., 2020. Comparison of serum human pregnancy-specific beta-1-glycoprotein 1 levels in pregnant women with or without preeclampsia. J. Obstet. Gynaecol. 40, 1074–1078. https://doi.org/10.1080/01443615.2019.1679734

Thakur, J., Packiaraj, J., Henikoff, S., 2021. Sequence, Chromatin and Evolution of Satellite DNA. Int. J. Mol. Sci. 22, 4309. https://doi.org/10.3390/ijms22094309

The Cancer Genome Atlas Research Network, 2015. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. N. Engl. J. Med. 372, 2481–2498. https://doi.org/10.1056/NEJMoa1402121

THE GTEX CONSORTIUM, 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318–1330. https://doi.org/10.1126/science.aaz1776

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Thygesen, J.H., Wolfe, K., McQuillin, A., Viñas-Jornet, M., Baena, N., Brison, N., D'Haenens, G., Esteba-Castillo, S., Gabau, E., Ribas-Vidal, N., Ruiz, A., Vermeesch, J., Weyts, E., Novell, R., Buggenhout, G.V., Strydom, A., Bass, N., Guitart, M., Vogels, A., 2018. Neurodevelopmental risk copy number variants in adults with intellectual disabilities and comorbid psychiatric disorders. Br. J. Psychiatry 212, 287–294. https://doi.org/10.1192/bjp.2017.65

Tjio, J.H., Levan, A., 1956. The Chromosome Number of Man. Hereditas 42, 1–6. https://doi.org/10.1111/j.1601-5223.1956.tb03010.x

Tochitani, S., Hayashizaki, Y., 2008. Nkx2.2 antisense RNA overexpression enhanced oligodendrocytic differentiation. Biochem. Biophys. Res. Commun. 372, 691–696. https://doi.org/10.1016/j.bbrc.2008.05.127

Torvund-Jensen, J., Steengaard, J., Askebjerg, L.B., Kjaer-Sorensen, K., Laursen, L.S., 2018. The 3'UTRs of Myelin Basic Protein mRNAs Regulate Transport, Local Translation and Sensitivity to Neuronal Activity in Zebrafish. Front. Mol. Neurosci. 11, 185. https://doi.org/10.3389/fnmol.2018.00185

Torvund-Jensen, J., Steengaard, J., Reimer, L., Fihl, L.B., Laursen, L.S., 2014. Transport and translation of MBP mRNA is regulated differently by distinct hnRNP proteins. J. Cell Sci. 127, 1550–1564. https://doi.org/10.1242/jcs.140855

Towler, C.M., Horne, C.H.W., Jandial, V., Campbell, D.M., MacGillivray, I., 1977. PLASMA LEVELS OF PREGNANCY-SPECIFIC β1-GLYCOPROTEIN IN COMPLICATED PREGNANCIES. BJOG Int. J. Obstet. Gynaecol. 84, 258–263. https://doi.org/10.1111/j.1471-0528.1977.tb12573.x

Towler, C.M., Horne, C.H.W., Jandial, V., Campbell, D.M., MacGillivray, I., 1976. PLASMA LEVELS OF PREGNANCY–SPECIFIC β1–GLYCOPROTEIN IN NORMAL PREGNANCY. BJOG Int. J. Obstet. Gynaecol. 83, 775–779. https://doi.org/10.1111/j.1471-0528.1976.tb00743.x

Trost, B., Engchuan, W., Nguyen, C.M., Thiruvahindrapuram, B., Dolzhenko, E., Backstrom, I., Mirceta, M., Mojarad, B.A., Yin, Y., Dov, A., Chandrakumar, I., Prasolava, T., Shum, N., Hamdan, O., Pellecchia, G., Howe, J.L., Whitney, J., Klee, E.W., Baheti, S., Amaral, D.G., Anagnostou, E., Elsabbagh, M., Fernandez, B.A., Hoang, N., Lewis, M.E.S., Liu, X., Sjaarda, C., Smith, I.M., Szatmari, P., Zwaigenbaum, L., Glazer, D., Hartley, D., Stewart, A.K., Eberle, M.A., Sato, N., Pearson, C.E., Scherer, S.W., Yuen, R.K.C., 2020. Genome-wide detection of tandem DNA repeats that are expanded in autism. Nature 586, 80–86. https://doi.org/10.1038/s41586-020-2579-z

