Autosomal Recessive Variants in Intellectual Disability and Autism Spectrum Disorder

by

Ricardo Simeon Harripaul

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy of Medical Science

Institute of Medical Sciences
University of Toronto

© Copyright by Ricardo Simeon Harripaul 2021

Autosomal Recessive Variants in Intellectual Disability and Autism Spectrum Disorder

Ricardo Simeon Harripaul

Doctor of Philosophy of Medical Science

Institute of Medical Sciences
University of Toronto

2021

Abstract

The development of the nervous system is a tightly timed and controlled process where aberrant development may lead to neurodevelopmental disorders. Two of the most common forms of neurodevelopmental disorders are Intellectual Disability (ID) and Autism Spectrum Disorder (ASD). These two disorders are intertwined, but much about the etiological relationship between them remains a mystery. Little attention has been given to the role recessive variants play in ID and ASD, perhaps related to the ability of recessive variants to remain hidden in the population. This thesis aims to help determine the role of recessive variants in ID and ASD and identify novel genes associated with these neurodevelopmental disorders. Technological advancements such as Next Generation Sequencing (NGS) and large population-scale sequencing reference sets have ushered in a new era of high throughput gene identification. In total, 307 families were whole-exome sequenced (192 consanguineous multiplex ID families and 115 consanguineous trio ASD families), representing 537 samples in total. This work identified 26 novel genes for ID such as *ABI2, MAPK8, MBOAT7, MPDZ, PIDD1, SLAIN1, TBC1D23, TRAPPC6B, UBA7,* and *USP44* in large multiplex families with an overall diagnostic yield of 51 %. ASD often has other comorbid features and the connection with ID and ASD has been observed in literature with an estimated 40-70% if ASD probands also have comorbid ID. Extending the use of consanguineous families in autosomal recessive variant identification in ASD, the diagnostic yield for this ASD trio cohort was ~26 %, with 28 of the known variants coming from previously

known genes. This study identified putative variants with functional mouse models that display behavioral and nervous system abnormalities involving genes such as *EPHB1, VPS16, CNGA2* and *EFR3B*. Of the neurodevelopmental genes from this cohort, *ZNF292* is the most associated with ID with or without syndromic features and ASD. The overall body of work highlights the clinical and genetic heterogeneity in neurodevelopmental disorders and demonstrates the important role that recessive variants have in neurodevelopment disorders, including ASD. The identification of novel genes will hopefully provide better diagnosis and clinical management for patients.

Acknowledgments

I would like to thanks my supervisor Dr. John Vincent for his mentorship and guidance throughout my PhD. Thank you for giving me the opportunity and the freedom to explore and learn.

To my Program Advisory Committee Dr. Lisa Strug, Dr. Phillip Awadalla, Dr. Leon French, and Dr. Joyce So for their perspectives, feedback, support and insights to improve my experimental design and myself. I would also like to thank Drs. Muhammad Ayub and Bita Bozorgmehr for their clinical insight and patient recruitment. I thank Dr. Arun Tiwari and Dr. Clement Zai for their help, honesty, listening and friendly discussion.

I would like to thank the MiND Lab and the lab members. Thank you, Anna Mikhailov, for managing the lab and making research easier for everyone. You were always quiet and patient. I also thank all the lab members Dr. Taimoor Sheikh, Stephen Pastore, Rebecca Lau, Farah Machcher and Dr. Nasim Vasli. Thank you to Fleming College for the great group of students that came to our lab Samantha, Manisha, Kayla and Ashlyn, Kamalben, Lauren, Shemika and Jenelle. Gene validation would not be possible without their hard work.

Thank you to Andy Wang and David Rotenberg of Krembil Centre for Neuroinformatics, for allowing me to use the cluster to progress my research as I did. They truly enable science at CAMH.

I would like to thank the funding support of the Canadian Institute of Health Research and the University of Toronto, Ontario Student Opportunity Trust Funds, Peterborough K.M. Hunter Charitable Foundation Graduate Awards for their financial support.

I would like to thank my friends Sejal, Kayla, Kap and Akhil for their support and encouragement. I would like to thank my parents Wayne and Tara Harripaul for their continual support and strength in this long journey. My sister Anastasia Harripaul-Yhap for support and listening. I would also like to thank Emily for her editing and patience with me.

I would also like to thank the families and patients that participated in these students as the research findings would not be possible without them.

Contributions

The content presented in this thesis is original work and the majority of the work, including study design, data analysis and manuscript preparation was done by the author Ricardo Harripaul. Below are the individuals who have contributed to the work presented in this thesis.

Dr. John Vincent (Supervisor) for guidance in the mentorship in the study design, interpretation of results and manuscript preparation.

Dr. Ghayda Mirzaa (collaborator) for her clinical insights into ZNF292 and the spectrum of clinical features in this specific gene.

Dr. Muhammad Ayub (collaborator) for his patient recruitment, clinical insights and aid in manuscript preparation and result interpretation.

Dr. Leon French (collaborator and Program Advisory Committee Member) for his mentorship and guidance in gene expression and neuroimaging analysis.

Table of Contents

Table of Contents vi

List of Tables xi

List of Figures xiii

List of Appendices xviii

List of Abbreviations xx

1 Introduction: Dissecting the Genetic Causes of Intellectual Disability by Genetic Sequencing 1

1.1 Pre-NGS High-throughput Sequencing Research for Gene Identification 2

1.2 X Chromosome –Targeted Sequencing Studies 3

1.3 Targeted Sequencing Studies 4

1.4 Whole-Exome Sequencing for Gene Identification 5

1.5 Whole Genome Sequencing for Gene Identification 7

1.6 Structural Variant Detection through NGS for Gene Identification in Intellectual Disability 9

1.7 NGS Strategies for Clinical Diagnostics 10

1.8 Aims and Hypothesis 11

1.8.1 Study One - Autosomal recessive variants in intellectual disability in consanguineous populations 11

1.8.2 Study Two – Whole exome sequence analysis in autism spectrum disorder trios from consanguineous populations 12

1.8.3 Study Three – *De novo* and inherited variants in ZNF292 underlie a neurodevelopmental disorder with features of autism spectrum disorder 12

2 Autosomal Recessive Variants in Intellectual Disability in Consanguineous Populations 14

2.1 Abstract 15

2.2 Introduction 15

2.3 Materials and Methods 17

2.3.1 Family recruitment 17

2.3.2 Autozygosity mapping 19

2.3.3 Structural Variant Analysis using Microarray 20

2.3.4 Whole exome sequencing 21

2.3.5 Sequencing alignment and variant calling 21

2.3.6 Variant Filtration 22

2.3.7 Gene list construction 22

2.3.8 Database searches across different neuropsychiatric and Neurodevelopmental Disorders and Functional and Animal Models 22

2.3.9 Anatomical expression analyses 23

2.3.10 Developmental brain expression analyses 23

2.3.11 Pathway analyses 24

2.4 Results 25

2.4.1 Variants Identified for Intellectual Disability 29

2.4.2 Gene list Construction 29

2.4.3 Homozygous loss-of-function Mutations 29

2.4.4 Missense mutations 33

2.4.5 Families with multiple variants 34

2.4.6 X-linked variants 36

2.4.7 Genes with 'hits' in multiple families 36

2.4.8 Families with mutations in previously identified genes for Metabolic or Hormonal Disorders 37

2.4.9 Three Novel Mutations in the Bilateral Fronto Parietal Polymicrogyria Gene *GPR56* 37

2.4.10 Structural Variant Analysis 43

2.5 Database Searches for Putative Genes 43

2.5.1 Sequencing Databases for Neuropsychiatric Disorders 43

2.5.2 Protein Interaction Analysis 44

2.5.3 Pathway analyses 44

2.6 Spatiotemporal Expressions Analyses 45

2.6.1 Anatomical expression analyses 45

2.6.2 Developmental brain expression analyses 45

2.7 Discussion 47

2.8 Addendum to Harripaul et al, 2018 51

3 Whole exome sequence analysis in autism spectrum disorder trios from consanguineous populations 52

3.1 Abstract 52

3.2 Introduction 53

3.3 Materials and Methods 55

3.3.1 Trio Family Recruitment 55

3.3.2 Whole Exome Sequencing 57

3.3.3 Sequence Alignment and Variant Calling 57

3.3.4 Annotation and Variant Prioritization 58

3.3.5 Construction of Gene List for RNA Expression and Gene Enrichment 61

3.3.6 Adult human brain gene expression data 61

3.3.7 Prenatal human gene expression data 61

3.3.8 Microarray gene expression data processing 62

3.3.9 Brain region enrichment analysis 62

3.3.10 Differential expression analyses ..... **Error! Bookmark not defined.**

3.3.11 Gene Ontology enrichment analysis ..... **Error! Bookmark not defined.**

3.3.12 Estimation of Cell-Type Proportions ..... **Error! Bookmark not defined.**

3.3.13 Developmental Expression and Neuroanatomy of Genes Involved in ASD and ID 62

3.4 Results 63

3.4.1 *De novo* Variant Identification 64

3.4.2 Autosomal Recessive Variant Identification 66

3.4.3 Developing Human Brain Atlas Cortical Layer Identification from Developing Human Brain 82

3.4.4 Cell-types Expression in ASD and ID 86

3.4.5 ASD Anatomical Mapping 90

3.4.6 Spatiotemporal Expression in PsychEncode and BrainSpan shows Striatum is Important in Developmental Trajectories 96

3.5 Discussion 108

4 *De novo* and inherited variants in *ZNF292* underlie a neurodevelopmental disorder with features of autism spectrum disorder 115

4.1 Abstract 115

4.2 Introduction 116

4.3 Materials and Methods 117

4.4 Results 120

4.5 Discussion 125

5 Concluding Summary, General Discussion and Future Directions 130

5.1 Concluding Summary 130

5.2 Novelty to the Field 132

5.3 Contribution to the Field 133

5.4 Unifying Discussion 135

5.4.1 Recessive variants contribute to the ASD and ID more than expected 135

5.4.2 Exome sequencing in addition to microarray is an excellent approach 136

5.4.3 Neurodevelopmental disorders represent a spectrum 136

5.5 Future Directions 139

5.5.1 More detailed clinical descriptions 139

5.5.2 Whole Genome Sequencing Potential for ASD and ID 140

5.5.3 Advances Utilizing RNA Sequencing 141

5.5.4 Translating and Functional Validation of Intellectual Disability and Autism Spectrum Disorder Genes 142

6 References 144

7 Appendices 161

7.1 Supplementary Material for Chapter 2 161

7.1.1 Methods 161

7.1.2 Supplemental Figures and Tables 167

7.2 Supplementary Material for Chapter 3 268

7.3 Supplementary Material for Chapter 4 292

7.3.1 Supplemental Materials and Methods 292

7.3.2 Supplemental Figures 319

7.4 Supplemental References 328

List of Tables

Table 2.1: Summary statistics for Iranian and Pakistani family cohorts.

Table 2.2: List of sequence variants identified for the Pakistani and Iranian families in the study. Families for which a single homozygous (or hemizygous for ChrX variants) mutation was identified, and no homozygous loss of function variants for the gene have been reported among ExAC database of control exomes (unless otherwise indicated). ExAC database minor allele frequencies (MAF) are indicated. Variants highlighted in bold are previously unreported in association with ID, and asterisks indicate genes/variants already reported by us in this cohort. A list of all variants identified including known syndromic ARID or XLID genes identified in the cohort is give in Table A- 3, and Table A- 3 for families where >1 variant was identified. For subjects where the phenotype is clearly NS-ARID, despite prior association of the gene with S-ARID (e.g. *LRP2* and *MECP2*), we have included the variant in this list. See Table A- 3 for *in silico* predictions of effects of variants.

Table 2.3: Candidate genes and sequence variants identified for the Pakistani and Iranian ID families in support of previously reported non-syndromic ID genes. Asterisks indicate genes/variants already reported by us in this cohort For subjects where the phenotype is clearly NS-ARID, despite prior association of the gene with S-ARID (e.g. *LRP2* and *MECP2*), we have included the variant in this list. See Table A- 1 for *in silico* predictions of effects of variants.

Table 2.4: Clinical Features of Families A, B, and C

Table 3.1: Demographic and Biological Gender Composition of Pakistani and Iranian Trio Families in this Study

Table 3.2: Percentage of Variants Identified by Inheritance Pattern for All Genes and Known Genes in this Study

Table 3.3:Autosomal *De novo* sequence variants identified in Pakistani, Iranian and Saudi ASD trio families by Genomic Coordinate (n=115, Pakistani Trios n = 58, Iranian Trios n= 41 , Saudi Trios n= 16).

Table 3.4: Autosomal Recessive genes and sequence variants identified in Pakistani and Iranian ASD trio families by Genomic Coordinate (n=115, Pakistani Trios n = 58, Iranian Trios n= 41 , Saudi Trios n= 16)

Table 3.5: *De novo* and Recessive X- Linked genes and sequence variants identified in Pakistani and Iranian ASD trio families by Genomic Coordinate (n=115, Pakistani Trios n = 58, Iranian Trios n= 41 , Saudi Trios n= 16)

Table 3.6: Developmental Cortical Layer Expression of ASD Genes Constructed from Literature

Table 3.7:Developmental Cortical Layer Expression in ID Genes Constructed from Literature

Table 3.8: Developmental Cortical Expression in Both ID, ASD from Literature and Genes Identified in Pakistani and Iranian Trio Families from this Study

Table 3.9: Cell-type Expression in Genes (n=148) Identified in this Pakistani and Iranian Trio Families

Table 3.10: Polygenic Cell-types in ASD Genes (n = 1068) from Literature

Table 3.11: Polygenic Cell-types of ID Genes (n = 1346) from Literature

Table 3.12: Polygenic Cell-types of ASD, ID and Pakistani and Iranian Trio Families Composite Genes (n = 2227)

Table 3.13 Multigene Expression Mapping in the Human Brain for Genes Identified Pakistani and Iranian Trio Families (n = 148)

Table 3.14 Multigene Expression Mapping of ASD Genes (n = 1068) Identified in Literature

Table 3.15 Multigene Expression Mapping for ID Genes (n = 1346) Identified in Literature

Table 3.16 Multigene Expression Mapping for ID and ASD Genes Identified in Literature and Pakistani and Iranian Trio Families (n = 2227)

Table 4.1 Summary of the clinical features of ZNF292 mutation-positive individuals (N=24)

List of Figures

Figure 2.1: Microarray and exome sequencing work flow. CNV, copy number variations; HBD, homozygosity-by-descent; WES, whole exome sequencing.

Figure 2.2: Pedigrees and HomozygosityMapper output for eight of the families

Figure 2.3: MRI from individual II:2 Family A. (A) MRI showing frontoparietal polymicrogyria transverse plane; (B) transverse plane showing cerebellum and cerebellar vermis; and (C) sagittal plane showing pons. MRI, magnetic resonance imaging.

Figure 2.4: (i) Pedigree drawings for Families A-C. Dark symbols indicate individuals affected by ID. Electropherograms showing the homozygous mutations identified in GPR56 in affected individuals (ii) and in heterozygous carriers (iii). For Family C, electropherograms from the reverse sequence is shown.

Figure 2.5: Spatiotemporal expression of ID genes in human brain development: a) Scaled expression trajectories of the ID Genes, averaged across brain regions. Solid lines are average expression for a group of genes. The lighter and thinner lines show trajectories of individual genes. Average expression for each gene group is depicted with points (green: glycosylation associated (14 genes), blue: hormone or metabolic (20), purple: remainder of the ID Genes (66) and red: remaining genes assayed). Local regression was used to smooth the z-scored expression values for individual genes and gene group averages (LOESS). A vertical dashed line marks birth. b) Heatmap showing brain samples that are enriched for specific expression of ID genes. For each brain region and brain combination, z-scored expression values for the ID genes was compared against all other genes (Wilcoxon one sided test, FDR corrected p-values). Significance levels are indicated by black (non-significant result), red ($p < 0.05$), orange ($p < 0.005$) and yellow ($p < 0.0005$). Missing values are shown in grey. Brain regions include: ventrolateral prefrontal cortex (VFC), dorsolateral prefrontal cortex (DFC), orbital frontal cortex (OFC), anterior (rostral) cingulate (medial prefrontal) cortex (MFC), posterior (caudal) superior temporal cortex (area 22c) (STC), inferolateral temporal cortex (area TEv, area 20) (ITC), posteroventral (inferior) parietal cortex (IPC), primary motor cortex (area M1, area 4) (M1C), primary somatosensory cortex (area S1, areas 3,1,2) (S1C), primary auditory cortex (core)

(A1C), primary visual cortex (striate cortex, area V1/17) (V1C), hippocampus (hippocampal formation) (HIP), amygdaloid complex (AMY), mediodorsal nucleus of thalamus (MD), striatum (STR) and the cerebellar cortex (CBC).

Figure 2.6: Cross-disorder overlap: The Venn diagrams shown indicate genes for which **a.** variants have been reported in other neuropsychiatric or neurodevelopmental disorders, either homozygous, compound heterozygous, or *de novo*, through searches of published gene-list; **b.** functional or animal models with relevant phenotypes have been reported. Genes newly reported here are in red text.

Figure 3.1: Workflow of experimental design where ASD probands and parents are collected, Sequenced and run through a variant prioritization schema.

Figure 3.2: Variant prioritization methods where variants that pass GATK best practices would be filtered for MAF less than 10^{-5} and then categorized into recessive or homozygous variants. The variants are then prioritized based on how damaging it would be to the protein and then filtered based on scoring algorithms to predict pathogenicity.

Figure 3.3: Spatiotemporal Expression of all genes identified in this cohort (n = 148) using the BrainSpan Atlas of the Developing Human Brain microarray data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

Figure 3.4: Spatiotemporal Expression of all genes (n = 148) identified in this cohort using the BrainSpan Atlas of the Developing Human Brain RNAseq A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

Figure 3.5: Spatiotemporal Expression of ASD genes (n = 1068) using the BrainSpan Atlas of the Developing Human Brain microarray data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

Figure 3.6: Spatiotemporal Expression of ASD genes (n = 1068) using the BrainSpan Atlas of the Developing Human Brain RNAseq data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

Figure 3.7: Spatiotemporal Expression of ID genes (n = 1346) using the BrainSpan Atlas of the Developing Human Brain microarray data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

Figure 3.8: Spatiotemporal Expression of ID genes (n = 1346) using the BrainSpan Atlas of the Developing Human Brain RNAseq data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

Figure 3.9: Spatiotemporal Expression of cohort, ID and ASD genes (n = 2227) using the BrainSpan Atlas of the Developing Human Brain microarray data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

Figure 3.10: Spatiotemporal Expression of Cohort, ID and ASD genes (n = 2227) using the BrainSpan Atlas of the Developing Human Brain RNAseq data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

Figure 3.11: Spatiotemporal Expression of genes identified in this cohort (n = 148) using psychENCODE RNAseq data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

Figure 3.12: Spatiotemporal Expression of ASD genes (n =1068) using psychENCODE RNAseq data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

Figure 3.13: Spatiotemporal Expression of ID genes (n = 1346) using psychENCODE RNAseq data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

Figure 3.14: Spatiotemporal Expression of ID, ASD and gene identified in this cohort (n = 2227) using psychENCODE RNAseq data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

Figure 3.15 *De novo* Known ID/ASD Variants with Functional Models from Genes Identified in Pakistani and Iranian Trio Families

Figure 3.16 Known Recessive ID/ASD Variants with Functional Mouse Models from Genes Identified in Pakistani and Iranian Trio Families

Figure 4.1 Genomic structure and distribution of variants in ZNF292. Most of the identified variants in *ZNF292* are truncating (frameshift, nonsense) located within the largest and most terminal exon (8) of the gene that encodes a ZNF292 DNA binding domain. The cDNA panel shows the coding and non-coding regions of the gene (in blue and yellow, respectively). The bottom panel shows the predicted protein domains including the zinc finger (C2H2 type) regions (shown in gray), the coiled coil domain (in pink) and the nuclear localization signal (in black). ZNF292 variants in the main cohort are shown, color-coded by type with nonsense variants shown in yellow, and frameshift variants in green

Figure 4.2 Photographs of individuals with pathogenic ZNF292 variants. (A, B) Photos of 17-027 showing a thin upper lip, smooth philtral folds, upturned nasal tip, sparse but long eyebrows with synophrys; (C-D) Photos of 18-007 at age 3.5 years showing epicanthal folds, mildly upslanted palpebral fissures, prominent forehead, and bulbous nose. D, hand photographs of the child showing ichthyosis; (E, F) Photo of 17-013 as a child (E), and as a teenager (F) showing laterally prominent ears, thick lips with a tented upper lip, short philtrum, prominent eyebrows with very prominent brow ridge and deep set eyes; (G, H) Frontal and lateral facial photograph of 17-005 at age 4 years 1 month showing mild micrognathia, short philtrum, and mildly down slanting palpebral fissures. All affected individuals have a prominent chin.

Figure 4.3 Brain MRI images of individuals with pathogenic variants in ZNF292. (A-B) T1-weighted and T2 weighted brain MR images of patient 17-003 showing mildly prominent ventricles (asterisks, F); (C-D) T1 weighted sagittal and axial images of 17-008 showing paucity of the white matter due to an in utero vascular insult and a thin corpus callosum (arrowhead; G); (E-H) T1 weighted and Constructive Interference in Steady State (CISS) images of patient 17-009 showing multiple abnormalities including hypoplasia of the cerebellar vermis and hemispheres, with marked asymmetry (arrow, I; asterisks, G), with possible clefting OF the cerebellum (arrow, G), as well as a deep infold within the cortical surface (arrowhead, F). Patient also has evidence of possible hemosiderin deposition that is asymmetric, suggesting a previous vascular insult/injury.

Figure 4.4 Expression of ZNF292 in developing human and mouse brains. A, *ZNF292* expression in the developing human brain (normalized RPKM data) showing high expression during early prenatal development that diminishes in the postnatal brain. *Data obtained from BrainSpan http://www.brainspan.org* . Codes: AMY, amygdala; CBC, cerebellum; HIP, hippocampus; MD, medial dorsal nucleus of the thalamus; NCTX, neocortex; NCTX, neocortex. B, *Zpf292* expression in the adult mouse brain showing the highest expression (indicated by higher intensity staining) in hippocampus and Purkinje cells of the cerebellum.

List of Appendices

7 Appendices 161

7.1 Supplementary Material for Chapter 2 161

7.1.1 Methods 161

7.1.1.1 Family Recruitment 161

7.1.1.2 HBD/Autozygosity mapping 161

7.1.1.3 CNV Analysis 162

7.1.1.4 Sequencing Alignment and Variant Calling 162

7.1.1.5 Variant Filtration 163

7.1.1.6 Anatomical expression analyses 163

7.1.1.7 Developmental brain expression analyses 164

7.1.1.8 Pathway analyses 165

7.1.1.9 References 165

7.1.2 Supplemental Figures and Tables 167

7.2 Supplementary Material for Chapter 3 268

7.3 Supplementary Material for Chapter 4 292

7.3.1 Supplemental Materials and Methods 292

7.3.1.1 Supplemental Clinical Data 294

7.3.1.1.1 Exome sequencing identified a *de novo* variant in *ZNF292* (c.6160_6161del, p. Glu2054Lysfs*14). 307

7.3.1.1.2 Exome sequencing identified a ZNF292 variant (c.6160_6161del, p. Glu2054Lysfs*14), which was confirmed by Sanger sequencing. 307

7.3.1.1.3 Clinical exome sequencing identified a *de novo ZNF292* variant (c.6661_6664delTTAG, p. Leu2221Serfs*10). 310

7.3.1.1.4 Exome sequencing detected a *de novo ZNF292* variant (c. 6578A>C, p.Tyr2193Ser). No additional genetic findings were reported. 315

7.3.2 Supplemental Figures.....319

7.4 Supplemental References.....328

List of Abbreviations

AR	Autosomal Recessive
ARID	Autosomal Recessive Intellectual Disability
ASD	Autism Spectrum Disorder
AUC	Area Under the Curve
AUROC	Area Under the Receiver Operator Curve
BPD	Bipolar Disorder
CAMH	Centre for Addition and Mental Health
CNV	Copy Number Variation
EE	Epileptic Encephalopathy
GWAS	Genome Wide Association Study
HBD	Homozygosity-by-Descent
HITS-CLIP	High-Throughput Sequencing of RNAs by Cross-Linking Immune-Precipitation
ID	Intellectual Disability
LoF	Loss of Function
NGS	Next Generation Sequencing
NSID	Nonsyndromic Intellectual Disability
RNAseq	RNA Sequencing
SCC	Specialized Computing Cluster
SCZ	Schizophrenia
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SV	Structural Variants
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
XLID	X-linked Intellectual Disability

1

Introduction: Dissecting the Genetic Causes of Intellectual Disability by Genetic Sequencing

This chapter is modified from the following published manuscript:

Harripaul, R., Noor, A., Ayub, M., & Vincent, J. B. (2017). The use of next-generation sequencing for research and diagnostics for intellectual disability. Cold Spring Harbor perspectives in medicine, 7(3), a026864.

Below is the link to the published manuscript:

http://perspectivesinmedicine.cshlp.org/content/7/3/a026864.short

Reprinted with permission from Cold Spring Harbor Perspectives in Medicine

1.1 Pre-NGS High-throughput Sequencing Research for Gene Identification

Large-scale sequencing attempts for Intellectual Disability (ID) gene/mutation screening did not begin with next-generation sequencing, but, for the obvious reasons of cost and labour, were limited in capacity and frequency. A few examples are described here. Fadi F Hamdan et al. (2011, p. 175) Sanger sequenced 197 genes involved in glutamate neurotransmission or associated with the glutamatergic synaptic machinery in 95 sporadic cases of nonsyndromic ID (NSID), finding an excess of *de novo* deleterious mutations, including *de novo* truncating mutations in *SYNGAP1, STXBP1*, and *SHANK3*. Another large scale Sanger sequencing project involved all coding genes (N=829) on the X-chromosome in 208 families with suspected X-linked ID (XLID), predominantly non-syndromic (Tarpey et al., 2009). Nine new genes implicated in XLID were reported, including the gene encoding the presynaptic protein *CASK*. Earlier, smaller scale sequencing efforts were also performed, for instance the screening of 47 X-chromosomal genes within a 7.4 Mb region of Xp11 and selected based on brain expression profile in 22 XLID families (Jensen et al., 2007). This study reported four new genes for XLID, however, a microarray hybridization approach was used for mutation screening here rather than Sanger sequencing, but still not a massively parallel sequencing approach that would be considered as "next-generation."

These studies provided a taste of what was to come, but with some warning as to the issues that would be faced with massively parallel sequencing. In the Tarpey et al. (2009) study, truncating mutations were identified in 19 genes that turned out not to be disease-related, either because they were present in the control population or did not segregate with ID in the families. Large numbers of missense changes were also identified, and the caution in interpretation for these has to be even greater due to the greater frequency of missense changes. Currently, we have the advantage of large databases of control exome or genome data with which to compare variants in affected individuals, e.g., gnomAD, NHLBI ESP and 1000 Genomes. Data from these can be used to exclude variants, if too common (or present as hemizygotes or homozygotes for X and autosomes respectively) in the controls.

1.2 X Chromosome –Targeted Sequencing Studies

The X chromosome was, for a long time, assumed to be the source of the majority of genes for ID, in part because males are more frequently affected with ID than females. It is estimated that X-linked intellectual Disability (XLID) accounts for 10%–12% of ID in males (Musante & Ropers, 2014; Ropers & Hamel, 2005), which explains why the majority of research for many years focused on the X chromosome, resulting in the identification of over 100 ID genes on the X chromosome. There are still many likely X-linked cases for which there is no molecular diagnosis. Hu et al. (2015) performed a study on 405 families which had clear familial evidence compatible with XLID (H. Hu et al., 2015). These families were termed unresolved because, despite microarray based CNV analysis and Sanger sequencing of a subset of XLID genes, there was no definitive cause for ID identified. More than half of apparent XLID families remain unresolved and this underscores the genetic heterogeneity present in ID. In the same study, H. Hu et al. (2015) then proceeded to massively parallel sequence 745 X-chromosomal genes to determine if there are other genes present in these families that cause ID. They found that 18% (74 of the 405 families) carried variants in previously identified XLID genes that were determined to be causative. The study also identified seven novel ID genes in 2% of their cohort such as *CLCN4*, *CNKSR2*, *FRMPD4*, *KLHL15*, *LAS1L*, *RLIM* and *USP27X*. The *CLCN4* gene had protein truncating and missense variants in five families. Electrophysiological studies in *Xenopus laevis* showed that these mutations impaired the function of the protein, and resulted in impaired neuronal differentiation. The CNSKR and FRMPD4 proteins were determined to interact with PSD95, which acts as a major scaffolding protein at the postsynaptic density. Their results provided a molecular diagnosis for the families involved, however, approximately half of XLID cases still remain unresolved. Another frontier to be explored is the noncoding regions on the X chromosome for regulatory variants which may contribute to the disease. This study also shows that even in a well characterized chromosome that has been for a long time associated with ID, there are still many loci/genes yet to be found related to ID, and emphasizes the inherent genetic heterogeneity in ID. X-chromosomal inheritance, however, cannot explain a large portion of the ID cases, and thus warrants a broader approach encompassing the autosome for ID research.

A recent study focused on targeted sequencing of 107 XLID genes in 50 patients with family history suggestive of XLID, and in 100 apparently sporadic ID cases (Tzschach et al., 2015). In 13 (26%) of the familial cases, likely pathogenic mutations were identified. However, only 5 pathogenic mutations were identified in the 100 sporadic patients.

1.3 Targeted Sequencing Studies

XLID has been a major focus in ID research, despite representing only 10%–12% of patients with ID (Ropers & Hamel, 2005). It has become apparent that autosomal forms of ID are more common than previously appreciated. Autosomal Recessive Intellectual Disability (ARID) has not been studied as extensively because Western countries tend to have smaller families and fewer consanguineous marriages. This has led to a gap in knowledge regarding the role recessive variants play in ID.

Najmabadi et al. (2011) sought to identify ARID genes by mapping 272 consanguineous Iranian families and determining the single linkage interval regions, in order to design a custom sequence capture system for these regions. Subsequently, 136 consanguineous families were sequenced using these custom generated gene panels which covered the genes in the single interval regions mapped in the consanguineous families.

The study identified 50 novel genes for ARID as well as 26 families that harbor known, mostly syndromic, forms of ARID. Najmabadi et al. (2011) also noted that only two of the novel genes were found in more than one family, which highlights the heterogeneity of ARID genes. As with XLID, identification of multiple ID families with mutations in a specific gene is important for validation as an ARID gene, and many of these genes are still awaiting supporting evidence from additional families. The variants detected for this cohort ranged from housekeeping genes such as *LARP7* (a negative transcriptional regulator of polymerase II) to variants in genes with no obvious link with ID, including *ADK*, which causes adenosine kinase deficiency. Several variants detected in the Najmabadi et al. (2011) study also had associations with brain-specific functions, including CACNA1G (a T-type calcium channel protein), and *ZBTB40* which has a role in glia differentiation.

This targeted approach was proven useful in identifying novel and rare variants associated with ID. It also had the benefit of reducing the cost of the massively parallel sequencing approach while increasing the number of samples screened. In recent years, targeted sequencing approaches have fallen out of favour due to the plummeting price of whole exome and whole genome sequencing. This approach, however, is biased to the targeted capture regions, and relies on having consanguineous families, preferably large, and with microarray data, and the good fortune to identify clear autozygous (homozygous-by-descent) regions. This bias may result in ID genes or unexpected variants being overlooked in some instances.

1.4 Whole-Exome Sequencing for Gene Identification

Whole Exome Sequencing (WES) targets only the 2% of the genome which codes for the proteins. Most disease-causing variants identified thus far occur within the protein coding regions of the genome, where variants are more likely to alter the protein's structure and thus resulting in dysfunction. This may also be due to difficulty in interpreting non-coding variants and will be an active area of research in the coming years. WES has become an excellent research and diagnostic tool which balances cost with disease-related information. Alongside pedigree and familial data, WES of mother/father/proband trios can be used to detect variants from different modes of inheritance as well as sporadic, *de novo* instances of disease mutations, in an unbiased fashion.

De novo mutations have been found to be more enriched in patients with ID than in control populations. In Western countries, where environmental and familial causes of ID are not obvious, *de novo* mutations may pose a plausible genetic model in many cases. Because of the lack of diagnostic, familial and environmental clues J. de Ligt et al. (2012) focused primarily on *de novo* mutations in their study after an assessment by clinical geneticists, 100 patients who lacked any diagnosis by copy number variant (CNV) analysis.. Both parents and probands were sequenced and the data was filtered based on mode of inheritance and predicted pathogenicity of the variants (based on algorithms such as SIFT and PolyPhen, as well as gene ontology). *De novo* mutations were identified in 53% of the patients sequenced and offered conclusive diagnosis for 13% of the patients sequenced. An additional 3% were found to be X-linked, and *de novo* mutations were found in 22 patients, with 3 being confirmed as having characteristic phenotypes among the confirmation series of 765 ID patients. All screening was done with high-

throughput sequencing, and variant prioritization as outlined by J. de Ligt et al. (2012). Some variants discovered in this study can lead to treatments such as antiepileptic and avoidance of sodium channel blockers for a patient with *SCNA2 de novo* mutations or dietary considerations for a patient with a *PDHA1* mutation. Clinical evaluation and counseling were offered to families once the study was completed to return findings and provide .

This study also showcased the usefulness of unbiased sequencing as a potential diagnostic tool for ID and other complex heterogeneous disorders. The need for a robust, scalable and integrated method for identifying causal variants, managing the data and disseminating the results as a national health care initiative is needed. The logistics of this type of undertaking are daunting especially when considering the scalable application of ethics, counseling and follow-up of patients afflicted with ID.

The use of WES in clinical practice has raised many questions concerning the appropriate methods to sift through genomic data for relevant pathogenic variants and the ethical delivery of results to patients and families in light of incidental findings. The Deciphering Developmental Disorders (DDD) group is a collaborative group in the United Kingdom that aims to translate high resolution genomic and phenotypic information into the National Health Service (NHS) (C. F. Wright et al., 2015). In conjunction with the UK NHS and the Irish Regional Genetics Service, the study C. F. Wright et al. (2015) recruited approximately 2000 families in the first year. Another 8000 families were recruited within the next 3 years. The first 1133 family trios collected and integrated recruitment, sample tracking, data generation, analysis, variant filtration and clinical feedback to patients with the approval of the UK Research Ethics Committee. A team of research coordinators was essential in procuring informed consent, and clinical geneticists entered phenotypic data after detailed assessments. After sample collection, whole genome microarrays were used for CNV detection and whole exome sequencing was performed on both parents and the proband. Variants detected in exome sequencing were disregarded if they were common (minor allele frequency >1%) and unlikely to disrupt protein function (e.g., synonymous changes filtered out). The pathogenic variants are then compared to the Developmental Disorders Genotype to Phenotype (DDG2P) in-house database. This database contains over 1000 genes that are continually being updated as patients are recruited and scientific literature advances.

Clinical results are evaluated by a multi-disciplinary team including clinical geneticists and genetic scientists in a weekly meeting. Results are provided to families and then are deposited on the DECIPHER database. In total, from ~80,000 genomic variants identified, 400 rare variants were predicted to be causal protein altering variants. The use of trios reduced the number of putative causal variants by tenfold and allowed a diagnostic yield of 27% by focusing on segregating and *de novo* variants.

Despite these advances, the majority of cases of ID remain with no identified genetic cause. WES represents 2% of the genome, but extending the unbiased approach of high-throughput sequencing has proven informative. With advances in whole genome sequencing (WGS), ID research may benefit from interrogating the non-coding regions of the genome.

1.5 Whole Genome Sequencing for Gene Identification

Whole Genome Sequencing (WGS) is the most comprehensive diagnostic test to date because it sequences the entire genome of an organism. C. Gilissen et al. (2014) recently sequenced 50 patients with severe ID and their parents, who had already undergone a comprehensive diagnostic workup including targeted sequencing, genomic microarray and WES without molecular diagnosis.

C. Gilissen et al. (2014) chose to focus on *de novo* CNVs and SNVs because of their known importance in ID and discovered on average 84 *de novo* CNVs and 82 *de novo* SNVs per genome. This accords well with previous studies. The authors found that there was a significant enrichment for loss of function (LoF) mutations, and mutations in known ID genes. WGS also allowed for the detection of structural variants and indels that may be outside the exons of ID genes. This was illustrated in a patient with a partial duplication of *TENM3* within the known ID gene *IQSEC2*. This duplication led to a gene fusion product that disrupted the function of *IQSEC2*, and was confirmed with RNA studies. Interestingly, three of the ten *de novo* SNVs were identified in a mosaic state in patients. WGS, as long as depth of coverage is sufficient, allows for the detection of mosaic variations as seen in a fraction of the reads from sequencing.

This study conclusively diagnosed 21 out of the 50 patients, resulting in a diagnostic yield of 42%. C. Gilissen et al. (2014) attempted to systematically characterize variants by selecting high confidence variants located in promoter, introns, and untranslated regions, such as those described by the ENCODE project, and validated 43 variants by Sanger Sequencing. The ENCODE project provided a rich set of resource for transcription factor binding sites and chromatin state segments in nine different human cell types (ENCODE Project Consortium, 2012) (ENCODE Project Consortium, 2012). However, data-mining using these resources still yielded no pathogenic information about the non-coding variants identified in this study. Scientific understanding of non-coding regulatory variation is still in its infancy and rigorous functional work will be needed to understand the role of non-coding genetic variation in disease.

There has been recent success in understanding regulatory variation in the human genome through computational techniques such as deep neural network learning algorithms and experimental functional work. L. Huang et al. (2012) used targeted massively paralleled sequencing to identify a non-coding, regulatory variant in the known ID gene *HCFC1*. Because no coding mutations had been identified despite the solid XLID mapping evidence, the authors hypothesized that the variant may reside outside of the coding region on the X chromosome. After extensive functional work, it was determined that this variant disrupted the YY1 transcription factor binding site. The authors also experimentally validated their claims to show that this variant reduced neurite growth and increased astrocyte production. Two further families with ID were also found to harbor this mutation.

WGS is the most exhaustive and complete diagnostic test available but it is hampered by the high cost of sequencing and difficulties in data analysis, management and storage. The real strengths of WGS are the sequencing of the non-coding regulatory regions of the genome, also the much greater utility of the data for identification of CNVs in comparison to WES data. WGS is still a cutting-edge technology and techniques to analyze WGS data are continually being developed but careful validation is required for clinical diagnostic implementation.

1.6 Structural Variant Detection through NGS for Gene Identification in Intellectual Disability

One of the major areas of development in NGS analysis is the detection of copy number variation (CNV). CNV analysis is still an active area of research in NGS variant analysis and has long been important in ID research.

Iqbal et al. (2012) encountered a case involving a patient with what they believed was a novel form of non-syndromic ID from a consanguineous family in Pakistan. The family had microarray SNP genotyping performed and there was a homozygous ~100 kb deletion on 2q37.1 found that did not segregate with the rest of the family. After homozygosity mapping, four autozygous regions were identified, but no obvious candidate genes within these regions. A targeted panel, designed for these homozygous regions was developed and the entire family was sequenced. There was a five-exon deletion in the *TPO* gene present which did segregate with the family. Reanalysis of the SNP microarray data (Affymetrix 250K) revealed that there was only a single SNP within this deletion region, and thus beyond the limits of CNV detection algorithms. This deletion could potentially have been identified using higher-density SNP arrays, but in this instance targeted exonic NGS provided the resolution required for detection. Subsequent validation was confirmed with qPCR. Iqbal et al. (2012) proved that the unbiased approach of NGS can be used to detect causal CNV variants for ID.

CNV analysis is routinely performed with CGH microarrays in diagnostic laboratories due to its low cost, robustness and reliability. The major drawback of this approach is that CGH microarrays can only detect unbalanced structural variants. Apparently Balanced Chromosomal Rearrangements (ABCR) occur in 1.54% of live births and contribute to 6% of abnormal phenotypes including multiple congenital abnormalities and ID (MCA/ID). Previous methods such as break point cloning using FISH and BAC, Southern Blot, inverse-PCR and long-range PCR are time consuming, labor-intensive and may not provide the desired resolution required for clinical diagnostics. Due to the above reasons, this area remains under investigated.

Schluth-Bolard et al. (2013) used WGS to characterize ABCR in four patients where CGH arrays had provided no diagnosis. NGS identified ABCR in three of the four patients including all breakpoints, inversions and balanced translocations. In three of the four patients, disruption of

genes (*SHANK2*, *TCF4*, *PPFIA1*, *RAB19*, and *KCNQ1*) was the probable cause of MCA/ID. The method was also validated with FISH, and PCR followed by Sanger Sequencing. The authors concluded that WGS is a reliable method for breakpoint mapping, but capture methods may have lower coverage around ABCR because they often occur near repetitive sequences, which are difficult to map back to the reference sequence.

Recently, many software programs have been developed to accurately detect CNVs and other structural variations such as CONIFER (O'Roak et al., 2012) and XHMM(Fromer et al., 2012; Fromer & Purcell, 2014; Poultney et al., 2013). It is an active area of research and large cohort studies need to be performed to determine the clinical relevance of NGS structural variation identification for ID.

1.7 NGS Strategies for Clinical Diagnostics

Due to a higher diagnostic yield and reduction in costs, NGS is increasingly being deployed in clinical laboratories for the diagnosis of genetic syndromes and cancer (Omoyinmi et al., 2014; Y. Yang et al., 2013). For individuals with ID, chromosomal microarray analysis (CMA) is recommended as the first-line diagnostic test as it offers much higher diagnostic yield (15%–20%) compared to G-banded karyotype analysis (3%)(D. T. Miller et al., 2010). However, the genetic etiology of 80% to 85% of patients still remains unknown. For these cases, NGS-based testing (targeted multi-gene panels, WES or WGS) has a great potential to achieve diagnosis. A recent study has shown a diagnostic yield of ~28% in patients affected with developmental delay using clinical WES (H. Lee et al., 2014). In another study, using clinical WES, molecular diagnosis was archived in ~25% cases with neurological phenotypes (Y. Yang et al., 2014). Using the NGS based targeted gene panel approach, Tan et al has identified the disease causing variants in 21% of the ID cases (Tan et al., 2015). These studies demonstrate the clinical utility of NGS based tests. Although, targeted gene panels and WES are the commonly used NGS tests in clinical laboratories, the WGS offers additional benefits as it has the potential to uncover all forms of genetic variation in one test (Pang, MacDonald, Yuen, Hayes, & Scherer, 2014), and consequently, it potentially offers a higher diagnostic yield. In a recent study of patients with severe ID, a diagnostic yield of 42% was observed, which is a significant improvement over diagnostic yield obtained by microarray, gene panels, or WES (C. Gilissen et al., 2014).

These observations indicate that NGS will undoubtedly have a remarkable impact for patients with ID and other genetic disorders. However, the widespread adoption of NGS technologies in diagnostic settings is hindered by several factors such as cost, processing time, clinical interpretation and storage of enormous amount of data. The American College of Medical Genetics and Genomics has issued the professional standards and guidelines for implementation of NGS in clinical laboratories (Rehm et al., 2013). In the light of these guidelines, Patwardhan et al. (2015) (Patwardhan et al., 2015) has evaluated the performance of the commercially available exome capture methods and sequencing technologies. Insufficient reads coverage was observed across several disease associated genes using each of the conventional exome capture and whole-genome platforms (Patwardhan et al., 2015). These findings highlight the need for further improvements in the exome capture methods, sequencing technologies and computational algorithms to improve the accuracy and sensitivity of NGS testing.

Finally, ethical concerns need also be considered for the application of NGS to ID in both clinical and research settings, in particular the possibility of incidental findings, as had been discussed in many articles [e.g., see H. C. Howard et al. (2015); Knoppers, Avard, Sénécal, and Zawati (2014). For gene-panel approaches, issues of incidental findings are obviously much less of a concern. Also, for trio-based NGS efforts aimed strictly at identifying *de novo* mutations, the concerns would possibly also be lower.

1.8 Aims and Hypothesis

1.8.1 Study One - Autosomal recessive variants in intellectual disability in consanguineous populations

Recessive variants have often been overlooked in neurodevelopmental disorders (NDD) including ID, due to the emphasis on association type studies and a majority of the research being performed in Western Countries. Association studies thus far have not had enough power to have a significant signal from studies. Historically, in literature, neurodevelopmental disorders and ID are characterized as having extreme phenotypic and genetic heterogeneity resulting in rare variants that require extremely large sample sizes to have sufficient power. The use of consanguineous families allows for enrichment of recessive variants and large families that allow the use of homozygosity mapping to identify causative variants. Next-Generation Sequencing has revolutionized and expedited disease gene identification and is currently the

method of choice for discovering novel molecular mechanisms. To focus on the interpretability of variant identification, whole-exome sequencing, where protein coding regions of the genome are sequenced, was used to sequence 192 large consanguineous multiplex families. The first study is driven by the hypothesis that autosomal recessive variants plays a significant and underappreciated role in ID and NDD. This study seeks to identify novel variants and their role in ID and NDD in both inbred and outbred populations. The identification of recessive variants should also be translatable to other populations with significant pleiotropy and genetic overlap with other disorders.

1.8.2 Study Two – Whole exome sequence analysis in autism spectrum disorder trios from consanguineous populations

There is a high degree of genetic overlap between ID and Autism Spectrum Disorder (ASD) and, very similar to ID, there is a gap in knowledge regarding the role in rare, recessive penetrating variants play in ASD. As mentioned, ASD is prevalent in North American countries, 1 in 66 in Canada, with a high degree of heritability. Most studies thus far have attempted to identify variants using genome wide association studies (GWAS) with large populations sizes. However, these studies have been mostly under powered in identifying recessive variants. This study has thus attempted to fill this gap in recessive variants for ASD by using whole exome sequencing and proband-parent trios from predominantly consanguineous regions to identify both novel genes and the role recessive variants play in ASD. The use of consanguineous trio families is expected to enrich for the recessive variants. In this study, I hypothesize that recessive variants also play an underappreciated and significant role in ASD, in addition, the variants identified in consanguineous families are translatable to outbred populations.

1.8.3 Study Three – *De novo* and inherited variants in ZNF292 underlie a neurodevelopmental disorder with features of autism spectrum disorder

There is significant genetic pleiotropy and heterogeneity in ID and ASD. Many of the critical genes for neurodevelopment are also critical in other molecular pathways. Genetic pleiotropy has presented significant barriers in understanding the role genes play in NDD. Deep phenotyping has to be performed to properly identify the role some genes play in NDD. The third study seeks to identify the phenotypic spectrum of ZNF292 in ID and ASD and more broadly NDD. ZNF292 encodes a zinc finger protein and notably this protein family acts as

transcription factors. The role of transcription factors in developing organisms can be diverse and dependent on the tissue where the protein is expressed. ZNF292 has been identified in previous studies as an enhancer of growth hormone expression in the pituitary gland of rat animal models. In addition, ZNF292 play a role in cancer, it has also been established where ZNF292 acts as a tumor suppressor. The role ZNF292 plays in neurodevelopment remains unknown, however, and this study seeks to characterize the clinical, neuroimaging and molecular features of a cohort of patients and further characterize the diverse clinical spectrum of ZNF292 in NDD.

2 Autosomal Recessive Variants in Intellectual Disability in Consanguineous Populations

This chapter is modified from the following published manuscripts:

Harripaul, R., Vasli, N., Mikhailov, A., Rafiq, M. A., Mittal, K., Windpassinger, C., ... & Johnson, M. (2018). Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. Molecular psychiatry, 23(4), 973-984.

Mittal, K., Rafiq, M. A., Rafiullah, R., Harripaul, R., Ali, H., Ayaz, M., ... & Rappold, G. A. (2016). Mutations in the genes for thyroglobulin and thyroid peroxidase cause thyroid dyshormonogenesis and autosomal-recessive intellectual disability. Journal of human genetics, 61(10), 867-872.

Sawal, H. A., Harripaul, R., Mikhailov, A., Vleuten, K., Naeem, F., Nasr, T., ... & Rafiq, M. A. (2018). Three Mutations in the Bilateral Frontoparietal Polymicrogyria Gene GPR56 in Pakistani Intellectual Disability Families. Journal of pediatric genetics, 7(02), 060-066.

Below is the link to the published manuscript:

https://www.nature.com/articles/mp201760

https://www.nature.com/articles/jhg201662/

https://www.thieme-connect.com/products/ejournals/html/10.1055/s-0037-1612591

Reprinted with permission from Nature Publishing Group and Thiemes

The three articles had significant contributions from myself in experimental design, data generation, data analysis, data interpretation, manuscript drafting and editing.

2.1 Abstract

Approximately 1% of the population worldwide are affected by intellectual disability (ID), the vast majority of whom currently receive no molecular diagnosis. Previous studies indicate high levels of genetic heterogeneity, with estimates of over 2500 autosomal ID genes, with the majority being autosomal recessive (AR). Here, we combined microarray genotyping, homozygosity-by-descent (HBD) mapping, copy number variation (CNV) analysis, and whole exome sequencing (WES) to identify disease genes/mutations for 192 multiplex Pakistani and Iranian consanguineous families ascertained for non-syndromic ID. We identified definite or candidate mutations (or CNVs) in 51% of families, in 72 different genes, including 26 not previously reported for ARID. The new ARID genes include 11 with loss-of-function mutations (*ABI2*, *CCDC82*, *MAPK8*, *C12ORF4*, *MPDZ, PIDD1*, *SLAIN1*, *TBC1D23*, *TRAPPC6B*, *UBA7*, and *USP44*), and missense mutations include the first reports of variants in *BDNF* or *TET1* associated with ID. The genes identified also showed overlap with *de novo* gene-sets for other neuropsychiatric disorders. Transcriptional studies show prominent expression in the prenatal brain. The high yield of AR mutations for ID signals that this approach has excellent clinical potential, will inform clinical diagnostics, including clinical whole exome and genome sequencing, for populations where consanguinity is common. As with other AR disorders, the relevance will also impact outbred populations.

2.2 Introduction

About 1% of the population worldwide is affected by intellectual disability (ID)(Maulik, Mascarenhas, Mathers, Dua, & Saxena, 2011) which can have a devastating impact on the lives of the affected individuals and their families and is a major challenge at the clinical level. Genetic factors are involved in the etiology of 25-50% of ID cases(McLaren & Bryson, 1987; Vissers, Gilissen, & Veltman, 2015). The clinical presentation and etiology of ID are complex and highly heterogeneous, leading to a poor rate of molecular diagnosis, resulting in inadequate clinical management and counselling.

ID can be divided into two groups: syndromic (S) ID, where comorbid illness or physical features are present, and nonsyndromic (NS) ID, where no such comorbidities are present. Of ~700 known ID genes (S and NS), fewer than 50 genes are mutated in NS autosomal recessive ID (NS-ARID)(Vissers et al., 2015). X-linked ID may account for only 10-12 % of ID cases(Ropers & Hamel, 2005). Dominant autosomal variants occurring *de novo* may contribute to a large proportion of sporadic cases, particularly in outbred populations(Joep de Ligt et al., 2012; Christian Gilissen et al., 2014; F. F. Hamdan et al., 2014). Autosomal recessive (AR) variants also play a significant role in ID because recessive variants can remain in the population in a heterozygous form. Estimates suggest that there may be more than 2500 autosomal ID genes in total— the majority being recessive(Musante & Ropers, 2014). In populations with high levels of consanguinity, most ID-causing mutations are recessive(Musante & Ropers, 2014). Even in outbred populations estimates are that 13-24% of ID cases are likely to be due to autosomal recessive causes(Musante & Ropers, 2014).

Despite an increased diagnostic yield among those who receive testing(Vissers et al., 2015) **the vast majority with ID currently receive no molecular diagnosis**(Rauch et al., 2006)**—a shortcoming that can impact health and lifespan. Recent studies indicate that the median age of death for males with ID is 13 years younger and for females 20 years younger, than in the general population**(Heslop et al., 2014)**.** Although advances in genotyping and sequencing technology have accelerated the rate of gene discovery for ID(Y. Yang et al., 2013), the majority of ID genes remain undetected. However, large scale ID family studies are making significant inroads(Najmabadi et al., 2011). Homozygosity mapping has been proven to be an effective method for gene identification in consanguineous populations(Ahmed et al., 2015; Bernkopf et al., 2014; Heidari et al., 2015; Law et al., 2014; Mir et al., 2009; Najmabadi et al., 2011; Muhammad Arshad Rafiq et al., 2011). Consanguineous marriages lead to a marked increase in the frequency of severe recessive disorders(Modell & Darr, 2002). Collectively, countries with levels of consanguinity higher than 10%, mainly from Africa, the Middle East and, South Asia(Hamamy, 2012; Saadat, 2007), represent a population of ~1.3 billion. In Pakistan, ~62.7% of the population engage in consanguineous marriages, of which ~80.4% are first cousin marriages(Hussain & Bittles, 1998) . The rate of consanguineous marriages in Iran is estimated at 40%(Najmabadi et al., 2007). Here we present a study of multiplex ID families from Pakistan (N=176) and Iran (N=16), using microarray-based genotyping to identify autozygous

regions (HBD shared by affected family members), coupled with whole exome sequencing (WES) to identify causal variants (see for workflow). In total, we have identified single candidate genes/variants for 88 families (50% were loss-of-function (LoF) mutations), and ten candidate pathogenic genomic variants (CNVs) in nine families.

2.3 Materials and Methods

2.3.1 Family recruitment

Institutional research ethics board consent was given for the study through the Centre for Addiction & Mental Health, Toronto, also through institutes at the recruiting sites (details in the Supplementary Methods). Families were recruited on the basis of diagnosis of ID in more than one individual (or ID and/or learning disability for N=13 families; or both ID and psychosis within the family, N=5), but with no obvious dysmorphic features or comorbidities, and with parental consanguinity. Typically, marriages were first or second cousin marriages, however in a number of families the exact relationship between the parents could not be established, but the marriage was within the same clan or caste. Written informed consent was obtained for all participants. Blood was drawn, and genomic DNA was extracted by standard methods. Cases of fragile X (tested using established methods(Steven Gazal, Sahbatou, Babron, Génin, & Leutenegger, 2014; Haddad, Mingroni-Netto, Vianna-Morgante, & Pena, 1996)), Down's syndrome and other clearly recognizable syndromes were excluded. Summary statistics for the families is given in Table 2.1.

Families were recruited from the following sites, following local REB approval and written consent: COMSATS Institute of Information Technology, Islamabad, Pakistan (N=8 families), National University of Sciences and Technology (NUST), Islamabad, Pakistan (N=14), International Islamic University, Islamabad, Pakistan (N=43), Quaid-i-Azam University, Islamabad, Pakistan (N=30), Lahore Institute of Research and Development, Lahore, Pakistan (N=81), Qazvin University of Medical Sciences, Qazvin, Iran (N=10), and Pasteur Institute of Iran, Tehran, Iran (N=6).

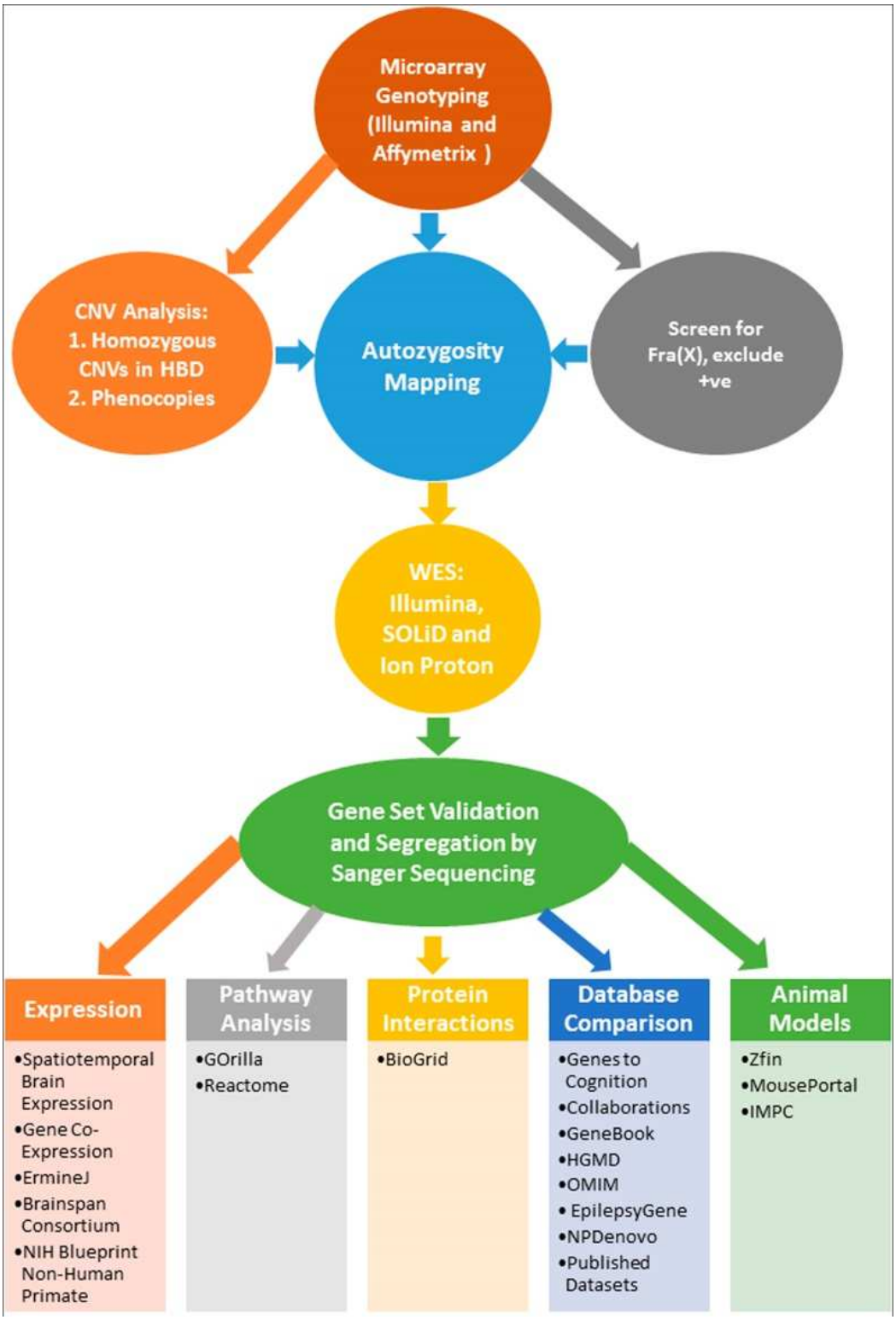


Figure 2.1: Microarray and exome sequencing work flow. CNV, copy number variations; HBD, homozygosity-by-descent; WES, whole exome sequencing.

2.3.2 Autozygosity mapping

Autozygosity mapping was performed using microarray data from the Illumina Human CoreExome, Affymetrix Mapping 500K *Nsp*I, Affymetrix CytoScan HD and Affymetrix SNP 5.0 or 6.0. All genotyping was performed as per manufacturer's protocol, and data processed using either the Affymetrix Genotyping Console (500K and 5.0), Affymetrix ChAS Software Suite (CytoScan HD and SNP 6.0) or the Illumina GenomeStudio platform (Illumina CoreExome). All data was exported to a plink format for analysis using HomozygosityMapper(Seelow, Schuelke, Hildebrandt, & Nürnberg, 2009) and FSuite(S. Gazal, Sahbatou, Babron, Genin, & Leutenegger, 2014). Details are provided in the Supplementary Methods.

Both of the below methods and the hg19 version of the genome were used to ensure a consistent and uniform genotyping.

Genotyping data was uploaded to the HomozygosityMapper server to determine putative homozygous-by-descent (HBD) regions based on the allele frequencies of the markers uploaded to the server from previous studies. HBD regions were identified by manual curation and only HBD regions larger than 1 Mb shared between all affected members (and not unaffected members) of the family were chosen. These regions were extracted based on SNP RS numbers and these dbSNP identifiers were converted to a genomic position for used to represent genomic regions with NGS data.

Previous methods used to identify HBD regions can vary among individual researchers due to the manual curation necessary to ensure the selected regions share both homozygosity and haploidentity between affecteds but not unaffecteds (i.e. to exclude false positives), and to try to limit false negative calls. These methods are also labor intensive and do not automatically take in family information. For these reasons, we used a software called FSuite for computational HBD mapping(S. Gazal et al., 2014). FSuite makes use of allele frequencies, genotyping and pedigree structure integrated into a Hidden Markov model to determine the likelihood of a region being HBD(S. Gazal et al., 2014). In-house developed scripts were used to convert genotyping calls from the Affymetrix Genotyping Console, ChAS and Illumina GenomeStudio output to a family based PLINK input for FSuite. We also used 1000 Genomes Phase Three reference for the allele frequencies, extracting allele frequencies from the South Asian ethnic group

specifically. This data was then used with standard run protocols of FSuite to generate HBD regions. Autozygosity mapping was done by extracting regions common between members of the family. These protocols were standardized and implemented on the Specialized Computing Cluster (SCC) at the Centre for Addiction and Mental Health (CAMH).

Table 2.1: Summary statistics for Iranian and Pakistani family cohorts.

Source Country	Total Families	Total DNA Samples	Total Affected	Mean Number of ID Affected/ Family	*F* Coefficient of inbreeding	No. of Families with Gene/Mutation Identified: ARID (XLID; *de novo*)	Number of Families with Pathogenic CNV Identified
Iran	16	193	58	4.375	0.058034056	6 (0;0)	0
Pakistan	176	1379	627	3.5625	0.084862578	81 (6; 1)	9[a]

[a]In family PK73 two different likely pathogenic CNV losses were identified, but only counted here as a single family with pathogenic CNV.

2.3.3 Structural Variant Analysis using Microarray

CNV analysis was performed to identify homozygous CNVs in HBD regions, or heterozygous CNVs that could indicate cases of intra-familial genetic heterogeneity, that should then be excluded from the HBD/WES analyses (details in the Supplementary Methods).

CNV analysis was performed to identify homozygous CNVs in HBD regions, or heterozygous/ homozygous CNVs that could indicate phenocopies that should then be excluded from the HBD/WES analyses. We have used five array types for homozygosity mapping including the Affymetrix SNP 5.0, SNP 6.0, CytoscanHD and Mapping 250K NspI, as well the Illumina HumanCoreExome platform. Due to the wide range of array types that were used in this study, we have used manufacturer-recommended and third-party software to call CNVs. The Affymetrix Genotyping Console was used to genotype and calls CNVs using the Affymetrix Birdsuite algorithm. CytoscanHD and SNP 6.0 arrays were analyzed and CNVs were called using the Chromosome Analysis Suite (ChAS). Illumina HumanCore Exome arrays were used for the largest proportion of families, and CNVs were called using the Illumina GenomeStudio cnvPartition plugin. PennCNV is an academically developed software that uses a Hidden Markov Method (HMM) to call CNVs(K. Wang et al., 2007). CNVs were called using PennCNV for all array types and represents a standard method for CNV calling across our study. CNVs derived from all algorithms were compared against the Database of Genomic Variants

(DGV) to screen for variants that showed 30% or lower overlap with DGV control variants, or were not in the database(MacDonald, Ziman, Yuen, Feuk, & Scherer, 2013).

2.3.4 Whole exome sequencing

WES was performed for one or more affected members from each family, using sequencing facilities at CAMH, Toronto. Three different next generation sequencing platforms were used across the study, taking advantage of newer platforms as they became available for research at CAMH (SOLiD 5500 platform (Life Technologies) for 49 families (51 individuals), using a protocol reported previously(Mir et al., 2014); Ion Proton platform/Ion Ampliseq™ Exome kit (Life Technologies) for 49 individuals from 30 families; Illumina HiSeq2500 platform, using ThruPLEX DNA-seq 96D kit (Rubicon, R400407) and SureSelect XT2 Target Enrichment (Agilent Technologies) system for 150 families).

2.3.5 Sequencing alignment and variant calling

We used an in-house pipeline to map and call variants on the different types of sequencing data for this study. The pipeline is summarized in Supplementary analysis and in 7.1.1.4.