Trotta, A., Di Forti, M., Mondelli, V., Dazzan, P., Pariante, C., David, A., Mulè, A., Ferraro, L., Formica, I., Murray, R.M., Fisher, H.L., 2013. Prevalence of bullying victimisation amongst first-episode psychosis patients and unaffected controls. Schizophr. Res. 150, 169–175. https://doi.org/10.1016/j.schres.2013.07.001

Trubetskoy, V., Pardiñas, A.F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T.B., Bryois, J., Chen, C.-Y., Dennison, C.A., Hall, L.S., Lam, M., Watanabe, K., Frei, O., Ge, T., Harwood, J.C., Koopmans, F., Magnusson, S., Richards, A.L., Sidorenko, J., Wu, Y., Zeng, J., Grove, J., Kim, M., Li, Z., Voloudakis, G., Zhang, W., Adams, M., Agartz, I., Atkinson, E.G., Agerbo, E., Al Eissa, M., Albus, M., Alexander, M., Alizadeh, B.Z., Alptekin, K., Als, T.D., Amin, F., Arolt, V., Arrojo, M., Athanasiu, L., Azevedo, M.H., Bacanu, S.A., Bass, N.J., Begemann, M., Belliveau, R.A., Bene, J., Benyamin, B., Bergen, S.E., Blasi, G., Bobes, J., Bonassi, S., Braun, A., Bressan, R.A., Bromet, E.J., Bruggeman, R., Buckley, P.F., Buckner, R.L., Bybjerg-Grauholm, J., Cahn, W., Cairns, M.J., Calkins, M.E., Carr, V.J., Castle, D., Catts, S.V., Chambert, K.D., Chan, R.C.K., Chaumette, B., Cheng, W., Cheung, E.F.C., Chong, S.A., Cohen, D., Consoli, A., Cordeiro, Q., Costas, J., Curtis, C., Davidson, M., Davis, K.L., de Haan, L., Degenhardt, F., DeLisi, L.E., Demontis, D., Dickerson, F., Dikeos, D., Dinan, T., Djurovic, S., Duan, J., Ducci, G., Dudbridge, F., Eriksson, J.G., Fañanás, L., Faraone, S.V., Fiorentino, A., Forstner, A., Frank, J., Freimer, N.B., Fromer, M., Frustaci, A., Gadelha, A.,

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

E., Ehrenreich, H., Escott-Price, V., Esko, T., Fanous, A.H., Gareeva, A., Gawlik, M., Gejman, P.V., Gill, M., Glatt, S.J., Golimbet, V., Hong, K.S., Hultman, C.M., Hyman, S.E., Iwata, N., Jönsson, E.G., Kahn, R.S., Kennedy, J.L., Khusnutdinova, E., Kirov, G., Knowles, J.A., Krebs, M.-O., Laurent-Levinson, C., Lee, J., Lencz, T., Levinson, D.F., Li, Q.S., Liu, J., Malhotra, A.K., Malhotra, D., McIntosh, A., McQuillin, A., Menezes, P.R., Morgan, V.A., Morris, D.W., Mowry, B.J., Murray, R.M., Nimgaonkar, V., Nöthen, M.M., Ophoff, R.A., Paciga, S.A., Palotie, A., Pato, C.N., Qin, S., Rietschel, M., Riley, B.P., Rivera, M., Rujescu, D., Saka, M.C., Sanders, A.R., Schwab, S.G., Serretti, A., Sham, P.C., Shi, Y., St Clair, D., Stefánsson, H., Stefansson, K., Tsuang, M.T., van Os, J., Vawter, M.P., Weinberger, D.R., Werge, T., Wildenauer, D.B., Yu, X., Yue, W., Holmans, P.A., Pocklington, A.J., Roussos, P., Vassos, E., Verhage, M., Visscher, P.M., Yang, J., Posthuma, D., Andreassen, O.A., Kendler, K.S., Owen, M.J., Wray, N.R., Daly, M.J., Huang, H., Neale, B.M., Sullivan, P.F., Ripke, S., Walters, J.T.R., O'Donovan, M.C., 2022. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature 604, 502–508. https://doi.org/10.1038/s41586-022-04434-5