Once putative variants are identified, Sanger sequencing was used to validate variants and determine if the variant segregates with the disease by testing parents, affected and unaffected individuals within the family. Additional members can be tested if the family had branches or other related affected families members

Illumina reads were aligned with BWA-mem (v 0.7.13) using the hg19 genome reference(H. Li, 2013). Once alignment was performed GATK (v 3.2.2) was used for base recalibration, indel realignment and the GATK Unified Genotyper was used for variant calling. Variants called were generated in the standard VCF version 4.1 and annotated with Annovar(H. Yang & Wang, 2015), Alamut (v 1.4.4 Interactive Biosoftware, Rouen, France) and EDGC Annotator (v 1.0.0 Eone-Diagnomics Genome Center, Songdo Ichon, South Korea) software.

SOLiD sequencing was aligned to the hg19 reference genome using Shrimp2(David, Dzamba, Lister, Ilie, & Brudno, 2011). Once alignment files were generated, variant calling and annotation were performed as previously described. Proton data was aligned using the TMAP algorithm and variants were called using the Ion Proton Variant Caller as part of the Torrent

Server analysis pipeline (ThermoFisher Scientific, Waltham, MA, US). Ion Reporter was used in addition to the annotation software mentioned above for Ion Proton VCF files (ThermoFisher Scientific, Waltham, MA, US).

2.3.6 Variant Filtration

Once annotation was performed, filtration of variants was performed by computationally and manually extracting variants that were in HBD regions provided by the autozygosity mapping of FSuite and HomozygosityMapper. The next step was to filter variants based on sequencing quality, homozygosity of allele and mutation type. Prioritization was performed by scoring truncating mutations such as stop and frameshift loss of function mutations higher than missense mutations and inframe indels. Missense mutations were scored based on Sift and Polyphen scores for prioritization. Allele frequencies were then used to filter out variants that were too common in the population (>1 in 10 000). All variants surviving filtration were checked against the Exome Aggregation Consortium (ExAC), Cambridge, MA (URL: http://exac.broadinstitute.org) accessed 03,2016) OMIM(Hamosh, Scott, Amberger, Bocchini, & McKusick, 2005) and Genecards(Safran et al., 2010) to determine expression level in tissues and other functional characteristics.

2.3.7 Gene list construction

We combined the genes listed in Table A- 2 with other genes for X-linked or autosomal recessive intellectual disability/mental retardation in OMIM. This full list is available as Table A- 3. This gene list was used in the pathway and gene expression analyses (7.1.1).

2.3.8 Database searches across different neuropsychiatric and Neurodevelopmental Disorders and Functional and Animal Models

We compared our variants to several databases including EpilepsyGene, Gene2Phenotype, Gene2Cognition, Schizophrenia Genebook, published disease-specific gene-sets(Christian Gilissen et al., 2014), HGMD(Stenson et al., 2003), OMIM(Hamosh et al., 2005), Zfin(Bradford et al., 2011), targets of FMRP through high-throughput sequencing of RNAs by cross-linking immune-precipitation (HITS-CLIP) (Darnell et al., 2011), and Mouseportal (http://www.sanger.ac.uk/science/collaboration/mouse-resource-portal, accessed May 2016). Searches were performed to identify genes/variants overlapping with our dataset.

2.3.9 Anatomical expression analyses

To look for commonalities in the anatomical and temporal expression profiles for the genes identified here and previously reported genes for ID, we used a number of human expression datasets, tools, and atlases.

The TopAnat Bgee tool was used to quantify tissue specific expression patterns before birth in human expression datasets (Bastian et al., 2008).TopAnat uses publicly available expression profiles from reference atlases of normal tissue. Within a tissue, it tests for expression enrichment using binary calls of expression (expressed/present versus not). The ID Gene list was converted to Ensembl identifiers using the Ensembl Biomart tool (Ensembl release 84). The embryo developmental stage, RNA-Seq and Affymetrix data source settings were used with the "Data quality" option was set to "All". We removed redundant tissue terms and only present data from tissues with direct expression information (17 human fetal tissues).

2.3.10 Developmental brain expression analyses

Human spatiotemporal transcriptomic datasets were downloaded from the BrainSpan website (http://www.brainspan.org/static/download.html, December 2015). In this developmental atlas, age ranged from 8 post conception weeks to 40 years old (46% female). Brain specimens were from normal healthy donors that passed several strict exclusion criteria to ensure consistency. For example, specimens with evidence of malformations, lesions, neuronal loss, neuronal swelling, glioneuronal heterotopias, or dysmorphic neurons and neurites were excluded. Complete details of the methods used in the BrainSpan atlas are available as a technical paper (http://help.brain-map.org/download/attachments/3506181/Transcriptome_Profiling.pdf).

Both RNA sequencing and exon microarray gene expression profiles were used to test for specific temporal and regional expression of the ID genes. In both datasets we removed data from brain regions with expression profiles for less than 7 donors (10 regions). These regions are early structures that were dissected at early timepoints (earlier than 14 post conception weeks). After this filtering, 465 expression profiles from 35 brains (27 timepoints) in 16 unique regions were used from the exon array data. In the RNA sequencing data, 553 expression profiles from 41 brains (30 timepoints) in 16 unique regions remained. Expression data was previously summarized to genes by the BrainSpan consortium (exon array: 22,255, RNA sequencing:

17,282 genes). The provided RPKM (Reads Per Kilobase of transcript per Million mapped reads) expression values were log scaled.

Human prefrontal cortex expression profiles(Colantuoni et al., 2011) were downloaded from the supplemental files in "Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex"(Jaffe et al., 2015). This microarray dataset contains gene expression profiles of the dorsolateral prefrontal cortex across development. Provided mappings were used to link probes to genes (31,699 probes). Expression was mean averaged across multiple probes for a given gene. We limited our analyses to donor brains under 5 years old at death (supported by our results in the BrainSpan atlas of the developing brain). After this processing, 67 samples (34 timepoints) with information for 18419 genes were used in our analyses.

Expression values were z-score normalized across samples in each of the above datasets. Specifically, for each gene, its expression values were scaled (mean=0, standard deviation=1). This allows identification of region and timepoint specific expression patterns that are independent of absolute expression levels.

2.3.11 Pathway analyses

We used the Gene Ontology enRIchment anaLysis and visuaLizAtion tool (GOrilla) to test for enrichment of Gene Ontology (GO) gene groups in the ID gene list(Ashburner et al., 2000; Eden, Navon, Steinfeld, Lipson, & Yakhini, 2009). All HGNC gene symbols were used as the background gene list and default options were used (biological process ontology). The GOrilla database was last updated on May 14, 2016 when we ran the analyses.

We have compared our variants to several databases including EpilepsyGene, Gene2Phenotype, Gene2Cognition, Schizophrenia Genebook, published disease-specific gene-sets(Christian Gilissen et al., 2014), HGMD(Stenson et al., 2003), OMIM(Hamosh et al., 2005), Zfin(Bradford et al., 2011) and Mouseportal (http://www.sanger.ac.uk/science/collaboration/mouse-resource-portal, accessed May 2016). Searches were performed to identify overlapping genes/variants with our dataset.

2.4 Results

Our study has identified single candidate genes/variants for 88 (81 autosomal recessive, 6 X-linked, and 1 *de novo* (heterozygous)) of the 192 families (Table 2.2,Table 2.3 & Table A- 3). Twenty-six of the genes identified in this cohort have not previously been reported for NS-ARID. An additional eleven genes were first reported previously from this cohort (*CC2D2A(Noor et al., 2008)*; *TCTN2(Sang et al., 2011)*; *TRAPPC9(Mir et al., 2009)*; *MAN1B1(Muhammad Arshad Rafiq et al., 2011)*; *FBXO31(Mir et al., 2014)*; *METTL23(Bernkopf et al., 2014)*; *FMN2(Law et al., 2014)*; *DCPS(Ahmed et al., 2015)*; *HMNT(Heidari et al., 2015)*; *NSUN2*(Muzammil Ahmad Khan et al., 2012); *MBOAT7(Johansen et al., 2016)*[l]). Most of these genes have since been reported in multiple ID families, including in outbred populations (e.g. *TRAPPC9*(Marangi et al., 2013); *MAN1B1*(Van Scherpenzeel et al., 2014)). Candidate CNVs were identified in at least nine families.

Table 2.2: List of sequence variants identified for the Pakistani and Iranian families in the study. Families for which a single homozygous (or hemizygous for ChrX variants) mutation was identified, and no homozygous loss of function variants for the gene have been reported among ExAC database of control exomes (unless otherwise indicated). ExAC database minor allele frequencies (MAF) are indicated. Variants highlighted in bold are previously unreported in association with ID, and asterisks indicate genes/variants already reported by us in this cohort. A list of all variants identified including known syndromic ARID or XLID genes identified in the cohort is give in **Table A- 3**, and **Table A- 3** for families where >1 variant was identified. For subjects where the phenotype is clearly NS-ARID, despite prior association of the gene with S-ARID (e.g. *LRP2* and *MECP2*), we have included the variant in this list. See **Table A- 3** for *in silico* predictions of effects of variants.

	Family	Gene	Disease; MIM #	Genome (ExAC MAF/S.Asian MAF)	cDNA (Protein)	Ref
1	**PJ7**	***LAMC1***		**Chr1:183083732A>T (0/0)**	**NM_002293.3:c.1088A>T (p.His363Leu)**	
2	AN49	*FMN2**	MRT47; 616193	Chr1:240370627_240370628delinsG (0/0)	NM_020066.4: c.2515_2517delinsCG (p. Thr839Argfs*48)	*(Law et al., 2014)*
3[a]	AS114	*FMN2*	MRT47; 616193	Chr1:240370641_240370642delCT (0/0)	NM_020066.4: c.2529_2530delCT (p. Ser844Cysfs*6)	
4	**PK113**	***AFF3***		**Chr2:100167973C>A (0/0)**	**NM_002285.2: c.3644G>T (p. Gly1215Val)**	
5	PJ12	*LRP2**	222448	Chr2:170027106C>T (2.47E-04/9.08E-04)	NM_004525.2: c.11335G>A (p. Asp3779Asn)	Vasli, 2016 #257}
6	IDH10	*HNMT**	MRT51; 616739	Chr2:138727776G>A (1.66E-05/an)	NM_006895.2: c.179G>A (p. Glu60Asp)	*(Heidari et al., 2015)*
7[b]	**AS61**	***ABI2***		**Chr2:204245039C>T (0/0)**	**NM_005759.4: c.394C>T (p. Arg132*)**	
8[c]	**PK34**	***UBA7***		**Chr3:49848458C>A (8.01E-04/5.45E-03)**	**NM_003335.2: c.1189G>T (p. Glu397*)**	
9	**PJ9**	***TBC1D23***		**Chr3:100037953delA (0/0)**	**NM_018309.4: c.1683delA (p. Glu562Argfs*3)**	
10	**AN50**	***ZBTB11***		**Chr3:101371344A>C (0/0)**	**NM_014415.3: c.2640T>G (p. His880Gln)**	
11	**AS70**	***MAP3K7***		**Chr6:91254333C>T (1.65E-05/1.21E-04)**	**NM_145331.2: c.1229G>A (p. Arg410Gln)**	
12[d]	PK11	*TUSC3*	MRT7; 611093	Chr8:15480694C>T (0/0)	NM_006765.3: c.244C>T (p. Arg82*)	Garshasbi, 2008 #86}
13	AN21	*TUSC3**	MRT7: 611093	Chr8:15,521,688_15,692,362del (N/A)	NM_178234 del exons 5-10	Khan, 2011 #66}

14	PK54	*IMPA1*		Chr8:82583195_82583196delinsC (0/0)	NM_005536.3: c.544_545delinsG (p. Leu182Valfs*54)	(Figueiredo et al., 2015)
15	AS19	*TRAPPC9**	MRT13; 613192	Chr8:141407724G>A (8.24E-06/6.06E-05)	NM_031466.7: c.1423C>T (p. Arg475*)	*(Mir et al., 2009)*
16[e]	**AS66**	***MPDZ***		**Chr9:13109992_13109995delGAAA (0/0)**	**NM_003829.4: c.5811_5814delTTTC (p. Phe1938Leufs*3)**	
17	**AS22**	***MAPK8***		**Chr10:49628265_49628265delC (0/0)**	**NM_002750.3: c.518delC (p. Arg174Glyfs*8)**	
18	**PK70**	***TET1***		**Chr10:70451328G>T (0/0)**	**NM_030625.2: c.6168G>T (p. Lys2056Asn)**	
19	**IDSG19**	**DMBT1**		**Chr10:124356559C>G (0/an)**	**NM_007329.2: c.2906C>G (p. Thr969Arg)**	
20[f]	**AS105 & 110**	***PIDD1***		**Chr11:799453G>A (0/0)**	**NM_145886.3: c.2587C>T (p. Gln863*)**	
21	**PK135**	***BDNF***		**Chr11:27679747A>G (1.65E-04/1.21E-03)**	**NM_001709.4: c.365T>C (p. Met122Thr)**	
22	AS102	*PGAP2*	MRT17; 614207	Chr11:3846254G>C (4.95E-05/3.64E-04)	NM_014489.3: c.713G>C (p. Arg238Pro)	Y
23	PK62	*TMEM135*		Chr11:87032300delTT (0/0)	NM_022918.3: c.1304_1305delTT (p. Phe435Serfs*31)	(Najmabadi et al., 2011)
24	**AS17**	***CCDC82***		**Chr11:96117377G>A (1.65E-05/6.05E-05)**	**NM_024725.3: c.535C>T (p. Arg179*)**	
25	PJ5	*DCPS**	ARS; 616459	Chr11:126208295G>A (8.25E-06/0)	NM_014026.3: c.636+1G>A (15 amino acid insertion)	*(Ahmed et al., 2015)*
26	**PK94**	***USP44***		**Chr12:95927147_95927160delinsA (0/0)**	**NM_032147.3: c.873_886delinsT (p. Leu291Phefs*8)**	
27	**AS23**	***SPATA13***		**Chr13:24797332C>T (4.71E-05/0)**	**NM_001166271.1: c.265C>T (p. Arg89Trp)**	
28	**PK68**	***SLAIN1***		**Chr13:78320722_78320722delA (0/0)**	**NM_144595.3: c.135delA (p. Thr49Hisfs*96)**	
29	**IDH3**	***TRAPPC6B***		**Chr14:39628712G>A (1.65E-05/an)**	**NM_177452.3: c.124C>T (p. Arg42*)**	
30	**PK79**	***VPS35***	**PARK17; 614203**	**Chr16:46705735T>G (0/0)**	**NM_018206.4: c.1406A>C (p. Gln469Pro)**	
31[g]	PK12	*GPT2*	MRT49; 616281	Chr16:46918865C>G (0/0)	NM_133443.2: c.238C>G (p. Gln80Glu)	Y
32	PK85	*GPT2*	MRT49; 616281	Chr16:46956326C>T (3.33E-05/1.89E-04)	NM_133443.3: c.1210C>T (p. Arg404*)	Y

33	AS72	*FBXO31**	MRT45; 615979	Chr16:87369054_87369059delinsT (0/0)	NM_024735.3: c.847_852delinsA (p. Cys283Asnfs*81)	*(Mir et al., 2014)*
34	**PK91**	***SYNRG***	**607291**	**Chr17:35896199C>T (8.66E-06/0)**	**NM_007247.4: c.3548G>A (p. Arg1183His)**	
35	**IDSG29**	***FBXO47***		**Chr17:37107906G>C (0/an)**	**NM_001008777.2: c.544C>G (p. Arg182Gly)**	
36	PK33	*NAGS*	NAGSD; 237310	Chr17:42085083C>T (7.74E-05/4.54E-04)	NM_153006.2: c.1393C>T (p. Arg465Trp)	Y
37	**AN53**	***SDK2***		**Chr17:71431658C>T (5.82E-05/0)**	**NM_001144952.1: c.1126G>A (p. Gly376Ser)**	
38	PK31	*METTL23**	MRT44; 615942	Chr17:74729449C>T (0/0)	NM_001080510: c.397C>T (p. Gln133*)	*(Bernkopf et al., 2014)*
39	**AN48**	***CAPS***		**Chr19:5914970G>A (2.17E-04/1.27E-04)**	**NM_004058.3: c.281G>A (p. Arg94Gln)**	
40	AN37	*TRMT1*		Chr19:13223779_13223810del (0/0)	NM_017722.3: c.657_688del (p. Gln219Hisfs*22)	(Najmabadi et al., 2011)
41	**PJ3**	***MBOAT7***		**Chr19:54684518_54684524del (0/0)**	**NM_024298.3: c.820_826del (p. Gly274Profs*47)**	
42	**PJ6**	***GPR64***		**ChrX:19055718T>C (8.45E-05/0)**	**NM_005756.3: c.191A>G (p. Asn64Ser)**	
43	AS101	*ARHGEF6*	MRX46; 300436	ChrX:135829705A>G (1.14E-05/9.91E-05)	NM_004840.2: c.296T>C (p. Val99Ala)	
44	**PK87**	***MAGEA11***		**ChrX:148796213C>T (0/0)**	**NM_005366.4: c.169T>C (p. Ser57Pro)**	
45	PK55	*MECP2*	MRXS13; 300055	ChrX:153296710C>T (0/0)	NM_004992.3: c.569G>A (p. Arg190His)	Y

a. [1] Affected individual in AS114 with *FMN2* mutation is a phenocopy, and genes/mutations for the other four affected individuals in this pedigree have yet to be established.

b. The homozygous nonsense mutation reported for family AS61 is located within a large HBD region shared between just two of the four affected individuals. Support for this gene comes from functional studies, including with *Abi2* null mice (Grove et al, 2004).

c. Two homozygotes for this variant are indicated in the ExAC South Asian population (N=8,251 individuals). N.B. affected individuals in PK34 have learning disability rather than ID. The South Asian cohort used in ExAC were part of the Pakistan Risk of Myocardial Infarction Study (PROMIS), and would not have excluded subjects in this category.

d. Homozygous nonsense mutation in *TUSC3* is not in shared HBD region, and does not segregate with other affected members of family PK11.

e. The *MPDZ* mutation reported here is not within a significant region of HBD, and is present as homozygous in the affected mother and the first-born affected son. The affected brother and sister are both heterozygous, but the father is wild type homozygous, suggesting non paternity for the first child.

f. Families AS105 and AS110 both carry the same mutation in *PIDD1*, and share a common haplotype, however, these families are apparently not closely related.

g. Missense change in *GPT2* is also predicted to affect a splice donor site (BDGP (www.fruitfly.org) splice predictor score: WT: 0.78; mut: 0.58).

an=not applicable, as Iranian exome data is not available in ExAC database.

2.4.1 Variants Identified for Intellectual Disability

Our study has identified single candidate genes/variants for 88 (81 autosomal recessive, 6 X-linked, and 1 *de novo* (heterozygous)) of the 192 families (Table 2 & 3, & Table A- 3). Twenty-six of the genes identified in this cohort have not previously been reported for NS-ARID. An additional eleven genes were first reported previously from this cohort (*CC2D2A(Noor et al., 2008)*; *TCTN2(Sang et al., 2011)*; *TRAPPC9(Mir et al., 2009)*; *MAN1B1(Muhammad Arshad Rafiq et al., 2011)*; *FBXO31(Mir et al., 2014)*; *METTL23(Bernkopf et al., 2014)*; *FMN2(Law et al., 2014)*; *DCPS(Ahmed et al., 2015)*; *HMNT(Heidari et al., 2015)*; *NSUN2*(Muzammil Ahmad Khan et al., 2012); *MBOAT7(Johansen et al., 2016)*[1]). Most of these genes have since been reported in multiple ID families, including in outbred populations (e.g. *TRAPPC9*(Marangi et al., 2013); *MAN1B1*(Van Scherpenzeel et al., 2014)). Candidate CNVs were identified in at least nine families.

2.4.2 Gene list Construction

We combined the genes listed in Table A- 3 with other genes for X-linked or autosomal recessive intellectual disability/mental retardation in OMIM. This gene list was used in the pathway and gene expression analyses (details in 7.1.1).

2.4.3 Homozygous loss-of-function Mutations

Homozygous Loss-of-Function Mutations

Homozygous truncating LoF mutations were found as single candidate variants in 43 families, including 11 genes that are previously unpublished in relation to NS-ARID, namely *ABI2*, *C12ORF4*, *CCDC82*, *MAPK8*, *MPDZ, PIDD1*, *SLAIN1*, *TBC1D23*, *TRAPPC6B*, *UBA7*, and *USP44.*

We identified the same nonsense mutation in *PIDD1* in two unrelated Pakistani families (AS105 and AS110;Figure 2.2). *PIDD1* encodes p53-Induced Death Domain Protein (PIDD; MIM 605247). Gln863* disrupts the death domain (DD), through which PIDD1 interacts with other DD proteins such as RIP1 or CRADD/RAIDD. Truncating mutations in *CRADD* have previously been reported for NS-ARID (MRT34(Puffenberger et al., 2012)), and thus our findings add support to the involvement of PIDD-related pathways, in the etiology of intellectual disability.

Homozygous truncating mutations in *TRAPPC9* have been reported for NS-ARID (Mir et al., 2009),(Jamra et al., 2011; Kakar et al., 2012; Marangi et al., 2013; Mochida et al., 2009; Philippe et al., 2009), thus it is of much interest that we have identified a nonsense mutation in a gene encoding a second member of the same protein trafficking particle complex, namely *TRAPPC6B*.

A nonsense mutation was identified in *SLAIN1* segregating fully in family PK68 (Figure 2.2). *SLAIN1* and *SLAIN2* encode microtubule plus-end tracking proteins that have been shown to be crucial to normal axonal growth in developing hippocampal neurons(van der Vaart et al., 2012).

We report here a homozygous nonsense mutation in the gene *UBA7* for family PK34. *UBA7* encodes ubiquitin-activating enzyme 7, believed to be involved in the ubiquitin conjugation pathway(Figueiredo et al., 2015). Family PK34 is one of three families in the study that did not meet criteria for ID, and instead were reported as having a learning disorder, and are considered relatively high functioning. We also report that this variant is present at a relatively high frequency in the South Asian population (MAF=0.0054; ExAC database), and two homozygotes for this variant were among the control group. Thus, this variant and gene could be a risk factor for a much milder form of cognitive disability, and thus potentially present in the control South Asian population (N>8,000) used in the ExAC database.

We also identified a LoF mutation in *TMEM135*, previously been reported in the large Iranian NS-ARID cohort by Najmabadi et al, 2011 (in which a missense mutation, Cys228Ser was reported)(Najmabadi et al., 2011), and in Wright et al, 2014 (missense)(C. F. Wright et al., 2015), also as a quantitative trait locus for intelligence(Franić et al., 2015).

Table 2.3: Candidate genes and sequence variants identified for the Pakistani and Iranian ID families in support of previously reported non-syndromic ID genes. Asterisks indicate genes/variants already reported by us in this cohort For subjects where the phenotype is clearly NS-ARID, despite prior association of the gene with S-ARID (e.g. *LRP2* and *MECP2*), we have included the variant in this list. See **Table A- 1** for *in silico* predictions of effects of variants.

	Family	Gene	Disease; MIM #	Genome (ExAC MAF/S.Asian MAF)	cDNA (Protein)	Ref
1	AN49	*FMN2**	MRT47; 616193	Chr1:240370627_240370628delinsG (0/0)	NM_020066.4: c.2515_2517delinsCG (p. Thr839Argfs*48)	*(Law et al., 2014)*
2[a]	AS114	*FMN2*	MRT47; 616193	Chr1:240370641_240370642delCT (0/0)	NM_020066.4: c.2529_2530delCT (p. Ser844Cysfs*6)	
3	PJ12	*LRP2**	222448	Chr2:170027106C>T (2.47E-04/9.08E-04)	NM_004525.2: c.11335G>A (p. Asp3779Asn)	(Vasli et al., 2015)
4	IDH10	*HNMT**	MRT51; 616739	Chr2:138727776G>A (1.66E-05/an)	NM_006895.2: c.179G>A (p. Glu60Asp)	(Heidari et al., 2015)
5[b]	PK11	*TUSC3*	MRT7; 611093	Chr8:15480694C>T (0/0)	NM_006765.3: c.244C>T (p. Arg82*)	(Garshasbi et al., 2008)
6	AN21	*TUSC3**	MRT7: 611093	Chr8:15,521,688_15,692,362del (N/A)	NM_178234 del exons 5-10	(Garshasbi et al., 2008)
7	PK54	*IMPA1*		Chr8:82583195_82583196delinsC (0/0)	NM_005536.3: c.544_545delinsG (p. Leu182Valfs*54)	(Sjøholt et al., 2004)
8	AS19	*TRAPPC9**	MRT13; 613192	Chr8:141407724G>A (8.24E-06/6.06E-05)	NM_031466.7: c.1423C>T (p. Arg475*)	(Mir et al., 2009)
9	AS102	*PGAP2*	MRT17; 614207	Chr11:3846254G>C (4.95E-05/3.64E-04)	NM_014489.3: c.713G>C (p. Arg238Pro)	(Krawitz et al., 2013)
10	PK62	*TMEM135*		Chr11:87032300delTT (0/0)	NM_022918.3: c.1304_1305delTT (p. Phe435Serfs*31)	
11	AS17	*CCDC82*		Chr11:96117377G>A (1.65E-05/6.05E-05)	NM_024725.3: c.535C>T (p. Arg179*)	(Riazuddin et al., 2016)
12	ZA5	*C12ORF4*		Chr12:4599717delA (0/0)	NM_020374.2: c.1537delT (p. Ser513Leufs*7)	(Philips et al., 2016)
13	PJ5	*DCPS**	ARS; 616459	Chr11:126208295G>A (8.25E-06/0)	NM_014026.3: c.636+1G>A (15 amino acid insertion)	(Vincent, 2013)
14[c]	PK12	*GPT2*	MRT49; 616281	Chr16:46918865C>G (0/0)	NM_133443.2: c.238C>G (p. Gln80Glu)	(Celis et al., 2015; Ouyang et al., 2019)
15	PK85	*GPT2*	MRT49; 616281	Chr16:46956326C>T (3.33E-05/1.89E-04)	NM_133443.3: c.1210C>T (p. Arg404*)	(Celis et al., 2015; Ouyang et al., 2019)
16	AS72	*FBXO31**	MRT45; 615979	Chr16:87369054_87369059delinsT (0/0)	NM_024735.3: c.847_852delinsA (p. Cys283Asnfs*81)	(Mir et al., 2014)

17	PK33	*NAGS*	NAGSD; 237310	Chr17:42085083C>T (7.74E-05/4.54E-04)	NM_153006.2: c.1393C>T (p. Arg465Trp)	(van Karnebeek & Stockler, 2012)
18	PK31	*METTL23**	MRT44; 615942	Chr17:74729449C>T (0/0)	NM_001080510: c.397C>T (p. Gln133*)	(Bernkopf et al., 2014; Reiff et al., 2014)
19	IAID3	*METTL23**	MRT44; 615942	Chr17: 74729211_74729215del (0/0)	NM_001080510.3: c.236_240del (p. Thr80Glyfs*20)	(Bernkopf et al., 2014; Reiff et al., 2014)
20	AN37	*TRMT1*		Chr19:13223779_13223810del (0/0)	NM_017722.3: c.657_688del (p. Gln219Hisfs*22)	(K. Zhang et al., 2020)
21	PJ3	*MBOAT7*		Chr19:54684518_54684524del (0/0)	NM_024298.3: c.820_826del (p. Gly274Profs*47)	(Johansen et al., 2016)
22	AS101	*ARHGEF6*	MRX46; 300436	ChrX:135829705A>G (1.14E-05/9.91E-05)	NM_004840.2: c.296T>C (p. Val99Ala)	(Kutsche et al., 2000)
23	PK55	*MECP2*	MRXS13; 300055	ChrX:153296710C4T (0/0)	NM_004992.3: c.569G4A (p. Arg190His)	(Amir et al., 1999)

Abbreviations: HBD, homozygosity-by-descent; ID, intellectual disability. Asterisks indicate genes/variants that we have previously reported in this cohort. For subjects in whom the phenotype was clearly NS-ARID, despite prior association of the gene with S-ARID (e.g. LRP2 and MECP2), we have included the variant in this list. See Table A- 3 for in silico predictions of the effects of the variants. an=not applicable, as Iranian exome data are not available in the ExAC database. a The affected individual in AS114 with FMN2 mutation is a genocopy, and the genes/mutations in the other four affected individuals in this pedigree have yet to be established. b The homozygous nonsense mutation in TUSC3 is not in a shared HBD region and does not segregate with other affected members of family PK11. c The missense change in GPT2 is also predicted to affect a splice donor site (BDGP (www.fruitfly.org) splice predictor score: WT: 0.78; mut: 0.58).

2.4.4 Missense mutations

We identified homozygous missense variants as single candidate variants in 43 families, including 16 for which the genes identified have not previously been reported for NS-ARID (*AFF3*, *BDNF*, *CAPS*, *DMBT1*, *DUOX2*, *EXTL3*, *FBXO47*, *LAMC1*, *MAP3K7*, *SDK2*, *SPATA13*, *SUMF2*, *SYNRG*, *TET1*, *VPS35*, and *ZBTB11*). Although LoF mutations are frequently more convincing than missense, a number of the homozygous missense changes we report here are or particular interest, due to the known nature or function of the protein or the likely effect of the amino acid substitution on protein function.

For instance, here we report a homozygous missense change Met122Thr in the gene for the brain-derived neurotrophic factor, *BDNF*, which has been implicated in many studies of neuropsychiatric disorders (Autry & Monteggia, 2012), and is a known gene target of *MECP2*, the Rett syndrome gene(Amir et al., 1999; Meloni et al., 2000). This variant replaces a methionine- a large, non-polar residue with an S-methyl thioether side chain- that is fully conserved across the vertebrate lineage, with threonine- a small, polar residue with a hydroxyl side group. This variant is reported with a frequency of 0.0001647 in ExAC (in South Asians MAF=0.001211), with no homozygotes, and has not previously been reported in any publications. The BDNF protein is important for the survival, differentiation, and development of neurons within the central nervous system (CNS). Hence, the identification of Met122Thr in connection with a cognitive disability will likely be of great interest, and functional studies of the effects of Met122Thr on BDNF function are indicated.

We report a homozygous Lys2056Asn change in *TET1* for family PK70 that is not present in any variant or mutation databases. Methylcytosine dioxygenases TET proteins play a role in the DNA methylation process and gene activation, and TET1 is an important regulator of neuronal differentiation. TET1 is implicated in SCZ and BD by down-regulating *GAD1*, *RELN*, and *BDNF* genes through epigenetic mechanisms in the prefrontal cortex, and the cerebellum in autism (E Dong, Ruzicka, Grayson, & Guidotti, 2015; Zhubi et al., 2014). Significant upregulation of *TET1* mRNA in the parietal cortex of psychosis patients has also been reported (Erbo Dong, Gavin, Chen, & Davis, 2012). *Tet1* knockout shows impaired hippocampal neurogenesis and cognitive deficits in mouse (R.-R. Zhang et al., 2013), and abnormal brain

morphology in zebrafish (C. Li et al., 2015). The large basic, charged, hydrophobic lysine residue at this position is conserved in mammals (or glutamic acid in non-mammalian vertebrates) and is replaced here by a small acidic asparagine residue.

The homozygous His880Gln change identified in the zinc finger/BTB domain gene *ZBTB11* in family AN50 disrupts a canonical Zn^{2+}-binding residue in one of the zinc fingers and is likely to result in an alteration in the specificity of DNA-binding/gene regulation. *AFF3* is an autosomal homolog of the X-linked ID gene, *AFF2* (MRX-FRAXE; MIM 309547). As with *AFF2*, *AFF3* is also associated with a fragile site (FRA2A) for which expansion of a CGG trinucleotide repeat triggers hypermethylation and gene silencing in association with neurodevelopmental disorders(Metsu et al., 2014). The Gly1215Val variant identified here is situated within a five-amino acid C-terminal motif (Gln-Gly-Leu-His-Trp) that is highly conserved across the vertebrate lineage.

2.4.5 Families with multiple variants

For eleven additional families between two and four putative damaging variants were identified after filtration that fulfilled criteria and segregated in the families, but without functional evidence in support of pathogenicity, it is not yet possible to narrow these variants down to a single candidate (Table A- 3).

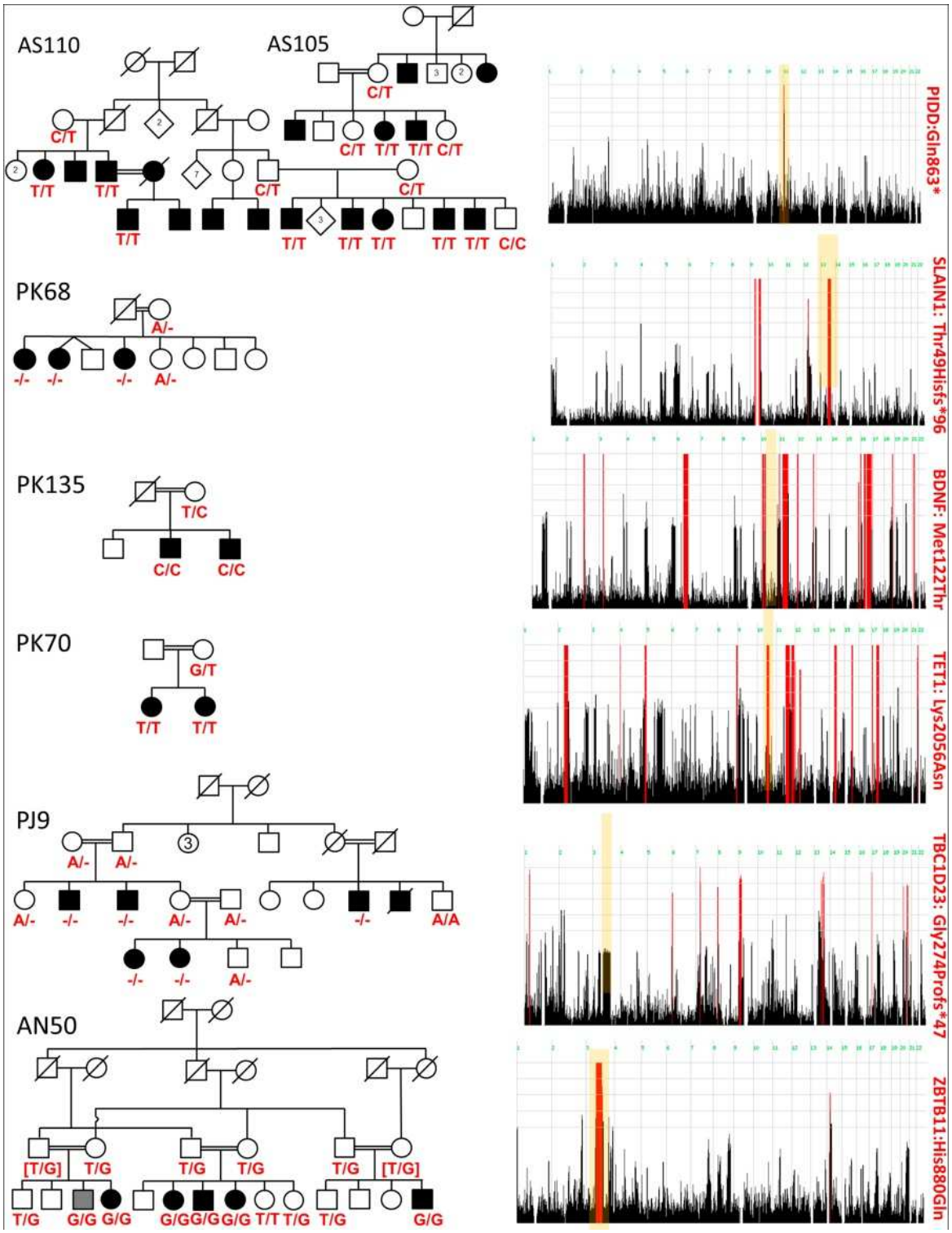


Figure 2.2: Pedigrees and HomozygosityMapper output for eight of the families

2.4.6 X-linked variants

A number of the families were compatible for both AR and X-linkage, and several variants or CNVs on the X chromosome have been identified, including two variants in *ATRX*, and a 6.7Mb interstitial duplication on Xp22.31-p22.2.

We report a missense mutation, Ser57Pro, in the X-chromosomal gene *MAGEA11*, segregating in family PK87. A Gln4Arg variant in MAGEA11 was recently reported among a cohort of X-linked ID families(H Hu et al., 2016). MAGEA11 shows protein interaction with TRMT1 — also reported here for NS-ARID— by yeast-two-hybrid(Rolland et al., 2014; Rual et al., 2005; Yu et al., 2011).

In another large multiplex and multi-branch family, the missense mutation Arg190His of the X-linked *MECP2* gene is present in a hemizygous form in a single male with mild ID, as well as in a heterozygous form in several females with mild ID or mild ID with psychosis or depression. Interestingly, for the female heterozygotes with ID, all appeared cognitively normal until age ~9 years, when cognitive regression started, and for the single male hemizygote, cognitive regression began much earlier, <5 years of age. The family also has several males with schizophrenia, with onset ~18 years but without cognitive regression amounting to intellectual disability, who are wild type for this variant (see Figure A- 2 2). The Arg190 residue is a critical DNA-binding amino acid within an AT-hook domain, and previously a *de novo* mutation at this residue, Arg190Cys, has been reported in a SCZ patient (McCarthy et al., 2014).

2.4.7 Genes with 'hits' in multiple families

Confidence in gene discovery relies on the identification of multiple affected families with mutations in the same gene, and/or replication in further studies. However, due to the anticipated high degree of genetic heterogeneity for NS-ARID, large sample sizes are typically required in order to find multiple families with mutations in the same gene. Comparison across different studies thus becomes vital. Of the ARID genes reported here for the first time, several are conspicuous by the presence of mutations in multiple families. In our study, the same nonsense mutation in *PIDD1* occurs in two apparently unrelated Pakistani families. We also provide confirmatory families/mutations for recently reported ARID genes such as *MBOAT7*, *IMPA1*, *TMEM135*, *PGAP2*, *GPT2* (2 families), *TDP2* (2 families), and *GMPPA*. Also, mutations in the

gene for the ID-associated brain malformation polymicrogyria, *GPR56*, are present in three families in this study, and could thus represent a relatively large proportion of ID families in these populations.

2.4.8 Families with mutations in previously identified genes for Metabolic or Hormonal Disorders

We report here three mutations in thyroid dyshormonogenesis (TDH) genes *TPO* (thyroid peroxidase), *TG* (thyroglobulin)(Mittal et al., 2016)), also *DUOX2* (thyroid oxidase 2). With adequate clinical resources available in most developed countries, many of these would likely have been diagnosed, and in some cases, for example, those with mutations in NAGS, *TPO*, *TG* and *DUOX2*, early treatment would have prevented ID development. Given the prevalence of mutations in these genes in a relatively modest number of families, it is likely that these disorders are relatively common causes of ID in populations where consanguinity is common but access to clinical diagnostics is poor. Mutations in *GNE* have previously been linked with sialuria (dominant; MIM 269921) — an extremely rare metabolic disorder— and Nonaka myopathy (recessive; MIM 605820), whereas here we report a homozygous missense mutation in a family with just ID and no myopathy. Although we were unable to get biochemical analysis on this family, we anticipate that this discovery represents a previously unreported recessive form of sialuria.

2.4.9 Three Novel Mutations in the Bilateral Fronto Parietal Polymicrogyria Gene *GPR56*

Family A

The parents were first cousins. There was no reported history of Intellectual Disability in the family. There are two male members affected in this family with very similar clinical presentation. In both boys the first cry at birth was delayed. They started to experience fits at age 2. Motor development was delayed and they walked around 6 years of age. They spoke only a few (5-6) single words. They started speaking these words at age 7. They never spoke full sentences. They were not toilet trained. They experienced frequent seizures which were tonic-clonic in nature. Medication used for seizures included Phenobarbitone, Sodium Valproate and

Clonazepam. They were totally dependent on others for eating, washing and other personal care. They had ataxia. The level of intellectual disability was severe to profound.

One of the brothers died in October 2015 after a febrile illness caused by pneumonia.

MRI was done on the older brother which showed classic "bilateral fronto-parietal polymicrogyria" with a gradient (anterior more than posterior). The microgyri were in a somewhat simplified gyral pattern (cobblestone).

Family B

This family has two branches and the two female patients are first cousins. The parents were first cousins. The clinical presentation in both the families is similar. In both the girls the first cry at birth was delayed. They started to experience fits within the first 2 years of life. Motor development was delayed and they walked around 5 years of age. They spoke only a few (5-6) single words. They started speaking these words at age 7. They never spoke full sentences. They were not toilet trained. They experienced frequent seizures which were tonic-clonic in nature. They were totally dependent on others for eating, washing and other personal care. They had ataxia. The level of intellectual disability was severe to profound.

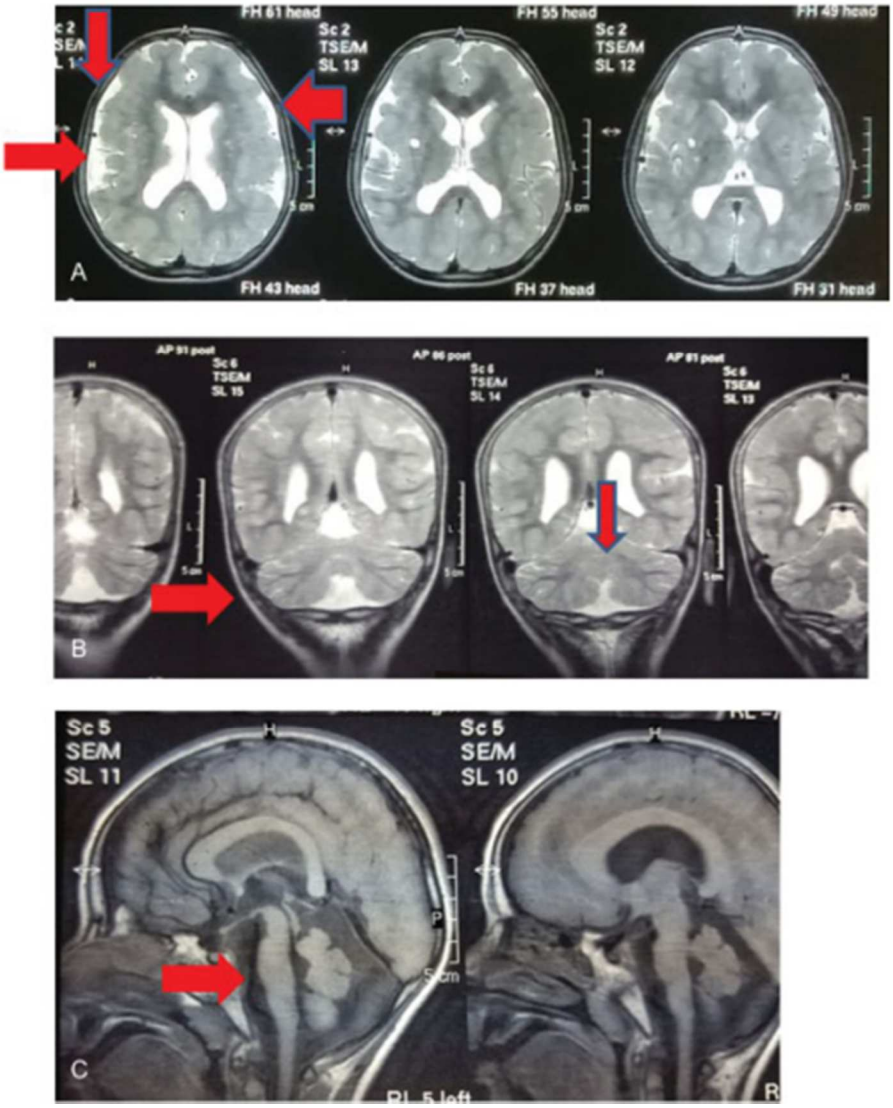


Figure 2.3: MRI from individual II:2 Family A. (A) MRI showing frontoparietal polymicrogyria transverse plane; (B) transverse plane showing cerebellum and cerebellar vermis; and (C) sagittal plane showing pons. MRI, magnetic resonance imaging.

Family C

The parents are cousins. Three children are affected; two females and one male. The disease was noted in the second year of life. Diagnoses included profound intellectual disability. By their teenage years, the patients still did not have any language at all. Their gait was unsteady, they had nystagmus and muscle spasticity. CT scan showed polymicrogyria.

In family A (Figure 2.4, Family A-i), a 5.5 Mb HBD region was identified flanked by SNPs rs3095600 and rs437947 (Chr16: 52,431,954-57,942,866), shared by affected individuals. Exome sequencing identified a novel homozygous missense mutation, (chr16:57693480T>C; NM_005682.5:c.1460T>C; p.Leu487Pro). Both affected individuals were homozygotes, and one normal female sibling and both parents heterozygous (Figure 2.4-Family A-ii). This variant is listed in the ClinVar database (Variation ID 158619) in association with BFPP in a single case. In Family B(Figure 2.4 Family B-i), a number of HBD regions were identified, including a 1.9Mb homozygous stretch on 16q21 was identified for individual III-6 from SNPs rs2306237 tors2133783 (Chr16: 56,973,335-58,881,897).Exome sequencing of III-6 identified a homozygous nonsense mutationChr16:57693446C>T; NM_005682.5:c.1426C>T; p.Arg476*, that was also confirmed in III-2(Figure 2.4-Family B-ii).This variant is listed in the ClinVar database (Variation ID 158618) in association with BFPP with multiple submitters. In family C (Figure 2.4 Family C-i), HBD mapping identified a 4.2 Mb region on 16q21. Because of the diagnosis of BFPP, and the presence of *GPR56* within this locus, Sanger sequencing was used and affected individuals were shown to be homozygous for a novel 13 bp insertion mutation, (Chr16: 57689345_57689346insCCATGGAGGTGCT; NM_005682.6:c.803_804insCCATGGAGGTGCT; p.Leu269Hisfs*21), in exon 7 of *GPR56*. Three unaffected siblings and both parents were heterozygous for this mutation (Figure 2.4-Family C-ii). This variant was not reported previously in the ClinVar database.

All three mutations segregated with disease status in the respective families, and all three mutations were absent from the Exome Aggregation Consortium (ExAC) database of 65,000 controls (including >8,000 South Asians).

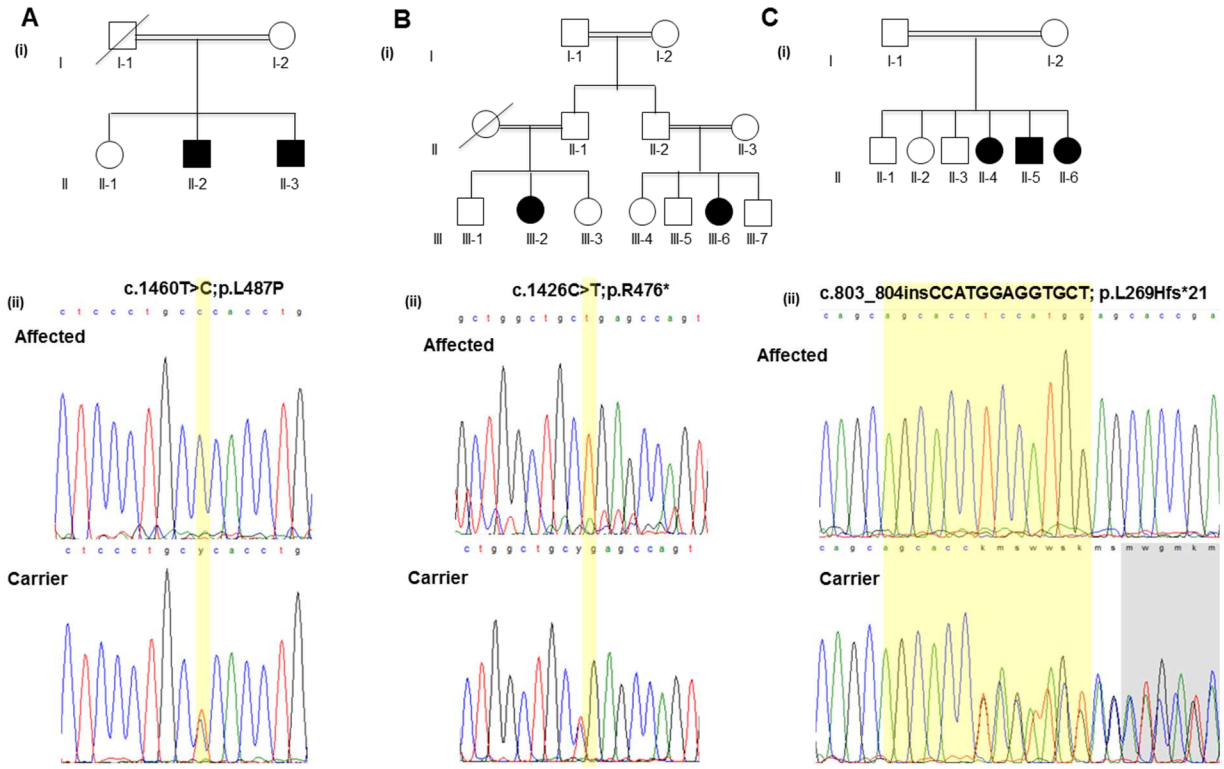


Figure 2.4: **(i) Pedigree drawings for Families A-C. Dark symbols indicate individuals affected by ID.** Electropherograms showing the homozygous mutations identified in GPR56 in affected individuals (ii) and in heterozygous carriers (iii). For Family C, electropherograms from the reverse sequence is shown.

Table 2.4: Clinical Features of Families A, B, and C

Family	Gender/Age	Head circumference	Cognitive delay	Motor delay	Dysconjugate gaze	Oromotor dyspraxia	Pyramidal signs	Cerebellar signs (ataxia)	Epilepsy	Age of onset	Type of seizure	Epileptic activity
A 1	M/14		Severe/profound	Yes	Yes	Yes	Yes	Yes	Yes	1 year	Tonic-clonic	Ongoing
A 2	M/9		Severe/profound	Yes	Yes	Yes	Yes	Yes	Yes	1 year	Tonic-clonic	Ongoing till death
			Severe/profound	Yes	Yes	Yes	Yes	Yes	Yes	1 year	Tonic-clonic	Ongoing
			Severe/profound	Yes	Yes	Yes	Yes	Yes	Yes	1 year	Tonic-clonic	Ongoing
C 1	F/23	52 CM	Severe/profound	Yes	Yes	Yes	Yes	Yes	Yes	2 year		Ongoing
C 2	M/21	57 CM	Severe/profound	Yes	Yes	Yes	Yes	Yes	Yes	2 year		Ongoing
C3	F/19	51.5 CM	Severe/profound	Yes	Yes	Yes	Yes	Yes	Yes	2 year		Ongoing

2.4.10 Structural Variant Analysis

In addition to HBD analysis, microarray data was used for CNV analysis, firstly to look for possible intra-familial genetic heterogeneity with large, probably pathogenic heterozygous loss/gain CNVs, and secondly to look for homozygous genic CNVs within mapped HBD regions. For the former, several candidates pathogenic CNVs were identified (Table A- 4), including an 8.4Mb deletion of 2q14.1-q14.3 (chr2:116583565-124954598) in one of the two affected individuals in family PK117. Deletions within this region have previously been reported for autism spectrum disorder(Pagnamenta et al., 2010) and a patient with mild holoprosencephaly spectrum phenotype(Kevelam et al., 2012). An overlapping, but slightly proximal deletion (~chr2: 114188161-119321989) has been reported to be compatible with normal phenotype(Barber, Maloney, Bewes, & Wakeling, 2006). In one family, PK28, all three affected males in one branch were shown to have a large (6.7Mb) interstitial duplication spanning cytobands Xp22.31-p22.2.

For homozygous CNVs in HBD regions, we have identified homozygous deletions in known ARID genes in several families (Table 2.3, and Table A- 3), including a 170 kb homozygous deletion spanning 9 out of 10 exons of NS-ARID gene, *TUSC3*, in family ANMR21(Muzammil A Khan et al., 2011), also a 51 kb deletion spanning exons 37 to 40 of the Cohen syndrome gene, *VPS13B*, in family ANMR51(Muhammad A Rafiq et al., 2015). Homozygous disrupting loss CNVs include a 50 kb deletion spanning exons 2 and 3 of *RAB8B* in PK95-7, however, this was not in a shared HBD region, and was thus considered as a potential case of intra-familial genetic heterogeneity.

2.5 Database Searches for Putative Genes

2.5.1 Sequencing Databases for Neuropsychiatric Disorders

It has been noted that genes relevant for ID are frequently also identified in individuals with autism spectrum disorders (ASD), and epilepsy (which both frequently also present with ID). In addition, there is growing support for overlap of ID genes with other neuropsychiatric disorders (Hall & Owen, 2015). For example, genes such as *NRXN1* and *ANK3* have been linked with SCZ, BD and ASD by CNV and/or genome-wide association studies are also known ARID

genes (*NRXN1*: PTHLS2, MIM 614325; *ANK3*: MRT37, MIM 615494). For this reason, we attempted to evaluate the genes identified herein variant databases or among gene lists for such disorders. We have screened various epilepsy or neuropsychiatric disease-specific databases or published datasets for the presence of variants in these genes. In the SCZ/control exome sequence database Genebook(Fromer et al., 2014; Purcell et al., 2014), none of the variants we have identified in Table 2.2 were present. However, for several of the genes, there is a reported increased burden of rare variants in SCZ cases versus controls ($p<0.05$ for *VPS35*, *SYNRG*, *DMBT1*, *ALPI*, and *NEU4*; $p<0.002$ for *SLC13A5*).

In epilepsy databases, we identified *ATRX*, *MECP2*, *SLC13A5*, and *ST3GAL3*, all of which have been reported in cases of epilepsy as well as ID. These four genes are either inherited as autosomal recessive or X-linked recessive. *ATRX* and *MECP2* are involved in chromatin binding and gene regulation(Pruitt, Tatusova, & Maglott, 2007). Mutations in *ATRX* and *MECP2* can have diverse effects on DNA methylation patterns and brain development.

2.5.2 Protein Interaction Analysis

We used BioGrid (http://thebiogrid.org) to identify protein interactors for each of the 67 different ARID genes identified among our population (Table A- 3) as well as *GRIN2B* (Table 2.3). Interacting proteins are listed, along with gene ontology processes, functions, and cellular compartments, in (Table A- 5).

2.5.3 Pathway analyses

In agreement with past findings of genetic heterogeneity, we observed limited overlap with the gene ontology gene sets. Of the 14,312 sets tested, 16 survive multiple test correction (q-value < 0.05) and a few top ranked terms are of interest (Table A- 5). Protein glycosylation which is known to be associated with intellectual disability was ranked 6th with 9 overlapping genes (corrected $p < 0.05$). As mentioned above, *TPO*, *DUOX2,* and *TG* are known to be involved in thyroid hormone generation (rank: 14, corrected $p < 0.05$). The smoothened signaling pathway is ranked 32nd with 4 overlapping genes (*NDST1*, *CC2D2A*, *TCTN2,* and *BBS7*).

2.6 Spatiotemporal Expressions Analyses

2.6.1 Anatomical expression analyses

With the exception of the spinal cord, all neural tissues were enriched in the expression of ID genes (corrected $p < 0.0001$,Table A- 7). The frontal cortex was the most enriched with 1.4 times the expected number of expressed genes. While the brain does express a majority of all genes, it appears that the ID genes show specificity. Of the 17 fetal tissues tested, the female gonad and testis have the highest number of expressed genes but do not show enriched expression of the ID list (corrected $p > 0.12$).

2.6.2 Developmental brain expression analyses

Across the developmental transcriptome, the ID genes are expressed at higher levels in the normal prenatal brain (Figure 2.5). The amygdaloid complex shows the most consistent enrichment (5 specimens with significantly higher expression of ID genes). Fetal brain samples from 21 and 24 post-conception weeks show the highest number of regions with significant expression (> 4 or more). In contrast, enriched expression in the postnatal brain samples is not observed. This prenatal pattern found in exon microarray expression profiles is also observed in RNA sequencing measurements in a largely overlapping set of samples from the same resource (BrainSpan, Figure A- 3). The RNA sequencing measurements show donor specific patterns with many enriched regions. These global patterns are not consistent across donors of the same age, suggesting the RNA sequencing data may have normalization artifacts that are not seen in the microarray measurements. Grouping the ID gene list into genes associated with glycosylation or hormones and metabolism shows above average prenatal expression for all groupings (Figure A- 3). Genes with metabolic or hormonal associated function have the most stable trajectory. We also observe higher expression for these groups in the prenatal brain in the BrainCloud (Colantuoni et al., 2011) resource that assayed the prefrontal cortex across human development (Figure A- 4).

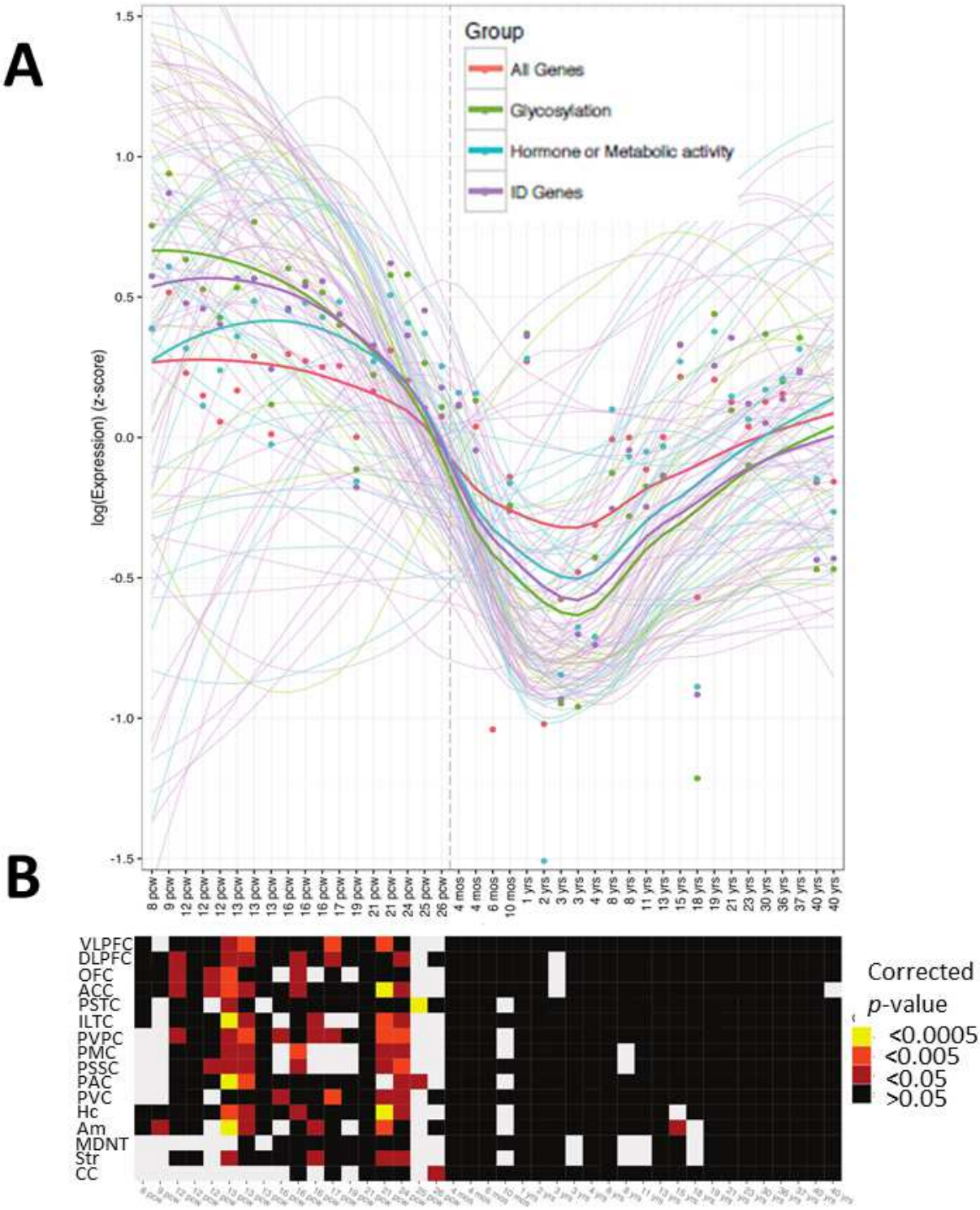

A
B
Group
All Genes
Glycosylation
Hormone or Metabolic activity
ID Genes
log(Expression) (z-score)
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
8 pcw
9 pcw
12 pcw
12 pcw
12 pcw
13 pcw
13 pcw
13 pcw
16 pcw
16 pcw
16 pcw
17 pcw
19 pcw
21 pcw
21 pcw
24 pcw
25 pcw
26 pcw
4 mos
4 mos
6 mos
10 mos
1 yrs
2 yrs
3 yrs
3 yrs
4 yrs
8 yrs
8 yrs
11 yrs
13 yrs
15 yrs
18 yrs
19 yrs
21 yrs
23 yrs
30 yrs
36 yrs
37 yrs
40 yrs
40 yrs
VLPFC
DLPFC
OFC
ACC
PSTC
ILTC
PVPC
PMC
PSSC
PAC
PVC
Hc
Am
MDNT
Str
CC
Corrected
p-value
<0.0005
<0.005
<0.05
>0.05

Figure 2.5: Spatiotemporal expression of ID genes in human brain development: a) Scaled expression trajectories of the ID Genes, averaged across brain regions. Solid lines are average expression for a group of genes. The lighter and thinner lines show trajectories of individual genes. Average expression for each gene group is depicted with points (green: glycosylation associated (14 genes), blue: hormone or metabolic (20), purple: remainder of the ID Genes (66) and red: remaining genes assayed). Local regression was used to smooth the z-scored expression values for individual genes and gene group averages (LOESS). A vertical dashed line marks birth. b) Heatmap showing brain samples that are enriched for specific expression of ID genes. For each brain region and brain combination, z-scored expression values for the ID genes was compared against all other genes (Wilcoxon one sided test, FDR corrected p-values). Significance levels are indicated by black (non-significant result), red ($p < 0.05$), orange ($p < 0.005$) and yellow ($p < 0.0005$). Missing values are shown in grey. Brain regions include: ventrolateral prefrontal cortex (VFC), dorsolateral prefrontal cortex (DFC), orbital frontal cortex (OFC), anterior (rostral) cingulate (medial prefrontal) cortex (MFC), posterior (caudal) superior temporal cortex (area 22c) (STC), inferolateral temporal cortex (area TEv, area 20) (ITC), posteroventral (inferior) parietal cortex (IPC), primary motor cortex (area M1, area 4) (M1C), primary somatosensory cortex (area S1, areas 3,1,2) (S1C), primary auditory cortex (core) (A1C), primary visual cortex (striate cortex, area V1/17) (V1C), hippocampus (hippocampal formation) (HIP), amygdaloid complex (AMY), mediodorsal nucleus of thalamus (MD), striatum (STR) and the cerebellar cortex (CBC).

2.7 Discussion

This study describes a cohort of 192 multiplex ID families from Pakistan and Iran. With combined microarray genotyping, homozygosity-by-descent (HBD) mapping, copy number variation (CNV) analysis, and whole exome sequencing (WES) we identified definite or candidate mutations (or CNVs) in 51% of families, in 72 different genes, including 26 not previously reported for ARID.