Umeton, R., Bellucci, G., Bigi, R., Romano, S., Buscarinu, M.C., Reniè, R., Rinaldi, V., Pizzolato Umeton, R., Morena, E., Romano, C., Mechelli, R., Salvetti, M., Ristori, G., 2022. Multiple sclerosis genetic and non-genetic factors interact through the transient transcriptome. Sci. Rep. 12, 7536. https://doi.org/10.1038/s41598-022-11444-w

Valgardsdottir, R., Chiodi, I., Giordano, M., Rossi, A., Bazzini, S., Ghigna, C., Riva, S., Biamonti, G., 2008. Transcription of Satellite III non-coding RNAs is a general stress response in human cells. Nucleic Acids Res. 36, 423–434. https://doi.org/10.1093/nar/gkm1056

van den Bent, M.J., 2007. Anaplastic Oligodendroglioma and Oligoastrocytoma. Neurol. Clin., Brain Tumors in Adults 25, 1089–1109. https://doi.org/10.1016/j.ncl.2007.07.013

Van Den Bossche, M.J., Strazisar, M., Cammaerts, S., Liekens, A.M., Vandeweyer, G., Depreeuw, V., Mattheijssens, M., Lenaerts, A.-S., De Zutter, S., De Rijk, P., Sabbe, B., Del-Favero, J., 2013. Identification of rare copy number variants in high burden schizophrenia families. Am. J. Med. Genet. B Neuropsychiatr. Genet. 162, 273–282. https://doi.org/10.1002/ajmg.b.32146

van der Lee, R., Wiel, L., van Dam, T.J.P., Huynen, M.A., 2017. Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts. Nucleic Acids Res. 45, 10634–10648. https://doi.org/10.1093/nar/gkx704

Varese, F., Smeets, F., Drukker, M., Lieverse, R., Lataster, T., Viechtbauer, W., Read, J., van Os, J., Bentall, R.P., 2012. Childhood adversities increase the risk of psychosis: a meta-analysis of patient-control, prospective- and cross-sectional cohort studies. Schizophr. Bull. 38, 661–671. https://doi.org/10.1093/schbul/sbs050

Vassos, E., Collier, D.A., Holden, S., Patch, C., Rujescu, D., St Clair, D., Lewis, C.M., 2010. Penetrance for copy number variants associated with

schizophrenia. Hum. Mol. Genet. 19, 3477–3481. https://doi.org/10.1093/hmg/ddq259

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Francesco, V.D., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.-R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z.Y., Wang, A., Wang, X., Wang, J., Wei, M.-H., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S.C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.-H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigó, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.-H., Coyne, M., Dahlke, C., Mays, A.D., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X., 2001.

The Sequence of the Human Genome. Science 291, 1304–1351. https://doi.org/10.1126/science.1058040

Vitart, V., Rudan, I., Hayward, C., Gray, N.K., Floyd, J., Palmer, C.N.A., Knott, S.A., Kolcic, I., Polasek, O., Graessler, J., Wilson, J.F., Marinaki, A., Riches, P.L., Shu, X., Janicijevic, B., Smolej-Narancic, N., Gorgoni, B., Morgan, J., Campbell, S., Biloglav, Z., Barac-Lauc, L., Pericic, M., Klaric, I.M., Zgaga, L., Skaric-Juric, T., Wild, S.H., Richardson, W.A., Hohenstein, P., Kimber, C.H., Tenesa, A., Donnelly, L.A., Fairbanks, L.D., Aringer, M., McKeigue, P.M., Ralston, S.H., Morris, A.D., Rudan, P., Hastie, N.D., Campbell, H., Wright, A.F., 2008. SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. Nat. Genet. 40, 437–442. https://doi.org/10.1038/ng.106

Vitsios, D., Dhindsa, R.S., Middleton, L., Gussow, A.B., Petrovski, S., 2021. Prioritizing non-coding regions based on human genomic constraint and sequence context with deep learning. Nat. Commun. 12, 1504. https://doi.org/10.1038/s41467-021-21790-4

Voce, D.J., Bernal, G.M., Wu, L., Crawley, C.D., Zhang, W., Mansour, N.M., Cahill, K.E., Szymura, S.J., Uppal, A., Raleigh, D.R., Spretz, R., Nunez, L., Larsen, G., Khodarev, N.N., Weichselbaum, R.R., Yamini, B., 2019. Temozolomide treatment induces lncRNA MALAT1 in an NF-κB and p53 co-dependent manner in glioblastoma. Cancer Res. 79, 2536–2548. https://doi.org/10.1158/0008-5472.CAN-18-2170

Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G., McCulloch, L.J., Ferreira, T., Grallert, H., Amin, N., Wu, G., Willer, C.J., Raychaudhuri, S., McCarroll, S.A., Langenberg, C., Hofmann, O.M., Dupuis, J., Qi, L., Segrè, A.V., van Hoek, M., Navarro, P., Ardlie, K., Balkau, B., Benediktsson, R., Bennett, A.J., Blagieva, R., Boerwinkle, E., Bonnycastle, L.L., Boström, K.B., Bravenboer, B., Bumpstead, S., Burtt, N.P., Charpentier, G., Chines, P.S., Cornelis, M., Couper, D.J., Crawford, G., Doney, A.S.F., Elliott, K.S., Elliott, A.L., Erdos, M.R., Fox, C.S., Franklin, C.S., Ganser, M., Gieger, C., Grarup, N., Green, T., Griffin, S., Groves, C.J., Guiducci, C., Hadjadj, S., Hassanali, N., Herder, C., Isomaa, B., Jackson, A.U., Johnson, P.R.V., Jørgensen, T., Kao, W.H.L., Klopp, N., Kong, A., Kraft, P., Kuusisto, J., Lauritzen, T., Li, M., Lieverse, A., Lindgren, C.M., Lyssenko, V., Marre, M., Meitinger, T., Midthjell, K., Morken, M.A., Narisu, N., Nilsson, P., Owen, K.R., Payne, F., Perry, J.R.B., Petersen, A.-K., Platou, C., Proença, C., Prokopenko, I., Rathmann, W., Rayner, N.W., Robertson, N.R., Rocheleau, G., Roden, M., Sampson, M.J., Saxena, R., Shields, B.M., Shrader, P., Sigurdsson, G., Sparsø, T., Strassburger, K., Stringham, H.M., Sun, Q., Swift, A.J., Thorand, B., Tichet, J., Tuomi, T., van Dam, R.M., van Haeften, T.W., van Herpt, T., van Vliet-Ostaptchouk, J.V., Walters, G.B., Weedon, M.N., Wijmenga, C., Witteman, J., Bergman, R.N., Cauchi, S., Collins, F.S., Gloyn, A.L., Gyllensten, U., Hansen, T., Hide, W.A., Hitman, G.A., Hofman, A., Hunter, D.J., Hveem, K., Laakso, M., Mohlke, K.L., Morris, A.D., Palmer, C.N.A., Pramstaller, P.P., Rudan, I., Sijbrands, E., Stein, L.D., Tuomilehto, J., Uitterlinden, A., Walker, M., Wareham, N.J., Watanabe, R.M., Abecasis,