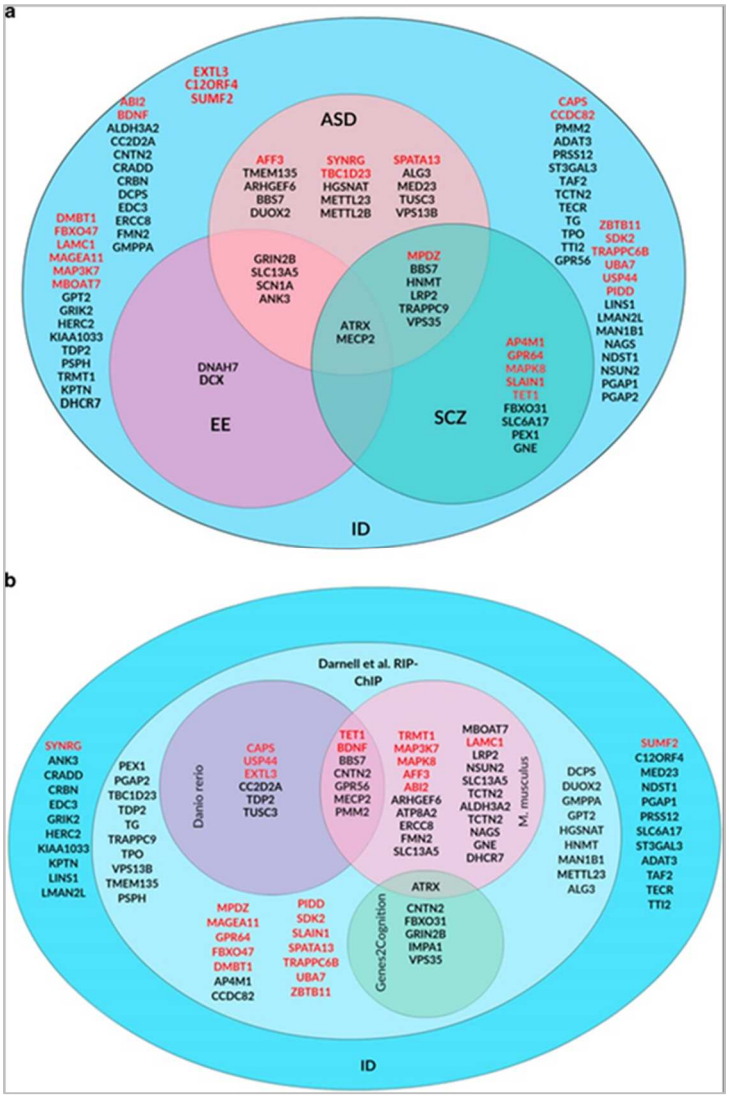


Figure 2.6: Cross-disorder overlap: The Venn diagrams shown indicate genes for which **a.** variants have been reported in other neuropsychiatric or neurodevelopmental disorders, either homozygous, compound heterozygous, or *de novo*, through searches of published gene-list; **b.** functional or animal models with relevant phenotypes have been reported. Genes newly reported here are in red text.

We note that 50% of the variants we report as single probable mutations are LoF changes. This compares well with the earlier study reported by Najamabadi et al, 2011(Najmabadi et al., 2011), in which 50 new genes for NS-ARID were reported, of which 40% had LoF mutations. In general, LoF mutations provide a higher degree of confidence of disease association over missense mutations.

There are a number of likely reasons that genes/mutations were not found for some families: 1. Intrafamilial etiologic heterogeneity- 2. poor WES depth of coverage at the etiologic gene/mutation; 3. causative mutations may be intronic or intergenic, and thus not picked up by WES. Whole genome sequencing may address this issue; 4. Common variants have been reported in association with many complex diseases including traits like intelligence or cognitive ability(Clarke et al., 2015; Davies et al., 2016). It is plausible that a proportion of ID cases with familial aggregation are caused by variants common in the general population but with low penetrance. These variants working together could potentially cause ID, however, this hypothesis has not been explored in ID. 5. non-genetic factors may be prevalent in some families, e.g. prenatal or perinatal insult(J. Huang, Zhu, Qu, & Mu, 2016).

Given that there may be an overlap in genetic etiology of neurodevelopmental and neuropsychiatric disorders, we cross-referenced our ID gene list with those from studies of other neuropsychiatric/neurodevelopmental disorders. We see multiple instances of genes identified for ARID that are implicated across disorders (see Figure 2.5). For instance, of the newly identified genes , *SLAIN1* is listed as one of the top-ranked genes for the burden of variants among a large cohort with SCZ (N=1,392)(Singh et al., 2016). Data from the Autism Sequencing Consortium(De Rubeis et al., 2014) reports *de novo* LoF mutations in *SPATA13* and *TBC1D23*, as well as *de novo* missense mutations in *ABI2*, *TET1,* and *SYNRG*. Li et al (2016)(J. Li et al., 2016) report variants in both *MPDZ* and *SPATA13* in ASD cohorts, and a *de novo* missense variant in *MAPK8* in SCZ(Fromer et al., 2014). Also, whereas heterozygous (typically *de novo*) mutations in *SCN1A* have been linked with ASD, ID, and epileptic encephalopathy, the missense *SCN1A* variant we report here is homozygous. A number of ARID genes (e.g. *NRXN1*, *CNTNTAP2*, *ANK3*) have already been implicated in neuropsychiatric disorders by GWAS or CNV studies (as heterozygous). We speculate that, in addition to pleiotropy with other neuropsychiatric/neurodevelopmental disorders, the mode of inheritance may also be variable, (e.g. *SPATA13*: homozygous missense variant in ID (this study) and *de novo* LoF variant in ASD(J. Li et al., 2016); *HNMT*: homozygous missense variant in ID(Heidari et al., 2015), and heterozygous splice mutation in SCZ (Genebook)). ID may be conceptualized as a much severe form of neurodevelopmental disorder, compared to ASD and SCZ.

Transcriptional analysis of our gene set combined with known ARID genes shows greater levels of transcription in the prenatal brain than the postnatal or adult brain (Figure 2.5) and in

particular higher levels in frontal cortex, hippocampus, and amygdala (Figure 2.5). ID is primarily a disorder of brain development, and thus it is reassuring to observe the ID genes showing relevant patterns of spatiotemporal expression. Pathway analysis of the gene set showed a few significant pathways; thyroid metabolism was prominent, as was protein glycosylation and the Smoothened signaling pathway. Involvement of a small number of pathways in a cohort of this size is probably an indication of the huge genetic heterogeneity in the number of genes and pathways involved in ID.

Our findings and those of other groups studying the genetics of NS-ARID will be invaluable for the development of targeted sequencing gene panels for diagnostic screening, and for the development of clinical whole exome and whole genome sequencing. In this regard both the replication of previous findings and new discoveries in this study are important. With replication, the confidence in the validity of the findings is enhanced and new discoveries open up the opportunities for further research in replication of the genetic findings but also experiments to study the functions of the newly associated genes. Together, these studies are also mapping genes across the human genome for which LoF mutations are viable, and their roles in human development are thus amenable for further study and comparison with similar mutations in model organisms. Furthermore, a more complete picture is being assembled of the molecular components and mechanisms that are important for the development of a fully functioning central nervous system, as well as documenting points in the mechanisms that are most vulnerable to genetic mutation. Through comparison with similar studies, we also note that there are much more such discoveries to be made to complete the picture. The ultimate challenge, to devise targeted therapeutic strategies for ID patients, is thus a step closer.

The discovery of disease-causing mutations in consanguineous families immediately creates opportunities for carrier screening among relatives and prevention of ID. This is a direct benefit to the families' communities and to public health. In addition, finding genetic causes for ID makes it possible to subgroup ID individuals by gene or pathway. This can lead to the development of cohorts, which can then be studied prospectively for the natural course of the disease, health complications, and also can be targeted for therapeutic strategies. This is a step towards personalized medicine in this important clinical population.

2.8 Addendum to Harripaul et al, 2018

Papers that have since provided supporting evidence/corroboration to the novel autosomal recessive candidate genes identified in Harripaul et al, 2018. *PIDD1* (Hu et al, 2019; Sheikh et al, Translational Psychiatry, *under review*), *DUOX2* (Bruellman et al, 2019; Peters et al, 2019), *NARS* (Manole et al, submitted to Nature Communications), *GPT2* (Ouyang et al, 2019), *ZBTB11* (Fattahi et al, 2018), *MBOAT7* (Johansen et al, 2016; Khan et al, 2019; Jacher et al, 2019), *TRAPPC6B* (Marin-Valencia et al, 2018), *ITFG2* (Gifty Bhat, *in preparation*), *TBC1D23* (Ivanova et al, 2017; Marin-Valencia et al, 2017); *SYNRG* (Saima Riazzuddin et al, *in preparation*). *CAPS2* (but not *CAPS*) Hu et al, 2019).

3 Whole exome sequence analysis in autism spectrum disorder trios from consanguineous populations

This chapter is unpublished but I contributed in the experimental design, addressing workflow and data quality issues, data generation, data processing, data analysis, data interpretation, manuscript drafting, editing and figure generation.

3.1 Abstract

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that affects 1 in 54 children in the US. This disorder has a staggering cost to healthcare, estimated to be $1.4 million USD per child, and rising to $2.4 million USD for cases with both ASD and intellectual disability (ID) over the course of the child's lifetime. Most recent DNA sequencing studies have focused on *de novo* variants with smaller family sizes, with little attention to autosomal recessive forms of ASD. This study uses whole exome sequencing of 115 trios from consanguineous geographical regions to enrich for recessive variants. In total, 148 variants in genes were identified for the 115 trios including 81 variants being Loss of Function (LoF). 68 (45.9 %) of 148 variants identified were recessive including 81 (54.7 %) were *de novo*. This cohort identified 38 known genes (31.3 %), with 28 of the autosomal recessive variants, 10 *de novo* heterozygous and six variants reside in X-linked genes. Cumulatively, 148 gene variants were identified in 115 trio families, with 68 recessive variants (45.9 %) and 81 *de novo* variants (54.7 %). Seven of the genes were previously been associated strictly with ID, and six have been associated with inborn errors of metabolism. Gene expression and cell-typing of the implicated genes indicated that the cortical plate, subplate and ventricular and excitatory and inhibitory neurons in critical learning and developmental regions such as the hippocampus and cerebral cortex were significantly enriched. The genes discovered in this cohort also showed significant co-expression with other known ASD and ID genes, with critical developmental expression points between 16 to 24 weeks post conception in the hippocampus, cerebral cortex and amygdaloid complex. As genes were added from other reported ASD and ID studies, other brain regions also became significant. Recessive variants represent an underappreciated class of variants that may provide additional diagnoses in ASD trio cohort studies..

3.2 Introduction

Autism Spectrum Disorder (ASD), is a developmental disorder which is characterized by repetitive behavior and impaired social communication and interaction with others. In Canada, 1 in 66 children are diagnosed with ASD each year with 1 in 42 being male and 1 in 165 being female (Ofner et al., 2018). This figure has risen to 1 in 54 children in the US(Maenner, Shaw, & Baio, 2020). ASD has a huge socioeconomic and emotional cost to families and communities involved with a loss of services due to healthcare cuts and aging out of school systems. 50 000 teens each year lose school-based services as they become adults(Baio, 2014) . This is compounded with the comorbidity of other psychiatric and developmental disorders that occur with ASD. Many disorders such as attention deficit and hyperactivity disorder (ADHD), seizures, anxiety, depression and gastrointestinal issues often accompany ASD. Intellectual disability (ID) occurs in 25-70 % of children with ASD and significantly increases the socioeconomic and care taking burden for both government, community and family (Baio, 2014; Dawson, Mottron, & Gernsbacher, 2008). It is estimated that it requires $3-5 million CAD to raise a child with ASD over their lifetime, with this figure increasing/changing due to other complexities contributed by comorbid features(Buescher, Cidav, Knapp, & Mandell, 2014).

The aetiology of ASD is not well understood, however, there is no single specific cause for ASD but it has many genetic and environmental factors. The genetic contribution to ASD have been established as 50-98% heritability through twin and adoption studies(Tick, Bolton, Happé, Rutter, & Rijsdijk, 2016). These studies have highlighted the extreme genetic heterogeneity in ASD, with many gene found associated to ASD but each gene only explaining a small fraction of cases of the disorder. Recent studies have provided major breakthroughs through identifying rare variants and Copy Number Variant (CNV) associated with ASD (Glessner et al., 2009; Marshall et al., 2008). Common variants have also been found to contribute to ASD with genome wide association study finding significant common variants with 18 381 individuals with ASD and 27 969 controls(Grove et al., 2019). These results show there is an interplay between rare and common variants(Weiner et al., 2017). It is estimated that common variants account for 20 % of ASD liability while *de novo* and copy number variations account for < 5%(Grove et al., 2019; Weiner et al., 2017). This also results in rare variants that individually explain small amounts of the hereditary for ASD, but together combine to explain the genetic and clinical heterogeneity seen in patients.

Many studies have that have focused on identifying rare variants, have been in Western countries and have typically identified *de novo* dominant and X-linked variants. This has identified some very penetrant genes such as *SHANK3* (Durand et al., 2007), *SHANK2* (Schmeisser et al., 2012), *DEAF1* (Rajab et al., 2015) and *PTCHD1* (Noor et al., 2010). but there still remains a large proportion of cases and heritability that requires an explanation. Recessive models have been underutilized in genetic studies but may represent an elegant model to explain a significant proportion of the heritability of ASD cases. Recessive variants remains hidden in heterozygous form, but emerges as either allozygous or autozygous forms. Allozygous is generally compound het, but can also be homozygous, where two damaging variants, on two different alleles, need to be present to fully manifest the phenotype. Variable penetrance and clinical manifestations may be explained by recessive variants due to the types of mutations present and no (or perhaps more subtle) phenotypes present in carriers; allowing these alleles to remain in the population. Research efforts in North American countries have yielded few recessive variants for ASD, despite the advent of Next Generation Sequencing (NGS) technologies, because North American families are typically smaller and analysis methods are not optimal for detecting recessive variants or compound heterozygotes. A number of recent papers have investigated recessive variants in ASD; however, these papers remain in the minority of ASD research (Chahrour et al., 2012; Doan et al., 2019; K. K. Lim & Chong, 2017; Morrow et al., 2008).

To date, the most effective method of identifying recessive variants is through the use of large, consanguineous families where recessive variants are passed down from generation to generation and are enriched in the population. The genes responsible for disease are usually inherited through shared stretches of homozygosity in the genome from parents to children; also known as homozygosity mapping. Families from the Middle East and South Asian, represent an attractive population for this type of study because consanguineous marriages are culturally and social embedded in the population (60% of marriages in Pakistan ((Hussein & Bittles, 1998)) and 40 % of marriages in Iran are consanguineous (Najmabadi et al., 2007). These cultural norms are across socioeconomic status, and, due to rapid population expansion and migration, various sub structures and founder populations exist.

In this study, we have used consanguineous trios predominantly from Iran and Pakistan with the aim of determining the contribution that autosomal recessive variants play in ASD and identify novel genes that cause ASD. These results will shed light on the role of recessive ASD

mutations in consanguineous populations, which represents over a billion people(Bittles & Black, 2010), and identify novel genes that can be translated to outbred populations, to seek new diagnostic approaches, clinical management and potential therapies from families burdened with this debilitating disorder.

3.3 Materials and Methods

3.3.1 Trio Family Recruitment

Institutional Research Ethics Board approval was received for this study through the Centre for Addiction and Mental Health (CAMH) and other institutional recruiting sites. All parents from Iranian cohorts filled out the Social Communication Questionnaire (SCQ) and children who met the specified cut-off were then assessed using the Autism Diagnostic Interview-Revised (ADI-R) and some with the Autism Diagnostic Observation Schedule (ADOS). The diagnosis of ASD was confirmed in all subjects using the ADI-R and ADOS. Twenty-seven trios were ascertained by Dr. Sasanfar by the Children's Health and Evaluation project sponsored by the Special Education Organization (SEO). Dr. Bita Bozorgmehr clinically assessed 10 trios (30 samples) and Dr. Abolfazel Heidari assessed three trios (9 samples). Fifty-four Pakistani trios were recruited from the University of Health Sciences and Children Hospital, Lahore, Pakistan. Pakistani ASD trios were assessed using the Childhood Autism Rating Scale (CARS). The clinical assessment has been performed by Drs. Saqib Mahmood, Shazia Maqbool and Ansa Rabia. A further eight Pakistani trios were recruited through COMSATS University, Islamabad, by Drs. Raheel Qamar and Zehra Ahga. The 13 Saudi Arabian trios were provided by Dr. Laila Al Ayadhi through the Autism Treatment Centre, King Saud University, Riyadh, Saudi Arabia and who had a confirmed diagnosis of ASD using the ADI-R, ADOS and 3DI scales. Overall, we have sequenced 115 trios: 61 from Pakistan, 42 from Iran and 12 from Saudi Arabia (Table 3.1). In total, we have sequenced 345 samples from different institutions with collaborator contributions listed in Table 3.1. Saudi and Iranian trios were tested for Fragile X and were negative for FRAXA.

Table 3.1: Demographic and Biological Gender Composition of Pakistani and Iranian Trio Families in this Study

Cohort ID (Recruited By)	Institute	Nationality	Number of Trios Sequenced (Number of	Male: Female

			Samples WES)	
IAU (Dr. Sansafar)	Special Education Organization	Iran	27(81)	23:4
IABB (Dr. Bita Bozorgmehr)	Shahid Beheshti Medical University	Iran	11(33)	9:2
IAH (Abolfazl Heidari)		Iran	3(9)	2:1
SMPA (Saqib Mahmood, Ansa Rabia, Shazia Maqbool)	University of Health Sciences and Children Hospital, Lahore	Pakistan	54 (162)	42:12
Autism (Raheel Qamar and Zehra Agha)	University of Health Sciences and Children Hospital, Lahore	Pakistan	8(24)	4:4
SADD (Laila Al Ayadhi)	King Saud University	Saudi Arabia	13(39)	12:1
Total			115(345)	91:24

3.3.2 Whole Exome Sequencing

Whole Exome Sequencing (WES) was performed using the Thruplex DNA-Seq (Rubicon Genomics) Library Preparation Kit with the Agilent SureSelect V5 Exome Capture kit. Samples were sheared using the Covaris ME220 Focused-Sonicator to ensure that the DNA is sheared to 200 bp. Fragment lengths of all samples where then analyzed using the Agilent 2100 Bioanalyzer System for fragment length distribution and quantification. Library preparation and adapter ligation was performed using the Thruplex DNA-seq kit. All trios, representing the proband, mother and father, were sequenced on the Illumina HiSeq 2500 or the Illumina NovaSeq sequencing system.

3.3.3 Sequence Alignment and Variant Calling

After sequencing, FastQC was used to assess the quality of sequences and determine the adapter contamination and read quality. Unligated adapters and low quality sequences were removed using FastX Toolkit (Updated Aug 4th 2017: https://github.com/agordon/fastx_toolkit)(Gordon & Hannon, 2010) and Trimmomatic (v 0.33: Updated Mar. 10th, 2015: https://github.com/timflutre/trimmomatic)(Bolger, Lohse, & Usadel, 2014). Cleaned fastq files were then aligned to the hg19 reference genome using BWA 0.7.17 ((H. Li, 2013)) and variants were called using the GATK Best Practices for variant calling and quality control (DePristo et al., 2011). Briefly, after alignment, duplicate reads were removed using Samtools version 1.4(H. Li et al., 2009) and the GATK v 4.1.7 (Van der Auwera et al., 2013) was used for Base Recalibration, Indel Realignment and then variant calling using the Global Haplotyper with the ScoreCNN machine learning method for variant filtering(Van der Auwera et al., 2013). This created raw Variant Call Format (VCF) files. A general overview of the process can be viewed in (Figure 3.1). The variants called per sample were then checked for their transition-transversion ratio which was approximately 2.5. After alignment, 92.3% of the reads were above Q30 with a minimum coverage of 20 X across the exome and mean coverage of 54 X for all samples.

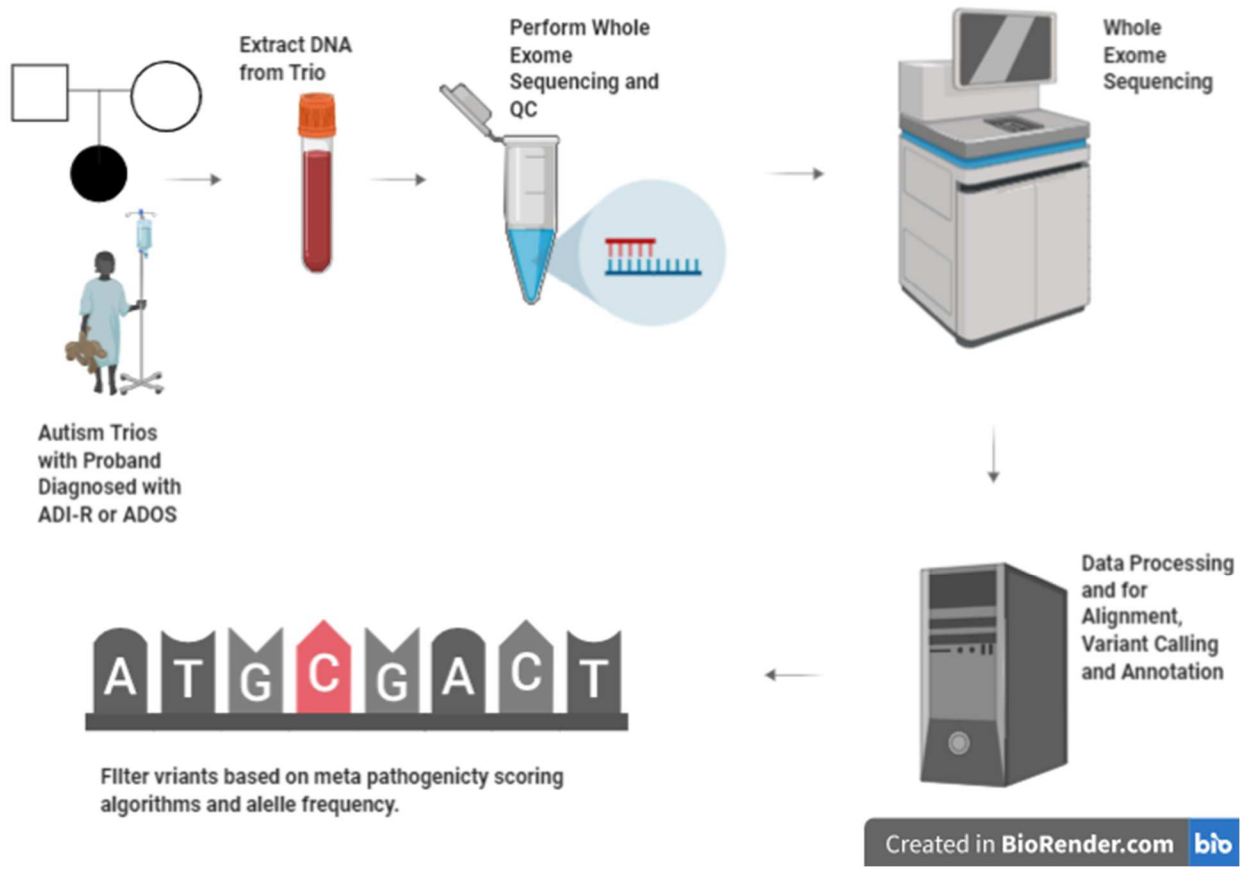


Figure 3.1: Workflow of experimental design where ASD probands and parents are collected, Sequenced and run through a variant prioritization schema.

3.3.4 Annotation and Variant Prioritization

VCF files were annotated using ANNOVAR (H. Yang & Wang, 2015)and integrated allele frequencies, from gnomAD (Karczewski & Francioli, 2017), The Greater Middle Eastern Variome Server (Scott et al., 2016) and ExAC (Lek et al., 2016) as well as functional pathogenicity scores from Polyphen2(IA Adzhubei et al., 2010), SIFT (Ng & Henikoff, 2003) and MutationTaster (Schwarz, Rödelsperger, Schuelke, & Seelow, 2010), MutationAssessor (Reva, Antipin, & Sander, 2010), FATHMM (Shihab et al., 2013), CADD (Rentzsch, Witten, Cooper, Shendure, & Kircher, 2019). Variants were assessed, firstly, if they passed the quality metrics as set by the Convolutional Neural Network scoring algorithm by GATK (Van der Auwera et al., 2013). Briefly, variants were compared against known variants from dbSNP in and a convolutional neural network was trained to filter out poor variant calls based on several metrics (Van der Auwera et al., 2013) and allele frequency in gnomAD of equal to or lower than

1×10^{-3} for autosomal recessive, or 1×10^{-5} for autosomal dominant/*de novo*, with a preference for variants that had no homozygotes in the gnomAD non-Neuro Cohort. Variants were then split into either homozygote or heterozygote categories and then prioritized by functional impact of mutation type. For example, loss of function (LoF) variants would carry more weight than missense variants leading to assessing stopgain, frameshift, splicing and missense variants in that order (Harripaul et al., 2018). The variants were then assessed for their functional pathogenicity score, where variants that met at least two functional prediction scores between, Polyphen2 (IA Adzhubei et al., 2010), and MutationTaster (Schwarz et al., 2010) and CADD (Rentzsch et al., 2019) were further investigated. Special attention was given to splicing with additional functional annotation scores being used for splicing events with the addition of FATHMM (Ivan Adzhubei, Jordan, & Sunyaev, 2013a) and dbscSNV (Jian, Boerwinkle, & Liu, 2014). Due to the large number of missense variants for autosomal dominant/*de novo* inheritance, stricter criteria were used where a variant would be considered if it was predicted as pathogenic by all software used including Polyphen2 (Ivan Adzhubei et al., 2013a; Ivan Adzhubei, Jordan, & Sunyaev, 2013b), MutationTaster (Schwarz et al., 2010), MutationAssessor (Reva et al., 2010), FATHMM (Shihab et al., 2013), CADD (Rentzsch et al., 2019) and M-CAP (Jagadeesh et al., 2016). The overall prioritization method is outlined in Figure 3.2Variants were also analyzed to determine if they are in genes associated with any previous studies reporting association with ASD or ID. Pathways that have higher priority were also considered such as neuron growth and guidance using gene expression and GOrilla ((Eden et al., 2009)). The LoF mutations were also given higher priority as it is assumed that LoF variants will disrupt protein function more severely and, depending on the position within the transcript, cause non-sense mediated mRNA decay (NMD). Finally, selected variants were Sanger-sequenced and segregation checked with both parents for validation.

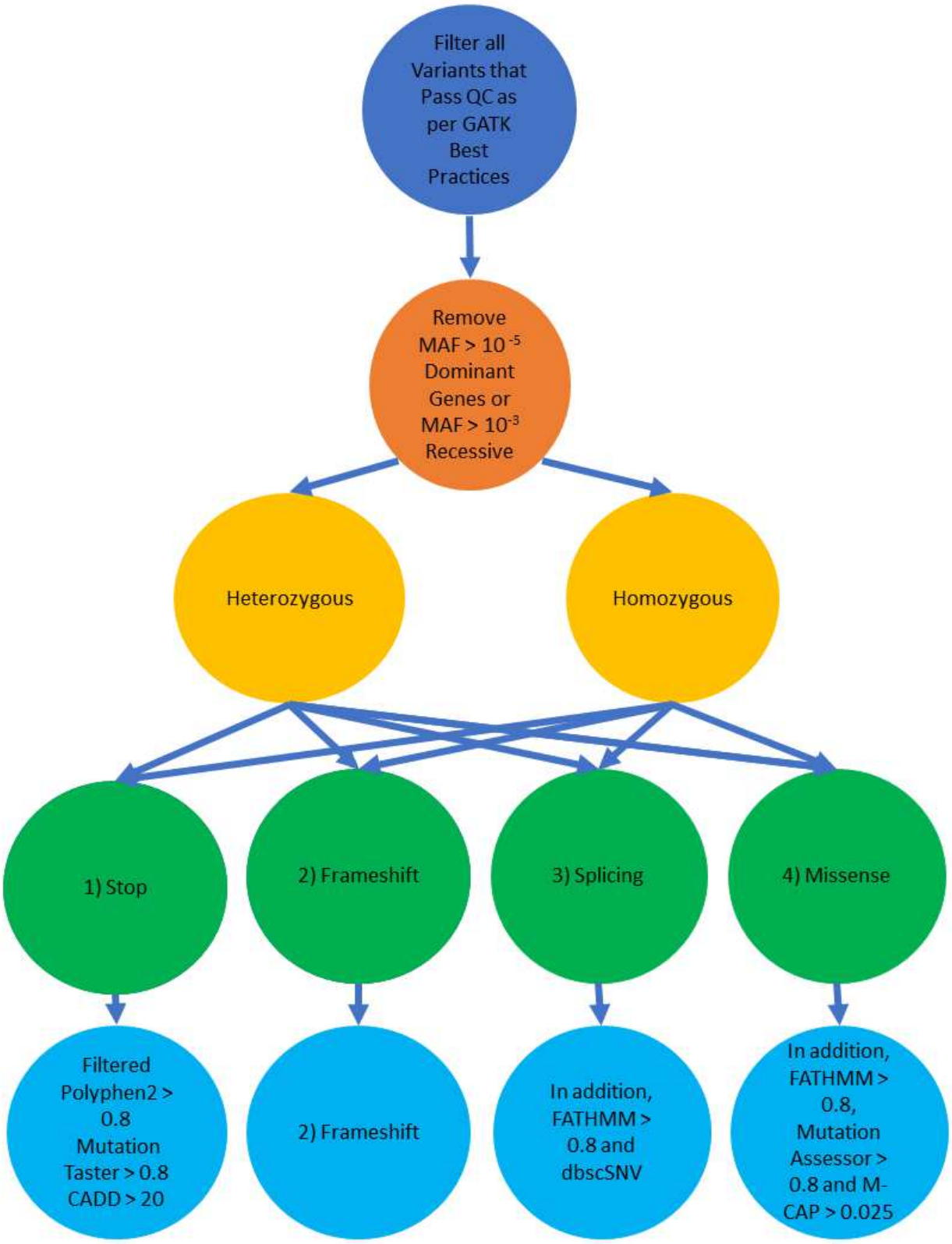


Figure 3.2: Variant prioritization methods where variants that pass GATK best practices would be filtered for MAF less than 10^{-5} and then categorized into recessive or homozygous

variants. The variants are then prioritized based on how damaging it would be to the protein and then filtered based on scoring algorithms to predict pathogenicity.

3.3.5 Construction of Gene List for RNA Expression and Gene Enrichment

A list of genes was constructed to represent the genes and molecular pathways that have been previously associated with ID and ASD. For the subsequent analysis, genes involved in this cohort, genes associated with ASD and genes associated with ID were treated separately with a final list of all combined genes. The gene lists were created by extracting all genes from OMIM (Hamosh et al., 2005), ClinVar (Landrum et al., 2014), the Deciphering Developmental Disorders Study (Study et al., 2017; C. F. Wright et al., 2015) and SFARI (Banerjee-Basu & Packer, 2010) databases in combination with several recent papers (Anazi et al., 2017; Guo et al., 2018; Harripaul et al., 2018; H Hu et al., 2016; Hao Hu et al., 2019; Jamra et al., 2011; Krumm et al., 2015; Najmabadi et al., 2011; O'Roak et al., 2012; Reuter et al., 2017; Riazuddin et al., 2017). These lists were separated and unique genes were taken for subsequent gene enrichment and RNA expression analysis (Harripaul et al., 2018; Harripaul et al., 2017a).

3.3.6 Adult human brain gene expression data

Adult human brain expression data was taken from the Allen Human Brain Atlas Portal (www.brain-map.org) and applied according to the methods described in (D. Howard et al., 2018). Briefly, the Atlas provides a transcriptional landscape of 16 adult control brains ranging from 24 to 57 years using custom Agilent 64 K microarrays and RNAseq. The Allen Institute normalized the expression values with a multistep process that adjusted for array batch effects. These gene expression data were used to map the gene expression for the gene lists constructed for this cohort, ID, and ASD across several tissues.

3.3.7 Prenatal human gene expression data

Similarly, the genes found to be associated with this cohort, ID and ASD phenotypes were also mapped against prenatal human gene expression data from (J. A. Miller et al., 2014). This dataset contains mid-gestational normal human brains which provides a reference for comparison of genes constructed from this cohort and other ID and ASD resources. This dataset used four

intact mid-gestational brains from 15-21 postnatal weeks (three females) that passed several exclusion criteria and were run on the same 64K Agilent custom microarrays. These methods were summarized and modified from (D. Howard et al., 2018).