G.R., Boehm, B.O., Campbell, H., Daly, M.J., Hattersley, A.T., Hu, F.B., Meigs, J.B., Pankow, J.S., Pedersen, O., Wichmann, H.-E., Barroso, I., Florez, J.C., Frayling, T.M., Groop, L., Sladek, R., Thorsteinsdottir, U., Wilson, J.F., Illig, T., Froguel, P., van Duijn, C.M., Stefansson, K., Altshuler, D., Boehnke, M., McCarthy, M.I., 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat. Genet. 42, 579–589. https://doi.org/10.1038/ng.609

Voineskos, A.N., Felsky, D., Kovacevic, N., Tiwari, A.K., Zai, C., Chakravarty, M.M., Lobaugh, N.J., Shenton, M.E., Rajji, T.K., Miranda, D., Pollock, B.G., Mulsant, B.H., McIntosh, A.R., Kennedy, J.L., 2013. Oligodendrocyte genes, white matter tract integrity, and cognition in schizophrenia. Cereb. Cortex N. Y. N 1991 23, 2044–2057. https://doi.org/10.1093/cercor/bhs188

Vollger, M.R., Guitart, X., Dishuck, P.C., Mercuri, L., Harvey, W.T., Gershman, A., Diekhans, M., Sulovari, A., Munson, K.M., Lewis, A.P., Hoekzema, K., Porubsky, D., Li, R., Nurk, S., Koren, S., Miga, K.H., Phillippy, A.M., Timp, W., Ventura, M., Eichler, E.E., 2022. Segmental duplications and their variation in a complete human genome. Science 376, eabj6965. https://doi.org/10.1126/science.abj6965

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H., Bucan, M., 2007. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 17, 1665–1674. https://doi.org/10.1101/gr.6861907

Wang, N., Li, R., Xue, M., 2019. Potential regulatory network in the PSG10P/miR-19a-3p/IL1RAP pathway is possibly involved in preeclampsia pathogenesis. J. Cell. Mol. Med. 23, 852–864. https://doi.org/10.1111/jcmm.13985

Wang, P., Xue, Y., Han, Y., Lin, L., Wu, C., Xu, S., Jiang, Z., Xu, J., Liu, Q., Cao, X., 2014. The STAT3-Binding Long Noncoding RNA lnc-DC Controls Human Dendritic Cell Differentiation. Science 344, 310–313. https://doi.org/10.1126/science.1251456

Wang, Q., Jia, Y., Wang, Y., Jiang, Z., Zhou, X., Zhang, Z., Nie, C., Li, J., Yang, N., Qu, L., 2019. Evolution of cis- and trans-regulatory divergence in the chicken genome between two contrasting breeds analyzed using three tissue types at one-day-old. BMC Genomics 20, 933. https://doi.org/10.1186/s12864-019-6342-5

Wang, S., Bleeck, A., Nadif Kasri, N., Kleefstra, T., van Rhijn, J.-R., Schubert, D., 2021. SETD1A Mediated H3K4 Methylation and Its Role in Neurodevelopmental and Neuropsychiatric Disorders. Front. Mol. Neurosci. 14.

Wang, Y., Liu, C., Hu, R., Geng, J., Lu, J., Zhao, X., Xiong, Y., Wu, J., Yin, A., 2023. Prenatal phenotype features and genetic etiology of the Williams-Beuren syndrome and literature review. Front. Pediatr. 11.

Warren, J., Im, M., Ballesteros, A., Ha, C., Moore, T., Lambert, F., Lucas, S., Hinz, B., Dveksler, G., 2018. Activation of latent transforming growth factor-β1, a conserved function for pregnancy-specific beta 1-glycoproteins. Mol. Hum. Reprod. 24, 602–612. https://doi.org/10.1093/molehr/gay044

Wat, M.J., Enciso, V.B., Wiszniewski, W., Resnick, T., Bader, P., Roeder, E.R., Freedenberg, D., Brown, C., Stankiewicz, P., Cheung, S.-W., Scott, D.A., 2010. Recurrent microdeletions of 15q25.2 are associated with increased risk of congenital diaphragmatic hernia, cognitive deficits and possibly Diamond--Blackfan anaemia. J. Med. Genet. 47, 777–781. https://doi.org/10.1136/jmg.2009.075903

Wei, H., Dong, X., You, Y., Hai, B., Duran, R.C.-D., Wu, X., Kharas, N., Wu, J.Q., 2021. OLIG2 regulates lncRNAs and its own expression during oligodendrocyte lineage formation. BMC Biol. 19, 132. https://doi.org/10.1186/s12915-021-01057-6

Weiss, M.M., Hermsen, M.A., Meijer, G.A., van Grieken, N.C., Baak, J.P., Kuipers, E.J., van Diest, P.J., 1999. Comparative genomic hybridisation. Mol. Pathol. MP 52, 243–251. https://doi.org/10.1136/mp.52.5.243

Whalen, S., Pollard, K.S., 2022. Enhancer Function and Evolutionary Roles of Human Accelerated Regions. Annu. Rev. Genet. 56, 423–439. https://doi.org/10.1146/annurev-genet-071819-103933

White, R., Krämer-Albers, E.-M., 2014. Axon-glia interaction and membrane traffic in myelin formation. Front. Cell. Neurosci. 7.

Wick, W., Hartmann, C., Engel, C., Stoffels, M., Felsberg, J., Stockhammer, F., Sabel, M.C., Koeppen, S., Ketter, R., Meyermann, R., Rapp, M., Meisner, C., Kortmann, R.D., Pietsch, T., Wiestler, O.D., Ernemann, U., Bamberg, M., Reifenberger, G., von Deimling, A., Weller, M., 2009. NOA-04 Randomized Phase III Trial of Sequential Radiochemotherapy of Anaplastic Glioma With Procarbazine, Lomustine, and Vincristine or Temozolomide. J. Clin. Oncol. 27, 5874–5880. https://doi.org/10.1200/JCO.2009.23.6497

Wiersma, D., Wanderling, J., Dragomirecka, E., Ganev, K., Harrison, G., An Der Heiden, W., Nienhuis, F.J., Walsh, D., 2000. Social disability in schizophrenia: its development and prediction over 15 years in incidence cohorts in six European centres. Psychol. Med. 30, 1155–1167. https://doi.org/10.1017/s0033291799002627

Willatt, L., Cox, J., Barber, J., Cabanas, E.D., Collins, A., Donnai, D., FitzPatrick, D.R., Maher, E., Martin, H., Parnau, J., Pindar, L., Ramsay, J., Shaw-Smith, C., Sistermans, E.A., Tettenborn, M., Trump, D., de Vries, B.B.A., Walker, K., Raymond, F.L., 2005. 3q29 Microdeletion Syndrome: Clinical and Molecular Characterization of a New Syndrome. Am. J. Hum. Genet. 77, 154–160. https://doi.org/10.1086/431653

Wittkopp, P.J., 2005. Genomic sources of regulatory variation in cis and in trans. Cell. Mol. Life Sci. CMLS 62, 1779–1783. https://doi.org/10.1007/s00018-005-5064-9

Xiao, R., Boehnke, M., 2009. Quantifying and correcting for the winner's curse in genetic association studies. Genet. Epidemiol. 33, 453–462. https://doi.org/10.1002/gepi.20398

Xiao, X., Chang, H., Li, M., 2017. Molecular mechanisms underlying noncoding risk variations in psychiatric genetic studies. Mol. Psychiatry 22, 497–511. https://doi.org/10.1038/mp.2016.241

Xu, C., Zhang, J., 2021. Mammalian circular RNAs result largely from splicing errors. Cell Rep. 36, 109439. https://doi.org/10.1016/j.celrep.2021.109439

Xu, K., Schadt, E.E., Pollard, K.S., Roussos, P., Dudley, J.T., 2015. Genomic and Network Patterns of Schizophrenia Genetic Variation in Human Evolutionary Accelerated Regions. Mol. Biol. Evol. 32, 1148–1160. https://doi.org/10.1093/molbev/msv031