3.3.8 Microarray gene expression data processing

Microarray probes were processed as per (D. Howard et al., 2018) where microarray probes were reannotated using Re-Annotator (Arloth, Bader, Röh, & Altmann, 2015) and samples with the same brain region were mean averaged to form a single expression profile. Expression levels were mean averaged and gene expression values were converted to ranks and z-score normalized.

3.3.9 Brain region enrichment analysis

Brain region enrichment was performed as per (D. Howard et al., 2018; Patel et al., 2019) where enrichment for the ID, ASD, and cohort gene sets within a brain region were calculated by ranking z-scores in an expression matrix per brain region. The gene sets of interest were then projected onto this ranking matrix. The area under the curve calculated as the area under the receiver operating characteristic (AUROC) curve statistic was used to quantify ID/ASD specific genes are more expressed. An AUROC statistic greater than 0.5 signifies that specific expression in a region of interest using theMann-Whitney test applied for statistical significance, and Benjamini-Hochberg was then used to correct the false discovery rate for testing within multiple brain regions.

3.3.10 Developmental Expression and Neuroanatomy of Genes Involved in ASD and ID

The temporal and neuroanatomical data were derived from the RNA-seq and microarray data representing two different resources. The Allen Institute of Brain Sciences has curated the BrainSpan Atlas of the Developing Brain, made up of gene expression data for up to 16 brains of both females and males from 8 weeks post-conception to 40 years old. Details can be found at http://help.brain-map.org/display/devhumanbrain/Documentation. PsychEncode provides tissue level RNA-seq data covering prenatal and postnatal development containing 36 human samples covering 8 weeks post-conception to 40 years old. Further details can be found here (Akbarian et

al., 2015). From these datasets, expression values were scaled across samples and z-scored across all brain regions. Mean expression was calculated for each gene in the gene list and brain region. The Wilcoxon signed-rank test was used to test differential expression from gene associated with ASD and ID to the background genes. Bonferroni corrections were used for adjusting p-values due to multiple testing biases.

3.4 Results

After alignment, 92.3% of the reads were above Q30 with a minimum coverage of 20 X across the exome and mean coverage of 54 X for all samples. In this study we sequenced the exomes of 115 trios (345 samples) and identified 38 previously associated genes with ID/ASD in 36 families resulting in a diagnostic yield of 31.3 %. We have found 81 *de novo* variants (Table 3.3) and 68 autosomal recessive variants (Table 3.4 and Table 3.2) in the dataset that included the 38 variants in genes that have been previously associated with ID and/or ASD (Table 3.2, Table 3.3, Table 3.4 and Table 3.5). In this cohort, we have identified 81 Loss of Function mutations (LoF), 45 of which were designated stop gain variants (Table 3.3). There were also a total of four genes identified to be known ASD/ID associated genes (three recessive and one *de novo*) causing splicing related defects representing 3.4% of the cohort with known disease associated genes (Table 3.3,Table 3.4 and Table 3.5).

Table 3.2: Percentage of Variants Identified by Inheritance Pattern for All Genes and Known Genes in this Study

	Percentage of Known Variants Compared to All Variants in Cohort	Percentage of Known Variants Compared to All Known Variants in Cohort	Percentage of All Inheritance Types to All Variants in Cohort
X-linked	4.10 (6/148)	15.8 (6/38)	12.1 (18/148)
Recessive	18.9	73.7 (28/38)	45.9 (68/148)

	(28/148)		
De novo	6.80 (10/148)	26.3 (10)	54.7 (81/148)
Yield	25.7 (38/148)		

The first column in Table 3.2 was calculated by comparing the previously known ID and ASD associated genes (38 associated genes) in the study cohort to all genes identified after variant filtering (148 genes). The second column was calculated by comparing the percentage of X-linked, recessive and *de novo* genes to all known genes identified in this cohort. The last column was calculated by comparing the X-linked, recessive and *de novo* genes to all variants found in this cohort.

3.4.1 *De novo* Variant Identification

De novo dominant mutations have been shown to be a major causal factor in ASD disease causality(Krumm et al., 2015). In this set of 148 identified variants, 81 variants were identified as *de novo* variants (78 autosomal and three X-linked) with 75 of the *de novo* variants being loss of function (LoF). Among the *de novo* variants identified, 10 variants have been identified in previous studies, and 10 of these variants were truncating LoF variants in the genes *CLMN, KMT2C, MECP2, MYT1L, NGLY1, PIK3CA, RAB3GAP2, SCN2A, USP45 and ZNF292*(Table 3.3 and Table 3.5).

In two of our ASD trios we identified *de novo* LoF mutations in the gene *ZNF292* that disrupt the zinc-finger DNA-binding domains. These two trios have contributed to a larger study of 25 families implicating *ZNF292* mutations in a spectrum of neurodevelopmental disorder ((Mirzaa et al., 2020)). Mutations in other genes that have previously been associated with ASD or ID were also identified. A known pathogenic mutation in *MECP2* (rs61750235) causes Rett Syndrome which is known to also have autistic features and this may be a case of variable expressivity (Table 3.5). *CLMN*, *KMT2C* and *MYT1L* were all previously found in studies of ID and have also been coming to the forefront in larger ASD studies (Grove et al., 2019). *De novo* mutations in *SCN2A* have been previously reported in multiple ASD studies ((Weiss et

al., 2003)). *NGLY1* is known to be involved in an inborn error of metabolism, where mutations in the N-glycanase 1 enzyme reduce enzyme function and the removal of glycan, which is an important step for the degradation of misfolded proteins (Enns et al., 2014). *PIK3CA* is another interesting gene which may cause ID in the form of hemimegalencephaly, which causes one side of the brain to enlarge and results in seizures and ID (J. H. Lee et al., 2012). Interestingly, *PIK3CA* has been most associated with cancers ((Samuels et al., 2004)), which underscores the extreme genetic pleiotropy and genetic heterogeneity associated with complex disease ((Dobyns & Mirzaa, 2019; Tatton-Brown et al., 2017)).

There were also several other genes that were listed as *de novo* variants that have not been associated with ASD (or ID) previously. These variants have met the criteria for filtering as described in the Materials and Method, and may represent novel putative variants for ASD. Despite no previous association, 37 variants have mouse models available that exhibit neurological or nervous system-related phenotypes (Figure 3.15). Where we have a LoF stop variant in *AEBP1*, an *Aebp1* mouse model has been generated that exhibited abnormal food intake and nursing (MGI:1197012) combined with a multi system phenotype involving many bodily systems including cardiovascular system, endocrine system and liver/biliary system. This gene causes defective collagen assembly and a variant ((Blackburn et al., 2018)). For *BAIAP2L2*, *baiap2l2* mice have abnormal startle reflexes. For *EFR3B*, *Efr3b* mice have both a neurological and a nervous system phenotype (MGI:2444851). See Figure 3.15). Sensory abnormalities such as impaired sensorimotor gating, prepulse inhibition, and startle reflex have often been associated with ASD (e.g.(Madsen, Bilenberg, Cantio, & Oranje, 2014).

SBF2 has been associated with Charcot-Marie Tooth Syndrome and was found in our cohort, and is supported by a mouse model that exhibits abnormal axon morphology, abnormal myelin sheath morphology (MGI:1921831). These genes may represent novel targets for ASD and may warrant further investigation.

In total, of the 148 variants identified, 38 variants have been previously associated with ID/ASD. Of these 38 variants that have been previously reported, 10 genes (eight were LoF) are *de novo* and associated with ASD/ID for a diagnostic yield of 31.3 % *de novo* variants and approximately 6.7% of variants in this cohort were previously reported *de novo* variants (Table 3.2 and Table 3.3).

3.4.2 Autosomal Recessive Variant Identification

This cohort yielded 68 recessive variants in 50 different families, of these 68 variants, 16 variants were LoF (Table 3.4). In total, this cohort had 28 variants that were previously associated with ID/ASD with six of these variants being LoF (Table 3.2 and Table 3.4).

There were several genes that had been previously reported in recent papers in autosomal recessive ID such as *TECPR2* (Anazi et al., 2017), *MTHFR* (Reuter et al., 2017), *MADD* (Anazi et al., 2017), *BTN3A2* (Anazi et al., 2017), *LINS1* (Najmabadi et al., 2011), *CC2D1A* (MIM 608443; MRT3; Basel-Vanagaite et al, 2006) and *ZNF335* (Hao Hu et al., 2019) (Table 3.4: Autosomal). Several other genes have been previously implicated in ID including *DEAF1*, *DRP2*, *RSRC1*, *DAGLA*, *WDR27*, *SHH*, *ASL* and *ASPA* (Table 3.4). Other known genes were also present in a recessive form including *CUL7*, *MSN (Kerin et al., 2012)*, *PLXNA3*, *SETD2* and *SGSM3* (Table 3.4).It is noteworthy that some variants have been associated with inborn errors in metabo which include ID, such as *ASPA*, *AGA*, *AIFM1*, *ASL* and *HTRA2* (Table 3.4Table 3.4: Autosomal). In one trio, a homozygous LoF mutation was found in *VPS13B*- the gene responsible for Cohen Syndrome (MIM 216550; COH1).

Mouse models were also used to identify putative variants that may potentially alter cognitive development that have not been identified previously in literature. *Cnga2* knockout mice (MGI:108040) have craniofacial and nervous system abnormalities which may represent a syndromic form of ASD/D. Pm20d1 (MGI:2442939), Rasal2 (MGI:2443881) and Vps16 (MGI:2136772) knockout mice all have behavioral/neurological and nervous system phenotypes. It is worth noting, that VPS16 is a vesicle mediated trafficking protein similar to *VPS13B*, and is related to dystonia. Functional models can be further viewed in Figure 3.16.

In total, of the 148 variants identified, there 38 variants have been previously associated with ID/ASD. Of these 38 variants that have been previously reported, 28 genes (seven were LoF) are recessive and associated with ASD/ID representing 73.6 % recessive variants for known ID/ASD genes and approximately 18.9% of variants in this cohort were previously known and reported recessive variants (Table 3.2 and Table 3.4)

There are also 18 variants that have come from the X chromosome including known genes such as *MECP2, PLXNA3, MSN, AIFM1* and *DPR2* (Table 3.5). A total, of five known genes in six

families along with other variants that fulfill the filtering criteria represent a distinct portion that is X related with a majority, 15 variants, coming in X-linked recessive forms and three in *de novo* forms (Table 3.5).

Of these 38 variants that have been previously reported, 5 genes (were LoF) are X-linked (five genes were recessive and there was a single *de novo* variant in *MeCP2*) and associated with ASD/ID contributing 15.7 % X-linked (13.1 % recessive and 2.6 % *de novo* previously reported variants) variants and approximately 4 % of variants in this cohort (3.3 % recessive and 0.7 % *de novo*) were previously reported X-linked variants (Table 3.5)

Table 3.3:Autosomal *De novo* sequence variants identified in Pakistani, Iranian and Saudi ASD trio families by Genomic Coordinate (n=115, Pakistani Trios n = 58, Iranian Trios n= 41 , Saudi Trios n= 16).

Family	Mode of Inheritance	Chromosome	Start	End	Reference	Alternative	Gene	Variant Type	Amino Acid Change	gnomAD Exome All Frequency	gnomAD Exome South Asian Frequency
IABB-16	*De novo*	chr1	152681631	.	C	CTGTAGCATTAAGATAGTTAATAGCTAATATTTACT	LCE4A	stopgain	NM_178356: exon1: c.80_81insTGTAGCATTAAGATAGTTAATAGCTAATATTTACT: p.S27_K28delinsSVALRX	0	0
IAH-3	*De novo*	chr1	231057224	.	AG	A	TTC13	stopgain	NM_001122835: exon14: c.1681delC: p. L561X	0	0
IAH-4	*De novo*	chr1	220366717	220366718	CA	-	RAB3GAP2**	frameshift deletion	NM_012414: exon13: c.1134_1135del: p. F378fs	0	0
IAH-4	*De novo*	chr1	225528168	225528168	A	-	DNAH14	frameshift deletion	NM_001373: exon67: c.10164delA: p. L3388fs	0	0
IAH-4	*De novo*	chr1	241731824	241731825	CC	-	KMO	frameshift deletion	NM_003679: exon10: c.834_835del: p. F278fs	0	0
SMPA-31	*De novo*	chr1	153907306	rs375088543	T	TGC	DENND4B	frameshift insertion	NM_014856: exon18: c.2702_2703insGC: p. Q901fs	0	0
SMPA-71	*De novo*	chr1	171810836	171810836	A	-	DNM3	frameshift deletion	NM_001136127: exon1: c.40delA: p. N14fs	0	0
RQPA-20	*De novo*	chr1	150483966	150483966	G	T	ECM1	stopgain	NM_001202858: exon7: c. G823T: p. E275X	0	8.12E-06

RQPA-20	*De novo*	chr1	160722912	160722912	C	A	SLAMF7	stopgain	NM_001282589: exon5:c.C560A: p. S187X	0	0
RQPA-20	*De novo*	chr1	112031311	112031311	A	-	TMIGD3	frameshift deletion	NM_001302680: exon2: c.286delT: p. W96fs	0	0
RQPA-20	*De novo*	chr1	28919246	28919246	G	A	RAB42	nonsynonymous SNV	NM_001193532: exon1: c. G223A: p. E75K	0	0
IAH-4	*De novo*	chr2	197531465	197531465	G	-	CCDC150	frameshift deletion	NM_001080539: exon7: c.785delG: p.C262fs	0	0
IAH-4	*De novo*	chr2	202552068	202552068	-	G	MPP4	frameshift insertion	NM_033066: exon5: c.305dupC: p. P102fs	0	0
IAU-25	*De novo*	chr2	25372563	25372563	C	T	EFR3B	stopgain	NM_014971: exon20: c.C2143T: p. R715X	0	0
IAU-29	*De novo*	chr2	1946865	1946865	C	-	MYT1L*/**	frameshift deletion	NM_001303052: exon9: c.394delG: p. E132fs	0	0
IAU-66	*De novo*	chr2	232320177	232320177	C	-	NCL	frameshift deletion	NM_005381: exon13: c.1991delG: p. G664fs	0	0
SMPA-53	*De novo*	chr2	166166958	166166958	C	T	SCN2A*/**	stopgain	NM_001040143: exon6:c.C823T: p. R275X	0	0
IABB-11	*De novo*	chr3	178921485	.	T	TAGA	PIK3CA*/**	stopgain	NM_006218: exon5: c.967_968insAGA: p. S323delinsX	0	0
IAH-3	*De novo*	chr3	17109533	.	T	TCTCTAAAATAAGAATAACAATGATCATAGGTGCCCTGTTTCCCTTCAGAACAATGAGGAACAATTGCTG	PLCL2	stopgain	NM_015184: exon5: c.2802_2803insCTCTAAAATAAGAATAACAATGATCATAGGTGCCCTGTTTCCCTTCAGAACAATGAGGAACAATTGCTG: p. F934delinsFLX	0	0

SMPA-10	*De novo*	chr3	108117585	108117585	C	A	MYH15	stopgain	NM_014981: exon36: c.G5092T: p. E1698X	0	0
IAU-25	*De novo*	chr4	48493038	48493038	G	T	ZAR1	stopgain	NM_175619: exon1: c.G730T: p. E244X	0	0
IAU-51	*De novo*	chr4	190862176	190862176	C	A	FRG1	stopgain	NM_004477: exon1: c.C12A: p.Y4X	0	0
IAU-56	*De novo*	chr4	123171555	123171555	G	T	KIAA1109	stopgain	NM_015312: exon35: c.G5749T: p. G1917X	0	0
RQPA-20	*De novo*	chr4	17608497	17608497	G	T	LAP3	stopgain	NM_015907: exon12: c.G1306T: p. E436X	0	0
RQPA-20	*De novo*	chr4	52938138	52938138	A	-	SPATA18	frameshift deletion	NM_001346102: exon4: c.163delA: p. S55fs	0	0
SMPA-75	*De novo*	chr5	137680632	137680632	-	T	FAM53C	frameshift insertion	NM_001135647: exon4: c.256dupT: p. N85fs	0	0
Autism-10	*De novo*	chr6	87969506	87969507	AG	-	ZNF292*/**	frameshift deletion	NM_015021: exon8: c.6159_6160del: p. T2053fs	0	0
IAH-3	*De novo*	chr6	34838816	.	C	CTTCTTTCAAGTAACATGATATTCTTTAAATTCATCTAGGTTGTTGAATTAGCTATAATTTGTTCATTTTT	UHRF1BP1	stopgain	NM_017754: exon18: c.3904_3905insTTCTTTCAAGTAACATGATATTCTTTAAATTCATCTAGGTTGTTGAATTAGCTATAATTTGTTCATTTTT: p.P1302_F1303delinsLLSSNMIFFKFIX	0	0
IAU-65	*De novo*	chr6	87969506	87969507	AG	-	ZNF292*/**	frameshift deletion	NM_015021: exon8: c.6159_6160del: p. T2053fs	0	0
RQPA-13	*De novo*	chr6	99883698	99883698	G	T	USP45*	stopgain	NM_001346024: exon16:c.C2165A: p.S722X,	0	0
IAU-71	*De novo*	chr7	65419159	.	C	G	VKORC1L1	nonsynonymous SNV	NM_001284342: exon2:c.C293G: p. A98G	0	0
SA-1	*De novo*	chr7	151949792	.	T	A	KMT2C*/**	nonsynonymous SNV	NM_170606: exon10: c.A1308T: p. R436S	0	0

SMPA-39	*De novo*	chr7	44144328	44144328	G	T	AEBP1	stopgain	NM_001129: exon1: c. G64T: p. G22X	0	0
RQPA-13	*De novo*	chr7	3366665	3366665	G	T	NFIC	splicing	NM_001245004: exon1: c.30+1G>T	0	0
IABB-10	*De novo*	chr8	6669525	.	T	TA	XKR5	stopgain	NM_207411: exon8: c.1254dupT: p.K419_I420delinsX	0	0
IABB-8	*De novo*	chr8	10467671	.	C	CTTCTGTTTTAGTT	RP1L1	stopgain	NM_178857: exon4: c.3936_3937insAACTAAAACAGAA: p.E1313_G1314 delensing	0	0
IABB-8	*De novo*	chr8	67341550	.	C	T	RRS1	stopgain	NM_015169: exon1: c.C184T: p. Q62X	0	0
IAU-79	*De novo*	chr8	67047342	.	C	G	TRIM55	nonsynonymous SNV	NM_033058: exon3: c.C459G: p.H153Q	0	0
IAU-56	*De novo*	chr8	10480504	10480504	G	A	RP1L1	stopgain	NM_178857: exon2: c.C208T: p. Q70X	0	0
IAH-3	*De novo*	chr9	6493889	.	A	AAAAGTTAACAGTTTTTATATCCTGTTACTAAGGGTATTTTGTCGATTGAGTAAAAGTGAGTGTTTTTCAC	UHRF2	stopgain	NM_152896: exon10: c.1561_1562insAAAGTTAACAGTTTTTATATCCTGTTACTAAGGGTATTTTGTCGATTGAGTAAAAGTGAGTGTTTTTCAC: p.I521_G522delins KSX	0	0
SMPA-38	*De novo*	chr9	140065894	140065894	C	A	TMEM210	stopgain	NM_001282477: exon3: c. G226T: p. E76X	3.09E-05	0.0002
SMPA-40	*De novo*	chr9	135863604	135863604	C	T	GFI1B	stopgain	NM_001135031: exon4:c.C259T: p. R87X	0	0
SMPA-39	*De novo*	chr10	45496158	45496159	AG	-	C10orf25	frameshift deletion	NM_001039380: exon1: c.149_150del: p. S50fs	1.22E-05	9.76E-05

IAH-3	*De novo*	chr11	9830504	.	C	CATTATACAAAAGAGTCATCTGAAGAATTTAAAAATGACCAAATGGAAATATAAATATAAAATATCAA	SBF2	stopgain	NM_030962: exon31: c.4200_4201insTTGATATTTTATATTTATATTTCCATTTGGTCATTTTTAAATTCTTCAGATGACTCTTTTGTATAAT: p.E1401_N1402delinsLIFYIYISIWSFLNSSDDSFVX	0	0
IAH-3	*De novo*	chr11	18361164	.	T	TAACCAAGGTTTGTAGTGATTTTTATGACTATACAAGTTTTATCTTAAAGCTCCACTGCC	GTF2H1	stopgain	NM_005316: exon5: c.567_568insAACCAAGGTTTGTAGTGATTTTTATGACTATACAAGTTTTATCTTAAAGCTCCACTGCC: p.T189_S190delinsTNQGLX, GTF2H1	0	0
IAU-71	*De novo*	chr11	10785300	.	CAAAG	C	CTR9	frameshift deletion	NM_014633: exon9: c.1069_1072del: p. K357fs	0	0
RQPA-20	*De novo*	chr11	46389550	46389550	A	C	DGKZ	splicing	NM_201533: exon4: c.379-2A>C	0	0
RQPA-20	*De novo*	chr11	22051034	22051034	A	G	PPIL2	splicing	NM_148175: exon21: c.1570-2A>G	0	0
IAH-3	*De novo*	chr12	14836132	.	T	TGAGATCTGATGGTTTTATAAGGGGCTCTTCCCTTTTTGCTTAG	GUCY2C	stopgain	NM_004963: exon4: c.454_455insCTAAGCAAAAAGGGAAGAGCCCCTTATAAAACCATCAGATCTC: p.D152_Y153delinsAKQKGKSPLX	0	0
IAU-73	*De novo*	chr12	89998008	89998008	G	C	ATP2B1	stopgain	NM_001001323: exon15:c.C2558G: p. S853X	0	0
IAU-56	*De novo*	chr12	11546888	11546888	G	A	PRB2	stopgain	NM_006248: exon3: c.C124T: p. Q42X	0	0

SMPA-39	*De novo*	chr12	49916625	49916625	C	A	SPATS2	stopgain	NM_001293286: exon11:c.C1100A: p. S367X	0	0
SMPA-39	*De novo*	chr12	8687235	8687235	-	A	CLEC4E	stoploss	NM_014358: exon6: c.658dupT: p.X220delinsL	1.22E-05	9.81E-05
IAH-4	*De novo*	chr14	56103239	56103239	G	T	KTN1	stopgain	NM_001079521: exon10: c. G1534T: p. E512X	0	0
SA-15	*De novo*	chr14	95682030	95682030	C	-	CLMN**	frameshift deletion	NM_024734: exon5: c.365delG: p. G122fs	0	0
IABB-8	*De novo*	chr15	36606831	.	G	C	TBCB	splicing	NM_001300971: exon1: UTR5	0	0
IAH-3	*De novo*	chr15	42107821	.	G	A	MAPKBP1	splicing	NM_001265611: exon12: c.1318-1G>A	0	0
IAH-3	*De novo*	chr16	75481538	.	G	GATCTT AATCA	TMEM170A	frameshift insertion	NM_001304996: exon3: c.283_284ins TGATTAAGAT: p. P95fs	0	0
RQPA-13	*De novo*	chr16	85682175	85682175	G	T	GSE1	stopgain	NM_001278184: exon2: c. G25T: p.E9X	0	0
RQPA-20	*De novo*	chr16	89006655	89006655	G	A	LOC100129697	nonsynonymous SNV	NM_001290330: exon1: c. G53A: p. R18K	0	0
IAH-3	*De novo*	chr17	20013765	.	C	CA	SPECC1	frameshift insertion	NM_001033553: exon3: c.173_174ins A: p. T58fs	0	0
IAU-25	*De novo*	chr17	1650438	1650438	A	G	SERPINF2	nonsynonymous SNV	NM_001165921: exon5: c. A301G: p. R101G	0	0

IAU-56	*De novo*	chr17	47895215	47895215	G	T	KAT7	stopgain	NM_001199158: exon6: c. G490T: p. G164X	0	0
RQPA-13	*De novo*	chr17	24880403	24880403	C	A	OSBPL3	splicing	NM_015550: exon14: c.1402-1G>T	0	0
RQPA-20	*De novo*	chr17	36638358	36638358	G	T	ARHGAP23	stopgain	NM_001199417: exon15: c. G2725T: p. E909X	0	0
IAH-3	*De novo*	chr18	59170279	.	G	GCCATTTTTGAAGCTGAGATTTTTTTACCTCAAATTTAAAAGCACAGGTGACTTGCACCTCACTTCT	CDH20	stopgain	NM_031891: exon5: c.755_756insCCATTTTTGAAGCTGAGATTTTTTTACCTCAAATTTAAAAGCACAGGTGACTTGCACCTCACTTCT: p. G252delinsGHFX	0	0
RQPA-13	*De novo*	chr18	12825849	12825849	G	T	PTPN2	stopgain	NM_001308287: exon4:c.C368A: p. S123X	0	0
IABB-11	*De novo*	chr19	25805804	.	T	A	NGLY1***	splicing	NM_001145294: exon3: c.121-2A>T	0	0
IAH-3	*De novo*	chr19	9677281	.	T	TTA	ZNF121	stopgain	NM_001008727: exon4: c.507_508insTA: p.K170_P171 delinsX	0	0
IAH-3	*De novo*	chr19	32954351	.	A	ACTGAATTCCAAGTTAATGGATGTGTGTTTTGCTTTTGCTTATTTGTAAAAATTTGAAAAATACTCATAATAAC	DPY19L3	stopgain	NM_001172774: exon13: c.1402_1403 insCTGAATTCCAAGTTAATGGATGTGTGTTTTGCTTTTGCTTATTTGTAAAAATTTGAAAAATACTCATAATAAC: p.I468_H469 delinsTEFQVNGCVFCFCLFVKIX	0	0

SA-13	*De novo*	chr19	43583862	.	C	T	TOMM34	splicing	NM_006809: exon3: c.228-1G>A	0	0
SMPA-36	*De novo*	chr19	51358216	51358216	G	-	KLK3	frameshift deletion	NM_001030047: exon1: c.5delG: p. W2fs	0	1.22E-05
SMPA-75	*De novo*	chr19	7734220	7734221	CT	-	RETN	frameshift deletion	NM_001193374: exon2: c.8_9del: p. A3fs	4.47E-05	0.0003
SMPA-77	*De novo*	chr19	44832906	44832907	AT	-	ZNF112	frameshift deletion	NM_001348284: exon3: c.1220_1221 del: p. Y407fs	0	4.07E-06
IAH-3	*De novo*	chr20	18445972	.	T	TCACCCAGGTTCAAGCAATTCTCCTGCCTCAACCTC	DZANK1	stopgain	NM_001099407: exon2: c.30_31ins GAGGTTGAGGCAGGAGAATTGCTTGAACCTGGGTG: p.I11_I12delinsEVEAGELLEPGX	0	0
IAH-3	*De novo*	chr22	27019240	.	G	GGTTTTATAAAGCTTTTGGATTTATCATCAGCCTTTAAAAGCCTGGAAGAACCACATAAA	CRYBA4	frameshift insertion	NM_001886: exon3: c.82_83insGTTTTATAAAGCTTTTGGATTTATCATCAGCCTTTAAAAGCCTGGAAGAACCACATAAA: p. E28fs	0	0
IAH-3	*De novo*	chr22	38493109	.	T	TCCACACATGAATAAATAAATGCATGAATTCGTGAATGCATTAAAAACAAACAATA	BAIAP2L2	frameshift insertion	NM_025045: exon7: c.541_542insTATTGTTTGTTTTTAATGCATTCACGAATTCATGCATTTATTTATTCATGTGTGG: p. K181fs	0	0
SMPA-36	*De novo*	chr22	17630575	17630576	CT	-	HDHD5	frameshift deletion	NM_017829: exon2: c.96_97del: p. R32fs	0	0

* Previously associated with Autism Spectrum Disorder
** Previously associated with Intellectual Disability
*** Previously associated with an inborn error of metabolism
All genomic coordinates are based on the UCSC hg19 build

Table 3.4: Autosomal recessive genes and sequence variants identified in Pakistani and Iranian ASD trio families by (hg19) genomic coordinate (n=115, Pakistani Trios n = 58, Iranian Trios n= 41 , Saudi Trios n= 16)

Family	Mode of Inheritance	Chromosome	Start	End	Reference	Alternative	Gene	Variant Type	Amino Acid Change	gnomAD Exome All Frequency	gnomAD Exome South Asian Frequency
SMPA-3	Recessive	chr1	11855404	11855404	C	A	MTHFR*/** Reuter et al, 2017	nonsynonymous SNV	NM_001330358: exon6: c. G905T: p. G302V	4.07E-06	3.25E-05
SMPA-3	Recessive	chr1	178425880	178425880	C	A	RASAL2	nonsynonymous SNV	NM_004841: exon11: c.C1813A: p. L605I	0.0008	0.006
SMPA-4	Recessive	chr1	152193673	152193682	GACGTTTC TG	-	HRNR	frameshift deletion	NM_001009931: exon3: c.423_432del: p. S141fs	4.467E-05	0.0004
SMPA-40	Recessive	chr1	197621364	197621364	C	A	DENND1B	splicing	NM_001195215: exon7: c.447+1G>T	0	0
SMPA-76	Recessive	chr1	166891853	166891854	CT	-	ILDR2	frameshift deletion	NM_199351: exon8: c.1187_1188del: p. E396fs	0	4.07E-06
SMPA-76	Recessive	chr1	205813938	205813938	C	T	PM20D1	splicing	NM_152491: exon4: c.576+1G>A	0	4.88E-05
Autism-8	Recessive	chr2	74759786	74759786	G	T	HTRA2*MIM 617248	nonsynonymous SNV	NM_013247: exon7: c.G1156T: p. D386Y	8.12E-06	6.50E-05
Autism-3	Recessive	chr3	157840002	157840002	C	T	RSRC1**	stopgain	NM_001271834: exon2:c.C109T: p. R37X	0	0
SA-15	Recessive	chr3	47164429	.	G	C	SETD2*	nonsynonymous SNV	NM_001349370: exon2:c.C1565G: p. S522C	0	0
SMPA-8	Recessive	chr3	169524635	.	C	T	LRRC34	splicing	NM_001172779: exon5: c.528+1G>A	0	0
SMPA-19	Recessive	chr3	38802185	.	C	A	SCN10A	stopgain	NM_001293306: exon7: c. G937T: p. G313X	0	0

SMPA-27	Recessive	chr3	134911657	134911657	T	C	EPHB1	nonsynonymous SNV	NM_004441: exon11: c.T2122C: p. F708L	4.07E-06	3.25E-05
IAU-66	Recessive	chr4	101108876	101108876	T	-	DDIT4L	frameshift deletion	NM_145244: exon3: c.540delA: p. K180fs	0	0
SMPA-35	Recessive	chr4	185033930	.	CAT	C	ENPP6	frameshift deletion	NM_153343: exon6: c.886_887del: p.M296fs	0	0
RQPA-12	Recessive	chr4	178360822	178360822	G	A	AGA***	nonsynonymous SNV	NM_000027: exon3: c.C302T: p. A101V	0	4.06E-06
IAH-4	Recessive	chr5	42769698	42769698	G	T	CCDC152	splicing	NM_001134848: exon4: c.194-1G>T	0	0
Autism-8	Recessive	chr5	150846823	150846823	-	CTGG	SLC36A1***	frameshift insertion	NM_001308150: exon6: c.483_484insCTGG: p. F161fs	8.00E-06	6.60E-06
IAU-56	Recessive	chr5	154199963	154199963	C	-	FAXDC2	frameshift deletion	NM_032385: exon9: c.915delG: p. Q305fs	5.35E-05	0.0004
Autism-9	Recessive	chr6	170058095	170058095	C	A	WDR27*	nonsynonymous SNV	NM_001350625: exon13: c. G1246T: p. G416W	0	0
IABB-8	Recessive	chr6	38851674	.	A	T	DNAH8	nonsynonymous SNV	NM_001206927: exon56: c. A8159T: p. D2720V	0	0
IAU-80	Recessive	chr6	26368279	rs72841509	G	A	BTN3A2** Anazi et al, MP2017	splicing	NM_001197247: exon2: UTR5	0	0
SA-15	Recessive	chr6	43011209	.	G	A	CUL7*	nonsynonymous SNV	NM_001168370: exon17:c.C3584T: p. P1195L	0	0
SMPA-48	Recessive	chr6	13634723	13634723	C	G	RANBP9	nonsynonymous SNV	NM_005493: exon11: c.G1735C: p. G579R	0	2.44E-05
IABB-7	Recessive	chr7	99720009	.	G	C	CNPY4	splicing	NM_152755: exon2: c.245+1G>C	0	0
IAU-7	Recessive	chr7	47944924	.	T	C	PKD1L1	splicing	NM_138295: exon11: c.1523-2A>G	0	0
IAU-79	Recessive	chr7	150174821	.	TCC	T	GIMAP8	frameshift deletion	NM_175571: exon5: c.1952_1953del: p. S651fs	0	0
SA-7	Recessive	chr7	65551627	.	C	T	ASL***	nonsynonymous SNV	NM_001024943: exon6:c.C502T: p. R168C	0	0