Xu, Y., Xi, J., Wang, G., Guo, Z., Sun, Q., Lu, C., Ma, L., Wu, Y., Jia, W., Zhu, S., Guo, X., Bian, S., Kang, J., 2021. PAUPAR and PAX6 sequentially regulate human embryonic stem cell cortical differentiation. Nucleic Acids Res. 49, 1935–1950. https://doi.org/10.1093/nar/gkab030

Yamazaki, T., Souquere, S., Chujo, T., Kobelke, S., Chong, Y.S., Fox, A.H., Bond, C.S., Nakagawa, S., Pierron, G., Hirose, T., 2018. Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. Mol. Cell 70, 1038-1053.e7. https://doi.org/10.1016/j.molcel.2018.05.019

Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., Lancet, D., Shmueli, O., 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 21, 650–659. https://doi.org/10.1093/bioinformatics/bti042

Yao, R.-W., Wang, Y., Chen, L.-L., 2019. Cellular functions of long noncoding RNAs. Nat. Cell Biol. 21, 542–551. https://doi.org/10.1038/s41556-019-0311-8

Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., Visscher, P.M., GIANT Consortium, 2018. Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. Hum. Mol. Genet. 27, 3641–3649. https://doi.org/10.1093/hmg/ddy271

Yoon, J., Mao, Y., 2021. Dissecting Molecular Genetic Mechanisms of 1q21.1 CNV in Neuropsychiatric Disorders. Int. J. Mol. Sci. 22, 5811. https://doi.org/10.3390/ijms22115811

Yue, P., Jing, L., Zhao, X., Zhu, H., Teng, J., 2019. Down-regulation of taurine-up-regulated gene 1 attenuates inflammation by sponging miR-9-5p via targeting NF-κB1/p50 in multiple sclerosis. Life Sci. 233, 116731. https://doi.org/10.1016/j.lfs.2019.116731

Zamore, P.D., Williamson, J.R., Lehmann, R., 1997. The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. RNA 3, 1421–1433.

Zarrei, M., MacDonald, J.R., Merico, D., Scherer, S.W., 2015. A copy number variation map of the human genome. Nat. Rev. Genet. 16, 172–183. https://doi.org/10.1038/nrg3871

Zeng, L., Zhang, P., Shi, L., Yamamoto, V., Lu, W., Wang, K., 2013. Functional Impacts of NRXN1 Knockdown on Neurodevelopment in Stem Cell Models. PLOS ONE 8, e59685. https://doi.org/10.1371/journal.pone.0059685

Zhang, F., Lupski, J.R., 2015. Non-coding genetic variants in human disease. Hum. Mol. Genet. 24, R102-110. https://doi.org/10.1093/hmg/ddv259

Zhang, J.-P., Robinson, D.G., Gallego, J.A., John, M., Yu, J., Addington, J., Tohen, M., Kane, J.M., Malhotra, A.K., Lencz, T., 2015. Association of a Schizophrenia Risk Variant at the DRD2 Locus With Antipsychotic Treatment

Response in First-Episode Psychosis. Schizophr. Bull. 41, 1248–1255. https://doi.org/10.1093/schbul/sbv116
Zhang, S., Wang, Y., Zhu, X., Song, L., Zhan, X., Ma, E., McDonough, J., Fu, H., Cambi, F., Grinspan, J., Guo, F., 2021. The Wnt Effector TCF7l2 Promotes Oligodendroglial Differentiation by Repressing Autocrine BMP4-Mediated Signaling. J. Neurosci. 41, 1650–1664. https://doi.org/10.1523/JNEUROSCI.2386-20.2021
Zhao, L., Triche, E.W., Walsh, K.M., Bracken, M.B., Saftlas, A.F., Hoh, J., Dewan, A.T., 2012. Genome-wide association study identifies a maternal copy-number deletion in PSG11 enriched among preeclampsia patients. BMC Pregnancy Childbirth 12, 61. https://doi.org/10.1186/1471-2393-12-61
Zhao, X., Bhattacharyya, A., 2018. Human Models Are Needed for Studying Human Neurodevelopmental Disorders. Am. J. Hum. Genet. 103, 829–857. https://doi.org/10.1016/j.ajhg.2018.10.009
Zhao, X., Collins, R.L., Lee, W.-P., Weber, A.M., Jun, Y., Zhu, Q., Weisburd, B., Huang, Y., Audano, P.A., Wang, H., Walker, M., Lowther, C., Fu, J., Gerstein, M.B., Devine, S.E., Marschall, T., Korbel, J.O., Eichler, E.E., Chaisson, M.J.P., Lee, C., Mills, R.E., Brand, H., Talkowski, M.E., 2021. Expectations and blind spots for structural variation detection from long-read assemblies and short-read genome sequencing technologies. Am. J. Hum. Genet. 108, 919–928. https://doi.org/10.1016/j.ajhg.2021.03.014
Zhao, X., Moore, D., 2018. Neural stem cells: developmental mechanisms and disease modeling. Cell Tissue Res. 371, 1–6. https://doi.org/10.1007/s00441-017-2738-1
Zhao, Y., Ding, M., Pang, H., Xu, X.M., Wang, B.J., 2014. Relationship between genetic polymorphisms in the DRD5 gene and paranoid schizophrenia in northern Han Chinese. Genet. Mol. Res. 13, 1609–1618. https://doi.org/10.4238/2014.March.12.13
Zheng-Bradley, X., Streeter, I., Fairley, S., Richardson, D., Clarke, L., Flicek, P., the 1000 Genomes Project Consortium, 2017. Alignment of 1000 Genomes Project reads to reference assembly GRCh38. GigaScience 6, gix038. https://doi.org/10.1093/gigascience/gix038
Zhu, F., Gong, F., Lin, G., Lu, G., 2013. DPY19L2 gene mutations are a major cause of globozoospermia: identification of three novel point mutations. Mol. Hum. Reprod. 19, 395–404. https://doi.org/10.1093/molehr/gat018
Zimmermann, W., Kammerer, R., 2021. The immune-modulating pregnancy-specific glycoproteins evolve rapidly and their presence correlates with hemochorial placentation in primates. BMC Genomics 22, 128. https://doi.org/10.1186/s12864-021-07413-8
Zody, M.C., Garber, M., Sharpe, T., Young, S.K., Rowen, L., O'Neill, K., Whittaker, C.A., Kamal, M., Chang, J.L., Cuomo, C.A., Dewar, K., FitzGerald, M.G., Kodira, C.D., Madan, Anup, Qin, S., Yang, X., Abbasi, N., Abouelleil, A., Arachchi, H.M., Baradarani, L., Birditt, B., Bloom, S., Bloom, T., Borowsky, M.L., Burke, J., Butler, J., Cook, A., DeArellano, K., DeCaprio, D., Dorris, L., Dors, M., Eichler, E.E., Engels, R., Fahey, J., Fleetwood, P., Friedman, C., Gearin, G., Hall, J.L., Hensley, G., Johnson, E., Jones, C., Kamat, A., Kaur, A., Locke, D.P., Madan, Anuradha, Munson, G., Jaffe, D.B., Lui, A.,

Macdonald, P., Mauceli, E., Naylor, J.W., Nesbitt, R., Nicol, R., O'Leary, S.B., Ratcliffe, A., Rounsley, S., She, X., Sneddon, K.M.B., Stewart, S., Sougnez, C., Stone, S.M., Topham, K., Vincent, D., Wang, S., Zimmer, A.R., Birren, B.W., Hood, L., Lander, E.S., Nusbaum, C., 2006. Analysis of the DNA sequence and duplication history of human chromosome 15. Nature 440, 671–675. https://doi.org/10.1038/nature04601