SMPA-11	Recessive	chr7	155604611	.	T	A	SHH**	nonsynonymous SNV	NM_000193: exon1: c.A206T: p. N69I	0	0
IABB-3	Recessive	chr8	100155379	100155379	AG	A	VPS13B***	frameshift deletion	NM_015243: exon13: c.1829delG: p. S610fs	0	0
SA-4	Recessive	chr9	5805613	5805613	T	C	ERMP1	nonsynonymous SNV	NM_024896: exon9: c.A1721G: p.H574R	0	0
IABB-2	Recessive	chr11	679701	.	G	GA	DEAF1*/**	frameshift insertion	NM_001293634: exon6: c.845dupT: p. F282fs	0	0
SMPA-8	Recessive	chr11	118403863	.	A	G	TMEM25	nonsynonymous SNV	NM_001144036: exon3: c. A302G: p. N101S	0	0
SMPA-20	Recessive	chr11	47305749	.	A	G	MADD** Anafi et al 2017, Hu et al, 2019	nonsynonymous SNV	NM_001135943: exon11: c. A1883G: p. Y628C	0	0
SMPA-21	Recessive	chr11	61495635	61495635	C	G	DAGLA*** SCA20	stopgain	NM_006133: exon7:c.C647G: p. S216X	0	0
SMPA-41	Recessive	chr11	47194985	47194985	G	A	ARFGAP2	nonsynonymous SNV	NM_001242832: exon6:c.C527T: p. A176V	0	0
SMPA-80	Recessive	chr11	67290028	67290028	C	T	CABP2	nonsynonymous SNV	NM_016366: exon2: c.G202A: p. A68T	0	6.60E-05
IABB-10	Recessive	chr12	69263140	.	C	G	CPM	nonsynonymous SNV	NM_001005502: exon6: c. G752C: p. G251A	0	0
IAU-79	Recessive	chr12	58021505	.	GGGAAGCC GACGAGCT CGTGGT	G	B4GALNT1	nonframeshift deletion	NM_001276468: exon9: c.1094_1114 del: p.365_372del	0	0
IAU-73	Recessive	chr13	95254182	95254182	A	T	GPR180	nonsynonymous SNV	NM_180989: exon1: c.A1T: p.M1L	1.02E-05	0
IAH-2	Recessive	chr14	102963429	.	TGTGGGGA CCGCCTG	T	TECPR2*** Anazi et al, 2017	frameshift deletion	NM_014844: exon18: c.3904_3917del: p. V1302fs	0	0
SMPA-8	Recessive	chr14	65032074	.	G	C	PPP1R36	splicing	NM_172365: exon5: c.270-1G>C	0	0

IABB-4	Recessive	chr15	101113962	101113962	T	-	LINS1** MRT27	frameshift deletion	NM_001040616: exon5: c.1116delA: p. E372fs	0	0
SMPA-76	Recessive	chr16	716004	716004	C	T	WDR90	nonsynonymous SNV	NM_145294: exon36: c.C4489T: p. R1497W	3.24E-05	0
IABB-14	Recessive	chr17	3392637	.	G	T	ASPA**	splicing	NM_001128085: exon5: c.634+1G>T	0	0
IABB-16	Recessive	chr19	50321659	.	G	A	MED25**	nonsynonymous SNV	NM_030973: exon1: c.G61A: p. V21M	0	0
IABB-16	Recessive	chr19	53411166	rs7255668	A	C	ZNF888	nonsynonymous SNV	NM_001310127: exon3: c. T409G: p. Y137D	0	0
SA-2	Recessive	chr19	14024452	14024452	G	T	CC2D1A*/**	splicing	NM_017721: exon6: c.748+1G>T	0	0
IABB-16	Recessive	chr20	2847201	.	C	A	VPS16	nonsynonymous SNV	NM_080413: exon20: c.C2069A: p. A690E	0	0
IABB-16	Recessive	chr20	16360737	.	A	G	KIF16B	nonsynonymous SNV	NM_001199865: exon19: c. T1910C: p. V637A	0	0
IABB-5	Recessive	chr20	44578999	.	C	T	ZNF335** Hu et al, 2019	nonsynonymous SNV	NM_022095: exon22: c.G3346A: p. G1116R	0	0
SMPA-9	Recessive	chr20	31531544	rs13045180	C	T	EFCAB8	stopgain	NM_001143967: exon23:c.C2893T: p. Q965X	0	0
SA-15	Recessive	chr22	46792575	.	G	A	CELSR1**	nonsynonymous SNV	NM_014246: exon13: c.C5770T: p. R1924C	0	0

* Previously associated with Autism Spectrum Disorder
** Previously associated with Intellectual Disability
*** Previously associated with an inborn error of metabolism
All genomic coordinates are based on the UCSC hg19 build

Table 3.5: *De novo* and recessive X- linked genes and sequence variants identified in Pakistani and Iranian ASD trio families by (hg19) genomic coordinate (n=115, Pakistani Trios n = 58, Iranian Trios n= 41 , Saudi Trios n= 16)

Family	Mode of Inheritance	Chromosome	Start	End	Reference	Alternative	Gene	Variant Type	Amino Acid Change	gnomAD Exome All Frequency	gnomAD Exome South Asian Frequency
Autism-6	Recessive	chrX	111019623	111019623	A	G	TRPC5	nonsynonymous SNV	NM_012471: exon11: c.T2840C: p. F947S	4.59E-05	0.0004
IABB-8	Recessive	chrX	152106616	.	A	G	ZNF185	splicing	NM_001178110: exon13: c.942-2A>G	0	0
IAU-24	Recessive	chrX	129290446	.	C	T	AIFM1***	nonsynonymous SNV	NM_001130847: exon2: c. G238A: p. A80T	0	0
IAU-45	Recessive	chrX	64947718	.	C	A	MSN*	nonsynonymous SNV	NM_002444: exon3: c.C139A: p. L47M	0	0
IAU-69	Recessive	chrX	55779942	.	T	C	RRAGB	nonsynonymous SNV	NM_006064: exon7: c.T730C: p. F244L	0	0
IAU-78	Recessive	chrX	127185449	.	C	T	ACTRT1	nonsynonymous SNV	NM_138289: exon1: c.G737A: p. G246E	0	0
SMPA-15	Recessive	chrX	105193588	105193588	C	T	NRK	stopgain	NM_198465: exon27: c.C4375T: p. Q1459X	0	0
SMPA-15	Recessive	chrX	108625363	108625363	C	T	GUCY2F	nonsynonymous SNV	NM_001522: exon17: c.G3134A: p. G1045E	6.25E-05	0.0006
SMPA-35	Recessive	chrX	134987552	.	G	A	SAGE1	splicing	NM_018666: exon5: c.454+1G>A	0	0
SMPA-41	Recessive	chrX	153689047	153689047	C	T	PLXNA3*	nonsynonymous SNV	NM_017514: exon2: c.C524T: p. S175L	4.62E-05	7.37E-05
SMPA-43	Recessive	chrX	13621531	.	GC	G	EGFL6	frameshift deletion	NM_001167890: exon5: c.497delC: p. A166fs	0	0
SMPA-43	Recessive	chrX	37535037	.	G	A	LANCL3	nonsynonymous SNV	NM_001170331: exon5: c. G1234A: p. E412K	0	0
SMPA-70	Recessive	chrX	150912120	rs147982375	G	A	CNGA2	nonsynonymous SNV	NM_005140: exon7: c.G1145A: p. R382Q	0	0

SMPA-84	Recessive	chrX	100507601	100507601	A	C	DRP2**	nonsynonymous SNV	NM_001171184: exon15: c. A1639C: p. S547R	0	0
RQPA10	Recessive	chrX	153296072	153296072	G	T	MECP2*/**	nonsynonymous SNV	NM_001110792: exon3:c.C1243A: p. P415T	0	0
IAH-3	*De novo*	chrX	71843114	.	C	CT	PHKA1	frameshift insertion	NM_001122670: exon18: c.1804_1805 ins A: p. G602fs	0	0
SMPA-38	*De novo*	chrX	47646871	47646871	G	A	ADGRF2	splicing	NM_153839: exon5: c.267+1G>A	4.07E-06	3.25E-05
SMPA-78	*De novo*	chrX	153296516	153296516	G	A	MECP2*/**	stopgain	NM_001110792: exon3:c.C799T: p. R267X	0	0

* Previously associated with Autism Spectrum Disorder
** Previously associated with Intellectual Disability
*** Previously associated with inborn errors of metabolism
All genomic coordinates are based on the UCSC hg19 build

3.4.3 Developing Human Brain Atlas Cortical Layer Identification from Developing Human Brain

The neocortex of the mammalian brain is attributed as having higher order brain functions including sensory perception, cognition, generation of motor commands, spatial reasoning and language. In light of this, the cortical developmental zones in the developing brain were interrogated using RNA-seq expression data. Gene lists were constructed by conglomerating genes (See Materials and Methods) that were involved in either ID or ASD from SFARI, OMIM and more recent literature. After stratifying cortical expression developmental zones that had an AUROC > 0.5, which implies higher than normal expression, the cortical plate, intermediate zone and subplate were shared among all gene lists including ASD and ID genes. In our cohort, genes with rare variants showed the subventricular, subpial granular zone, ventricular zone and subplate zone all exhibited above normal expression (**Table 3.6**). Despite having more genes than those from our cohort, the ASD gene list showed cortical expression in the intermediate zone, subplate zone, marginal zone. In a similar manner, the intermediate zone, subplate zone, marginal zone and cortical plate were significantly expressed in the ASD gene list (Table 3.6: Developmental Cortical Layer Expression of ASD Genes Constructed from Literature

). Despite having more genes associated with ID than ASD and significantly more genetic heterogeneity, the developmental cortical developmental zones that were significantly expressed were the cortical plate, the subplate zone, subventricular zone and marginal zone (Table 3.7). The entire list of genes for ID and ASD also provided similar results with the subplate zone, intermediate zone, cortical plate and marginal zone (Table 3.8). Through all gene sets, it seems that the cortical plate, subplate and intermediate zone had the most highly expressed genes in their developmental developmental zones.

Table 3.6: Developmental Cortical Layer Expression of ASD Genes Constructed from Literature

zone	AUROC	pValue	adjusted_P	laminar_order
intermediate zone	0.580	3.18e-18	2.23e-17	3
subplate zone	0.563	1.19e-11	5.95e-11	4
marginal zone	0.534	2.59e-04	5.18e-04	6
cortical plate	0.531	9.04e-04	9.04e-04	5
subpial granular zone	0.456	2.02e-06	6.06e-06	7
subventricular zone	0.450	5.29e-08	2.12e-07	2
ventricular zone	0.437	7.45e-12	4.47e-11	1

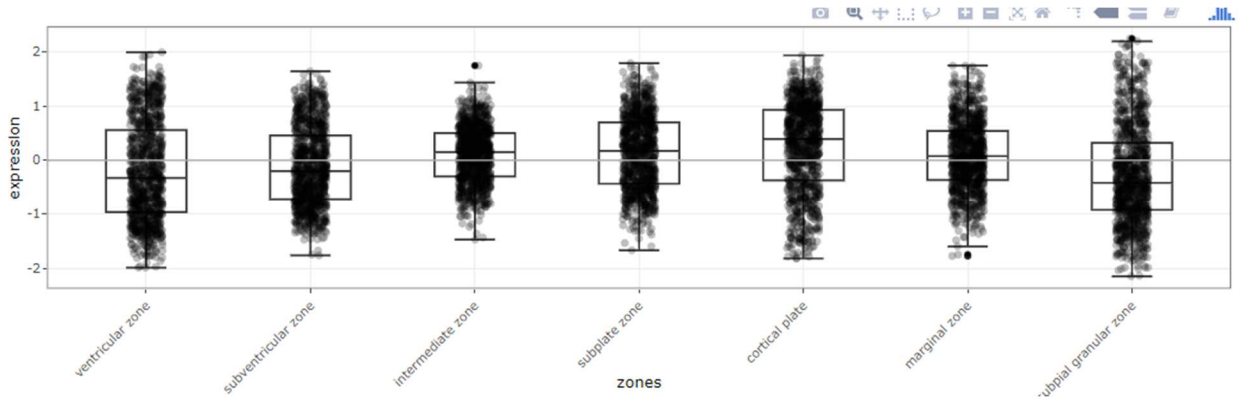


Table 3.7:Developmental Cortical Layer Expression in ID Genes Constructed from Literature

zone	AUROC	pValue	adjusted_P	laminar_order
cortical plate	0.534	7.01e-05	4.21e-04	5
subplate zone	0.531	2.35e-04	1.18e-03	4
subventricular zone	0.517	4.42e-02	1.77e-01	2
intermediate zone	0.515	6.89e-02	2.07e-01	3
ventricular zone	0.504	6.39e-01	6.39e-01	1
marginal zone	0.489	2.13e-01	4.26e-01	6
subpial granular zone	0.436	3.57e-14	2.50e-13	7

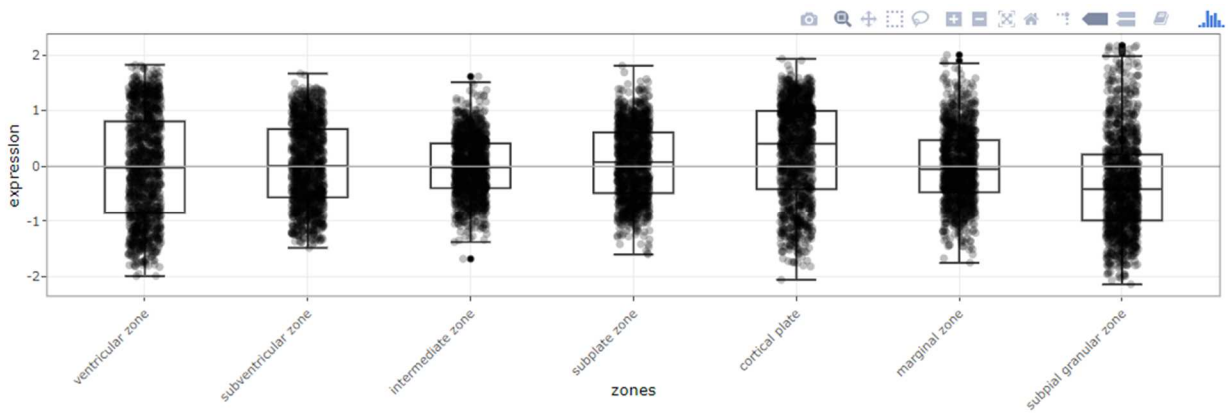
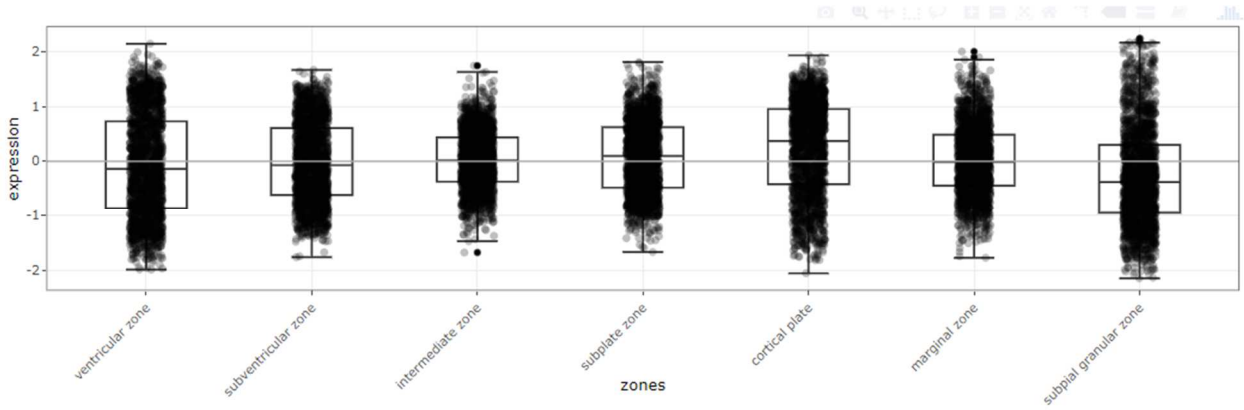


Table 3.8: Developmental Cortical Expression in Both ID, ASD from Literature and Genes Identified in Pakistani and Iranian Trio Families from this Study

zone	AUROC	pValue	adjusted_P	laminar_order
subplate zone	0.536	4.68e-08	2.81e-07	4
intermediate zone	0.529	1.58e-05	7.90e-05	3
cortical plate	0.528	2.08e-05	8.32e-05	5
marginal zone	0.500	9.71e-01	9.71e-01	6
subventricular zone	0.493	3.24e-01	6.48e-01	2
ventricular zone	0.484	1.32e-02	3.96e-02	1
subpial granular zone	0.454	3.79e-12	2.65e-11	7



3.4.4 Cell-types Expression in ASD and ID

Gene expression was assessed for ID and ASD genes identified in this study and in the literature to determine if these genes were expressed at a higher level in a specific cell type. Cell-type expression for genes found in this cohort had an excitatory, inhibitory and serotonergic neurons in the midbrain, hindbrain, thalamus and cerebral cortex, as well as in granular interneurons in the cerebellum (Table 3.9). This was a common theme in the other gene list, as ASD cell-types showed increased expression in the excitatory and inhibitory neurons in the hindbrain, midbrain cerebral cortex with the addition of the hippocampus and thalamus (Table 3.10). The ID gene list was a larger list but also depicted enhanced expression in cell-types related to the excitatory and inhibitory neurons of the hindbrain, midbrain, cerebral cortex with thalamus and hippocampus similar to the ASD cell-type findings (Table 3.10). The ID genes also represented afferent nuclei of the cranial nerves and cholinergic neurons of the hindbrain (*These tables are not full tables but the top hits that represent the most highly expressed cell types. Cluster_ID is

the cell type cluster obtained from an iterative approach for cluster identification from Zeisel et al. (2018). The clustering was an iterative approach involving projected PCA cell types from gene expression, then manual annotation and then a Support Vector Machine classification. The adjusted_P is the Benjamini-Hochberg adjusted p-value

Table 3.11*These tables are not full tables but the top hits that represent the most highly expressed cell types. Cluster_ID is the cell type cluster obtained from an iterative approach for cluster identification from Zeisel et al. (2018). The clustering was an iterative approach involving projected PCA cell types from gene expression, then manual annotation and then a Support Vector Machine classification. The adjusted_P is the Benjamini-Hochberg adjusted p-value

Table 3.11: Polygenic Cell-types of ID Genes). Together, a common theme of excitatory and inhibitory neurons in the hindbrain and midbrain seem common place with some differences associated with excitatory and inhibitory neurons of the thalamus, cerebral cortex and hippocampus. These brain regions are somewhat expected as these regions control memory, learning and emotion and the additional number of genes may result in expression being mapped for these regions where as our cohort had much fewer genes in the gene list.

Table 3.9: Cell-type Expression in Genes (n=148) Identified in this Pakistani and Iranian Trio Families

cluster_id	description	auc	pValue	adjusted_P
MEINH5	Inhibitory neurons, midbrain	0.602	0.000291	0.0771
HBSER1	Serotonergic neurons, hindbrain	0.583	0.003070	0.8100
MEGLU1	Excitatory neurons, midbrain	0.581	0.003820	1.0000
MEGLU7	Excitatory neurons, midbrain	0.579	0.004750	1.0000
CBINH2	Granular layer interneurons, cerebellum	0.579	0.005030	1.0000
HBSER3	Serotonergic neurons, hindbrain	0.573	0.009260	1.0000
HBGLU1	Excitatory neurons, hindbrain	0.572	0.009750	1.0000
HBINH4	Inhibitory neurons, hindbrain	0.572	0.010300	1.0000
HBGLU10	Excitatory neurons, hindbrain	0.571	0.011000	1.0000
HBINH6	Inhibitory neurons, hindbrain	0.571	0.011400	1.0000
HBSER2	Serotonergic neurons, hindbrain	0.569	0.014000	1.0000
MEINH6	Inhibitory neurons, midbrain	0.568	0.014700	1.0000
HBINH2	Inhibitory neurons, hindbrain	0.568	0.015300	1.0000
MEGLU5	Excitatory neurons, midbrain	0.566	0.018600	1.0000
DEGLU4	Excitatory neurons, thalamus	0.565	0.021200	1.0000
MEGLU8	Excitatory neurons, midbrain	0.564	0.022100	1.0000
MEINH2	Inhibitory neurons, midbrain	0.564	0.022100	1.0000
MEINH3	Inhibitory neurons, midbrain	0.564	0.022400	1.0000
HBINH7	Inhibitory neurons, hindbrain	0.564	0.022400	1.0000
MEINH7	Inhibitory neurons, midbrain	0.564	0.022700	1.0000
TEGLU6	Excitatory neurons, cerebral cortex	0.562	0.026800	1.0000
DEGLU1	Excitatory neurons, thalamus	0.561	0.028300	1.0000
HBGLU4	Excitatory neurons, hindbrain	0.561	0.028800	1.0000

*These tables are not full tables but the top hits that represent the most highly expressed cell types. Cluster_ID is the cell type cluster obtained from an iterative approach for cluster identification from Zeisel et al. (2018). The clustering was an iterative approach involving projected PCA cell types from gene expression, then manual annotation and then a Support Vector Machine classification. The adjusted_P is the Benjamini-Hochberg adjusted p-value

Table 3.10: Polygenic Cell-types in ASD Genes (n = 1068) from Literature

cluster_id	description	auc	pValue	adjusted_P
HBGLU2	Excitatory neurons, hindbrain	0.620	1.78e-30	4.72e-28
HBINH5	Inhibitory neurons, hindbrain	0.613	2.30e-27	6.05e-25
HBGLU3	Excitatory neurons, hindbrain	0.607	1.59e-24	4.12e-22
HBINH9	Inhibitory neurons, hindbrain	0.606	2.45e-24	6.32e-22
TEGLU4	Excitatory neurons, cerebral cortex	0.601	4.45e-22	1.12e-19
TEGLU21	Excitatory neurons, hippocampus CA1	0.596	5.16e-20	1.28e-17
TEGLU17	Excitatory neurons, cerebral cortex	0.594	2.35e-19	5.83e-17
HBADR	Adrenergic cell groups of the medulla	0.593	5.78e-19	1.43e-16
TEGLU10	Excitatory neurons, cerebral cortex	0.593	6.70e-19	1.64e-16
MEGLU6	Excitatory neurons, midbrain	0.589	2.12e-17	5.07e-15
TEGLU22	Excitatory neurons, amygdala	0.587	8.41e-17	1.98e-14
HBGLU9	Excitatory neurons, hindbrain	0.587	1.06e-16	2.48e-14
DEGLU3	Excitatory neurons, thalamus	0.586	1.45e-16	3.38e-14
TEGLU13	Excitatory neurons, cerebral cortex	0.586	1.95e-16	4.50e-14
HBINH3	Inhibitory neurons, hindbrain	0.584	9.07e-16	2.05e-13
HBGLU7	Excitatory neurons, hindbrain	0.583	1.48e-15	3.32e-13
HBINH4	Inhibitory neurons, hindbrain	0.583	2.07e-15	4.62e-13
TEGLU20	Excitatory neurons, cerebral cortex	0.582	3.15e-15	6.96e-13
TEGLU14	Excitatory neurons, cerebral cortex	0.582	3.23e-15	7.11e-13
TEGLU16	Excitatory neurons, cerebral cortex	0.582	4.14e-15	9.07e-13
HBSER2	Serotonergic neurons, hindbrain	0.580	1.99e-14	4.26e-12
HBINH2	Inhibitory neurons, hindbrain	0.579	3.40e-14	7.17e-12
DEGLU5	Excitatory neurons, midbrain	0.579	4.61e-14	9.68e-12

*These tables are not full tables but the top hits that represent the most highly expressed cell types. Cluster_ID is the cell type cluster obtained from an iterative approach for cluster identification from Zeisel et al. (2018). The clustering was an iterative approach involving projected PCA cell types from gene expression, then manual annotation and then a Support Vector Machine classification. The adjusted_P is the Benjamini-Hochberg adjusted p-value

Table 3.11: Polygenic Cell-types of ID Genes (n = 1346) from Literature

cluster_id	description	auc	pValue	adjusted_P
HBGLU6	Excitatory neurons, hindbrain	0.613	1.74e-31	4.35e-29
HBINH5	Inhibitory neurons, hindbrain	0.613	1.84e-31	4.58e-29
MEGLU6	Excitatory neurons, midbrain	0.610	5.67e-30	1.39e-27
TEGLU4	Excitatory neurons, cerebral cortex	0.610	9.05e-30	2.21e-27
HBGLU7	Excitatory neurons, hindbrain	0.604	4.63e-27	1.10e-24
HBCHO2	Cholinergic neurons, hindbrain	0.604	5.26e-27	1.24e-24
HBCHO3	Afferent nuclei of cranial nerves VI-XII	0.604	5.39e-27	1.27e-24
DEGLU1	Excitatory neurons, thalamus	0.604	7.57e-27	1.76e-24
HBINH2	Inhibitory neurons, hindbrain	0.603	1.27e-26	2.93e-24
HBGLU2	Excitatory neurons, hindbrain	0.602	3.84e-26	8.76e-24
HBGLU1	Excitatory neurons, hindbrain	0.601	2.34e-25	5.29e-23
HBSER5	Serotonergic neurons, hindbrain	0.600	3.00e-25	6.69e-23
HBSER2	Serotonergic neurons, hindbrain	0.600	3.51e-25	7.79e-23
TEGLU10	Excitatory neurons, cerebral cortex	0.599	1.66e-24	3.64e-22
TEGLU15	Excitatory neurons, cerebral cortex	0.593	6.80e-22	1.43e-19
TEINH12	Non-border Cck interneurons, cortex/hippocampus	0.593	7.51e-22	1.58e-19
TEGLU23	Excitatory neurons, hippocampus CA3	0.592	1.85e-21	3.83e-19
TEGLU18	Excitatory neurons, cerebral cortex	0.591	2.97e-21	6.12e-19
HBINH6	Inhibitory neurons, hindbrain	0.591	3.00e-21	6.15e-19
HBGLU9	Excitatory neurons, hindbrain	0.591	4.15e-21	8.47e-19
TEGLU14	Excitatory neurons, cerebral cortex	0.591	6.54e-21	1.33e-18
HBCHO1	Cholinergic neurons, hindbrain	0.590	9.56e-21	1.92e-18
TEINH11	R-LM border Cck interneurons,	0.590	1.37e-20	2.74e-18

*These tables are not full tables but the top hits that represent the most highly expressed cell types. Cluster_ID is the cell type cluster obtained from an iterative approach for cluster identification from Zeisel et al. (2018)The clustering was an iterative approach involving projected PCA cell types from gene expression, then manual annotation and then a Support Vector Machine classification. The adjusted_P is the Benjamini-Hochberg adjusted p-value

Table 3.12: Polygenic Cell-types of ASD, ID and Pakistani and Iranian Trio Families Composite Genes (n = 2227)

cluster_id	description	auc	pValue	adjusted_P
HBINH5	Inhibitory neurons, hindbrain	0.605	3.63e-42	9.62e-40
HBGLU2	Excitatory neurons, hindbrain	0.599	1.10e-37	2.88e-35
TEGLU4	Excitatory neurons, cerebral cortex	0.592	5.60e-33	1.44e-30
MEGLU6	Excitatory neurons, midbrain	0.589	5.72e-31	1.46e-28
HBGLU7	Excitatory neurons, hindbrain	0.588	7.77e-30	1.94e-27
HBGLU6	Excitatory neurons, hindbrain	0.587	1.44e-29	3.59e-27
HBINH2	Inhibitory neurons, hindbrain	0.586	1.31e-28	3.20e-26
HBCHO3	Afferent nuclei of cranial nerves VI-XII	0.586	1.67e-28	4.06e-26
HBSER2	Serotonergic neurons, hindbrain	0.584	8.76e-28	2.12e-25
HBGLU9	Excitatory neurons, hindbrain	0.584	1.55e-27	3.74e-25
HBCHO2	Cholinergic neurons, hindbrain	0.583	4.98e-27	1.19e-24
TEGLU10	Excitatory neurons, cerebral cortex	0.582	5.50e-26	1.28e-23
HBSER5	Serotonergic neurons, hindbrain	0.581	6.12e-26	1.42e-23
DEGLU1	Excitatory neurons, thalamus	0.581	7.97e-26	1.84e-23
DEGLU3	Excitatory neurons, thalamus	0.578	7.41e-24	1.65e-21
HBGLU1	Excitatory neurons, hindbrain	0.578	9.26e-24	2.05e-21
MEINH2	Inhibitory neurons, midbrain	0.576	5.75e-23	1.25e-20
TEGLU14	Excitatory neurons, cerebral cortex	0.575	2.70e-22	5.72e-20
TEGLU21	Excitatory neurons, hippocampus CA1	0.575	2.90e-22	6.12e-20
HBINH6	Inhibitory neurons, hindbrain	0.575	4.43e-22	9.26e-20
HBINH9	Inhibitory neurons, hindbrain	0.574	1.41e-21	2.92e-19
HBADR	Adrenergic cell groups of the medulla	0.573	3.22e-21	6.63e-19
HBCHO1	Cholinergic neurons, hindbrain	0.573	4.31e-21	8.84e-19

*These tables are not full tables but the top hits that represent the most highly expressed cell types. Cluster_ID is the cell type cluster obtained from an iterative approach for cluster identification from Zeisel et al. (2018). The clustering was an iterative approach involving projected PCA cell types from gene expression, then manual annotation and then a Support Vector Machine classification. The adjusted_P is the Benjamini-Hochberg adjusted p-value

3.4.5 ASD Anatomical Mapping

The Allen Brain Atlas was also interrogated to determine which regions of the brain are enriched for the expression in our dataset, and in comparison, with other datasets for ASD and ID genes. The analysis revealed that several regions are significantly enriched in the cohort in this study such at the medial subdivision of the central nucleus, lateral septal nucleus, CP in the caudal hippocampus, the agranular insular cortex, the lateral subdivision of BNST and the deep developmental zones of the caudal entorhinal cortex.

Many of the regions that showed significance were associated with adaptive learning, social cognition, memory and processing/integrating information from different regions of the brain. The medial subdivision of the central nucleus is the major output pathway of the amygdala, which is responsible for processing emotion and memories, and connects to the brainstem for innate physiological responses to emotion (A. Wright, 1997). The base of the Nucleus of Stria Terminalis is particularly interesting, as it has a very prominent sexually dimorphism. The Bed of the Stria Terminalis is twice the size in males as in females and contains twice the number of somatostatin neurons (Swaab, 2007). It functions as a major output center for the amygdala. The agranular insular cortex is involved in memory and learning and mediates, long term retention, appetitive, aversive and novelty-driven learning and has been identified as a location of hypoactivity and dysfunctional connectivity in ASD (Gogolla, 2017). This suggests that the insular cortex is involved in multisensory and affective processing (Gogolla, 2017). The entorhinal cortex was also present in all gene sets and acts like a hub for a constructing a network in the brain for the processing of memory, time and navigation, memory formation and consolidation (Gogolla, 2017; Tsao et al., 2018). Like the insular cortex, many of the regions seem to be involved in the integrating and processing of information that is being affected with the genes discovered causing dysfunction and lower activity in these regions. The lack of integration and processing of information is a hallmark of ASD and many of the social deficits found in ASD ((Thye, Bednarz, Herringshaw, Sartin, & Kana, 2018)). Together, the regions often overlap with the striatum as a major component of ASD brain region expression and can explain a number of symptoms including social deficits, intellectual disability, repetitive behaviors and difficult processing information.

Table 3.13 Multigene Expression Mapping in the Human Brain for Genes Identified Pakistani and Iranian Trio Families (n = 148)

Name	Donors	AUC	P-Value
medial subdivision of central nucleus	1	0.58	0.20
lateral septal nucleus	4	0.56	0.43
mediodorsal nucleus of thalamus	4	0.56	0.43
medioventral part of putamen	1	0.56	0.43
magnocellular (medial) division of MD	2	0.56	0.43
CP in caudal hippocampus	1	0.56	0.45
anterior hypothalamic nucleus	3	0.55	0.59
posterior hypothalamic nucleus	2	0.55	0.59
agranular insular cortex (area Iag)	2	0.55	0.59
layer II of caudal entorhinal cortex	2	0.55	0.59
multiform (lateral) division of MD	2	0.55	0.59
dorsal subdivision of VLC	1	0.55	0.59

*Donors represent human donors from the BrainSpan project who donated their brains to research

Table 3.14 Multigene Expression Mapping of ASD Genes (n = 1068) Identified in Literature

Name	Donors	AUC	P-Value
agranular insular cortex (area Iag)	2	0.59	$p < 0.00001$
IZ in dysgranular insular cortex	4	0.58	$p < 0.00001$
lateral subdivision of BNST	1	0.58	$p < 0.00001$
medioventral part of putamen	1	0.58	$p < 0.00001$
deep developmental zones of caudal entorhinal cortex	1	0.58	$p < 0.00001$
reuniens nucleus (medioventral nucleus) of thalamus	1	0.57	$p < 0.00001$
SP in temporal polar cortex	1	0.57	$p < 0.00001$
SP in medial temporal-occipital cortex	2	0.57	$p < 0.00001$
layer V of caudal entorhinal cortex	1	0.57	p <

			0.00001
pyramidal layer of caudal subiculum	2	0.57	$p < 0.00001$
basomedial nucleus (accessory basal nucleus)	4	0.57	$p < 0.00001$
SP in caudal midinferior temporal cortex (area TF)	2	0.57	$p < 0.00001$

*Donors represent human donors from the BrainSpan project who donated their brains to research

Table 3.15 Multigene Expression Mapping for ID Genes (n = 1346) Identified in Literature

Name	Donors	AUC	P-Value
medioventral part of putamen	1	0.59	p < 0.00001
deep developmental zones of caudal entorhinal cortex	1	0.59	p < 0.00001
CP in caudal hippocampus	1	0.58	p < 0.00001
lateral subdivision of BNST	1	0.58	p < 0.00001
outer CP in superolateral temporal cortex	4	0.57	p < 0.00001
body of caudate	4	0.57	p < 0.00001
outer CP in primary auditory cortex	3	0.57	p < 0.00001
lateral subdivision of area 9	1	0.57	p < 0.00001
inner CP in posterior frontal cortex (motor cortex)	4	0.57	p < 0.00001
intermediate portion of DFC (area 9/46)	1	0.57	p < 0.00001
head and face region of M1	1	0.57	p < 0.00001
indusium griseum	1	0.57	p < 0.00001

*Donors represent human donors from the BrainSpan project who donated their brains to research

Table 3.16 Multigene Expression Mapping for ID and ASD Genes Identified in Literature and Pakistani and Iranian Trio Families (n = 2227)

Name	Donors	AUC	P-Value
medioventral part of putamen	1	0.59	p < 0.00001
deep developmental zones of caudal entorhinal cortex	1	0.59	p < 0.00001

CP in caudal hippocampus	1	0.58	$p < 0.00001$
lateral subdivision of BNST	1	0.58	$p < 0.00001$
outer CP in superolateral temporal cortex	4	0.57	$p < 0.00001$
body of caudate	4	0.57	$p < 0.00001$
outer CP in primary auditory cortex	3	0.57	$p < 0.00001$
lateral subdivision of area 9	1	0.57	$p < 0.00001$
inner CP in posterior frontal cortex (motor cortex)	4	0.57	$p < 0.00001$
intermediate portion of DFC (area 9/46)	1	0.57	$p < 0.00001$
head and face region of M1	1	0.57	$p < 0.00001$
indusium griseum	1	0.57	$p < 0.00001$

*Donors represent human donors from the BrainSpan project who donated their brains to research

3.4.6 Spatiotemporal Expression in PsychEncode and BrainSpan shows Striatum is Important in Developmental Trajectories

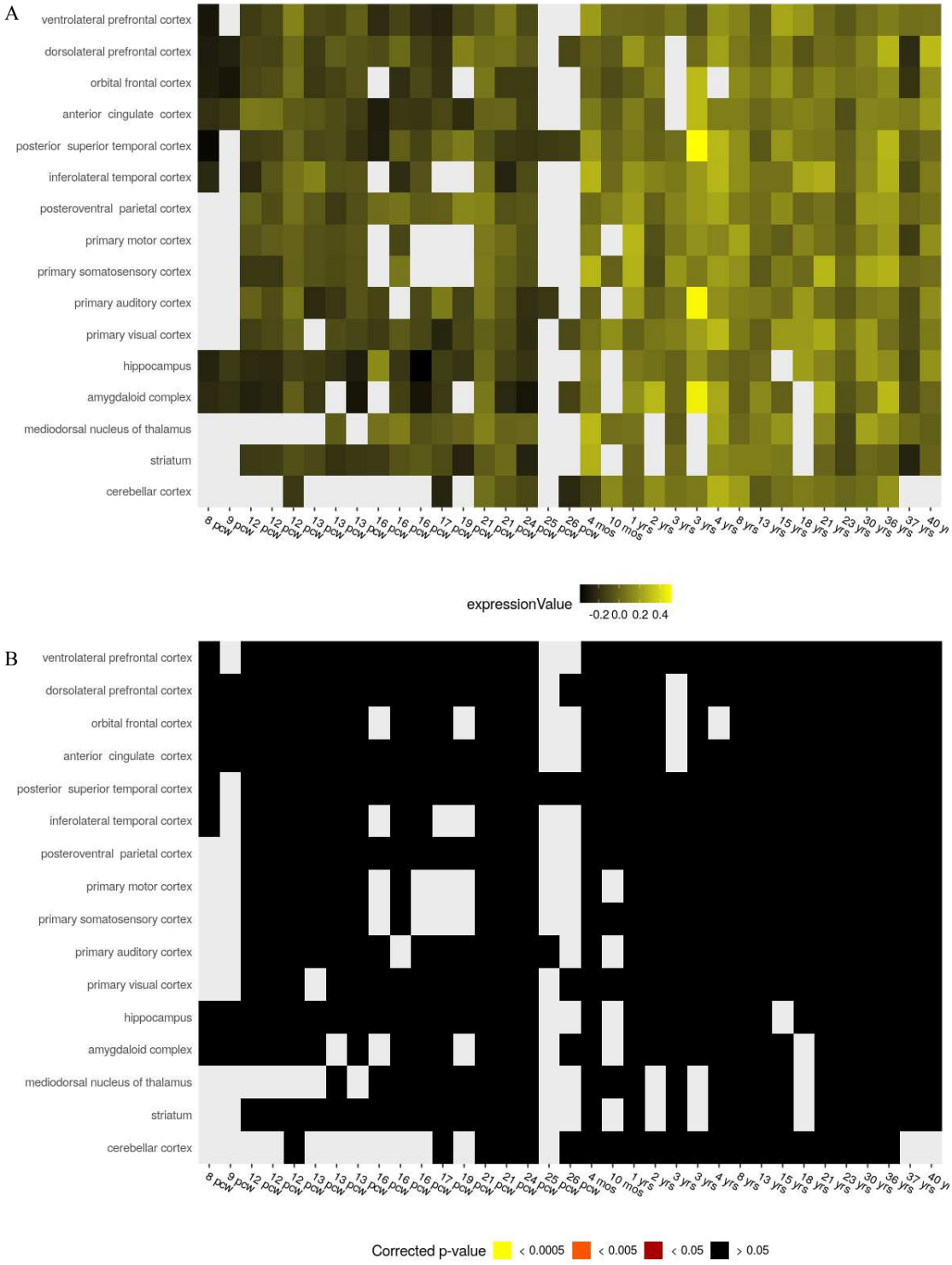


Figure 3.3: Spatiotemporal Expression of all genes identified in this cohort (n = 148) using the BrainSpan Atlas of the Developing Human Brain microarray data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

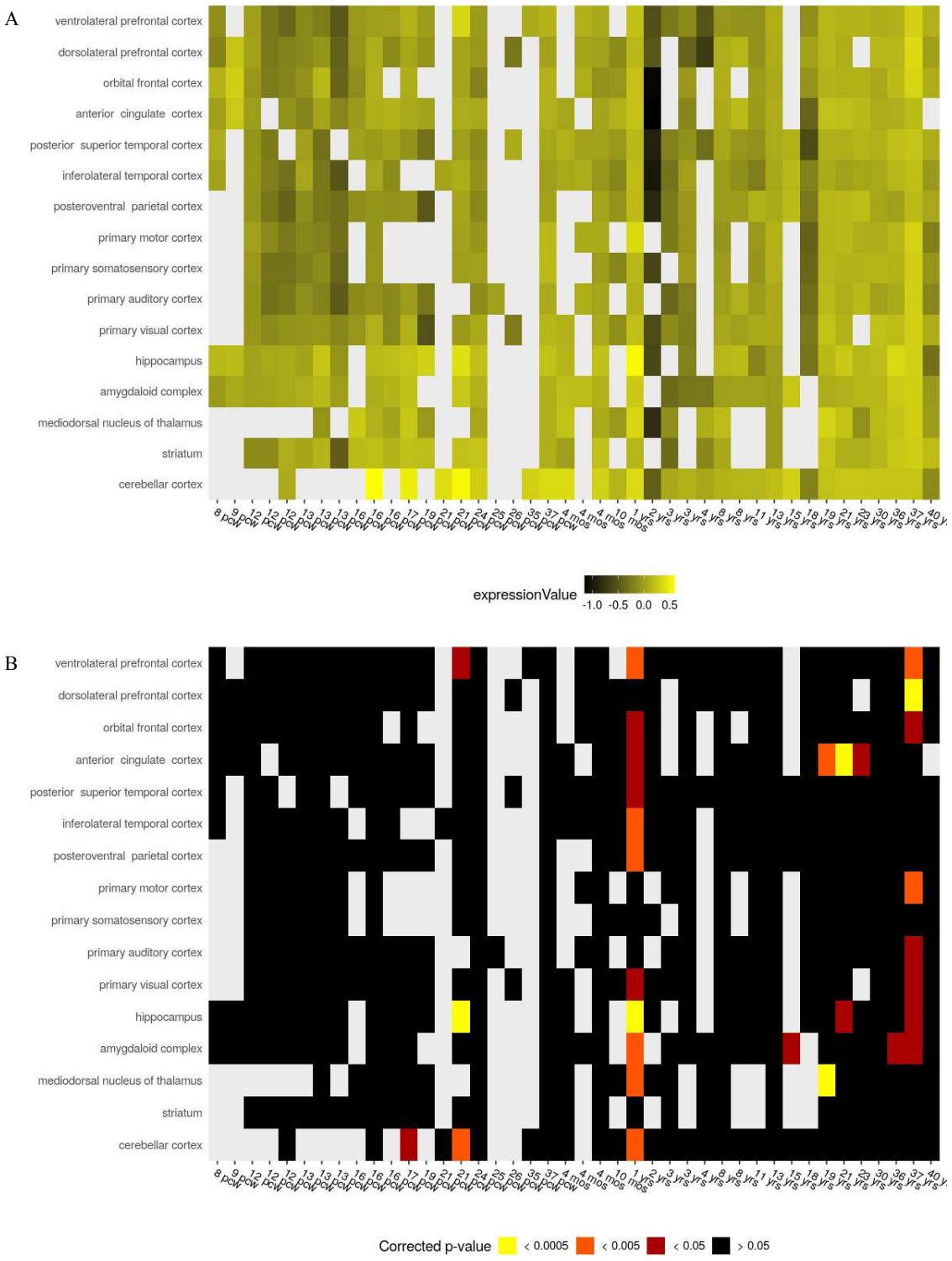


Figure 3.4: Spatiotemporal Expression of all genes (n = 148) identified in this cohort using the BrainSpan Atlas of the Developing Human Brain RNAseq A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

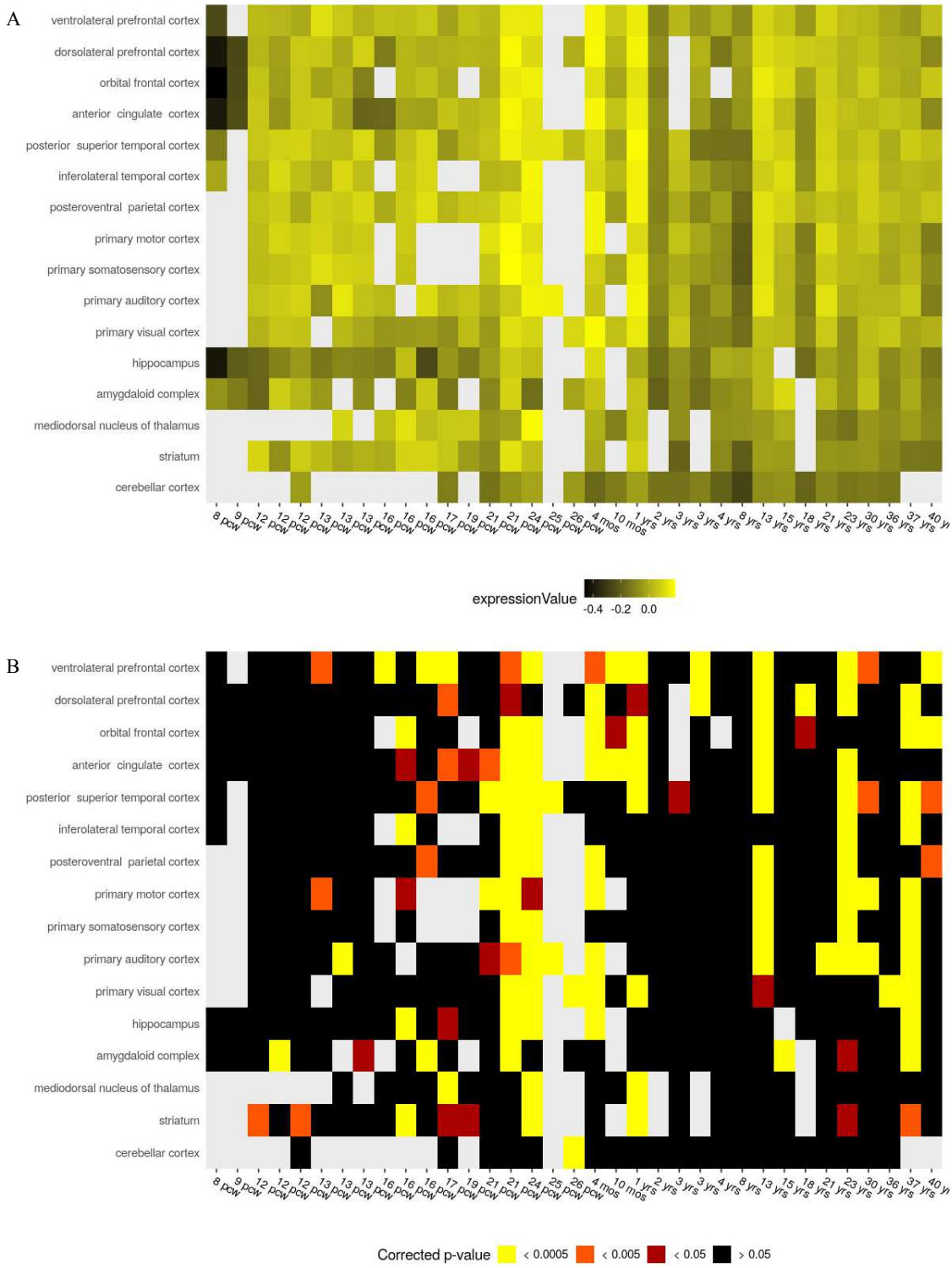


Figure 3.5: Spatiotemporal Expression of ASD genes (n = 1068) using the BrainSpan Atlas of the Developing Human Brain microarray data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

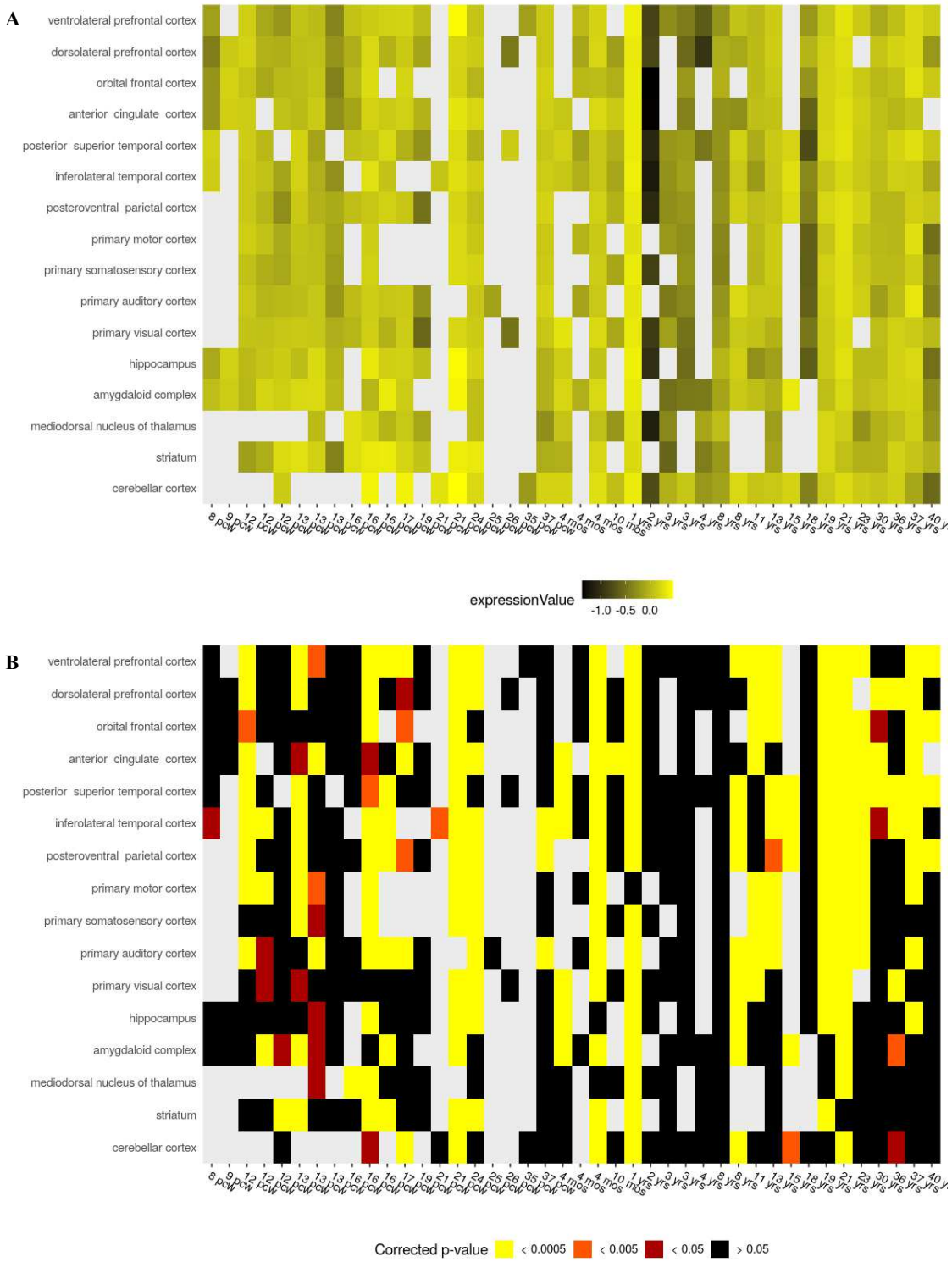


Figure 3.6: Spatiotemporal Expression of ASD genes (n = 1068) using the BrainSpan Atlas of the Developing Human Brain RNAseq data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

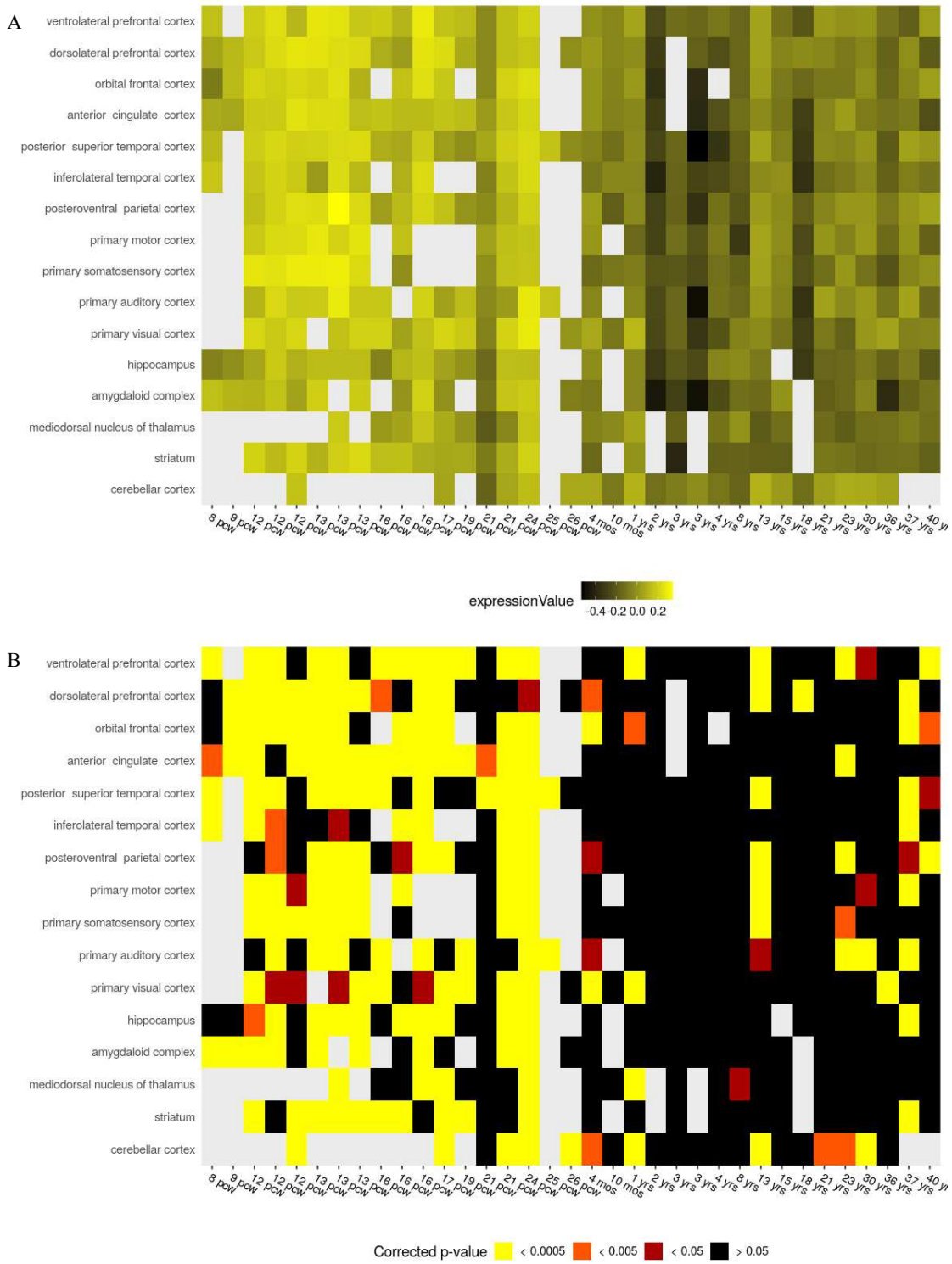


Figure 3.7: Spatiotemporal Expression of ID genes (n = 1346) using the BrainSpan Atlas of the Developing Human Brain microarray data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

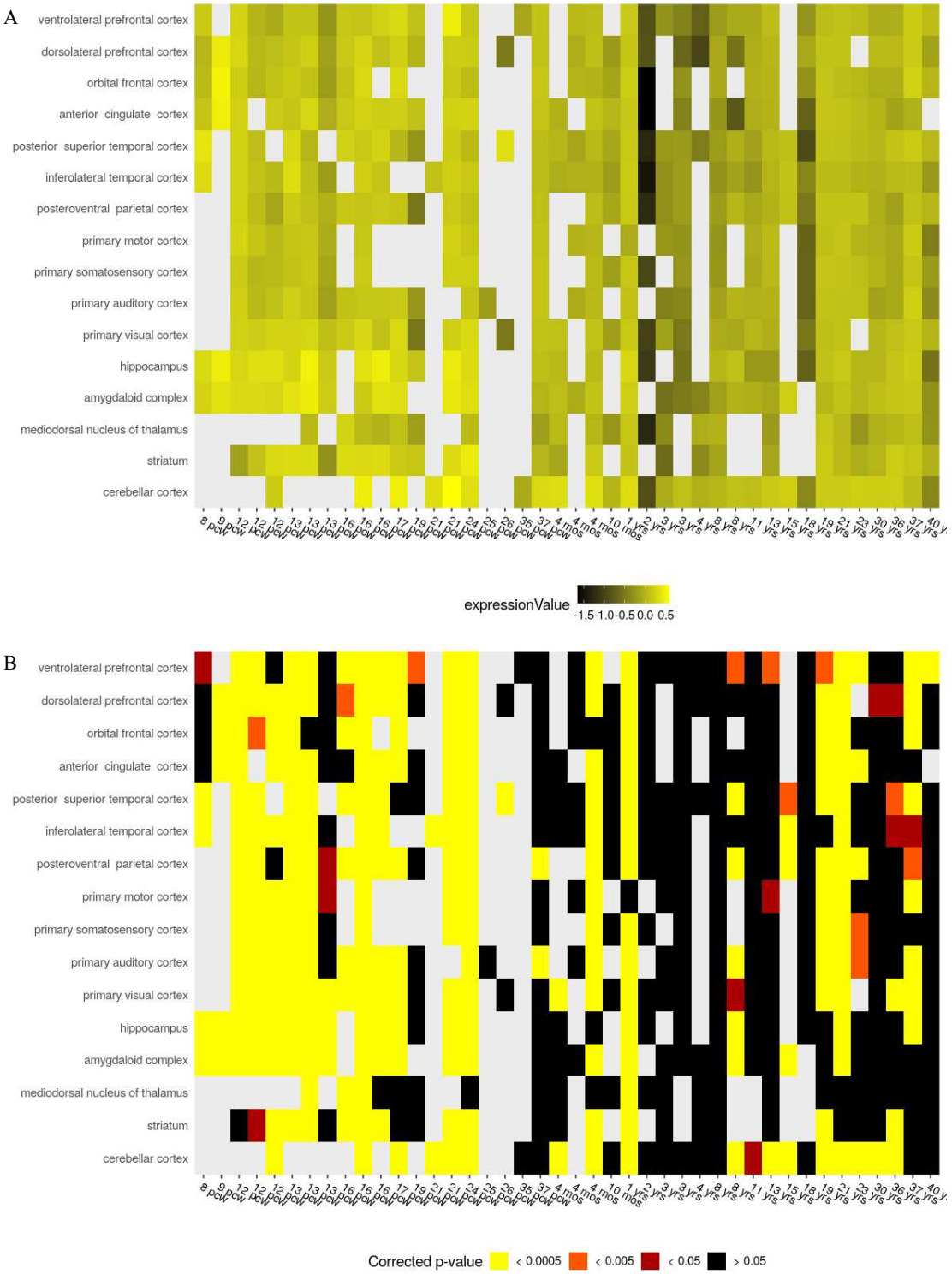


Figure 3.8: Spatiotemporal Expression of ID genes (n = 1346) using the BrainSpan Atlas of the Developing Human Brain RNAseq data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

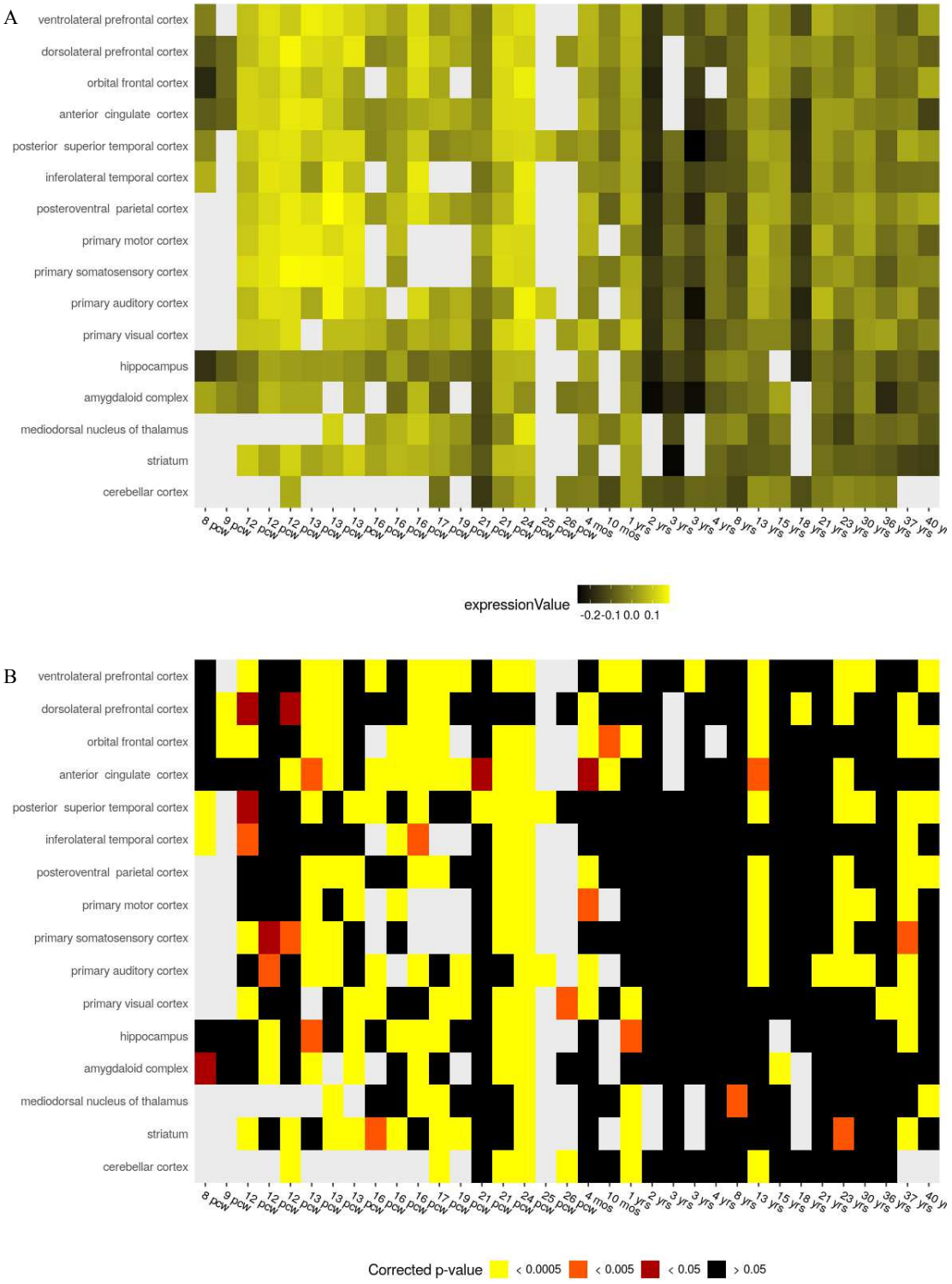


Figure 3.9: Spatiotemporal Expression of cohort, ID and ASD genes (n = 2227) using the BrainSpan Atlas of the Developing Human Brain microarray data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

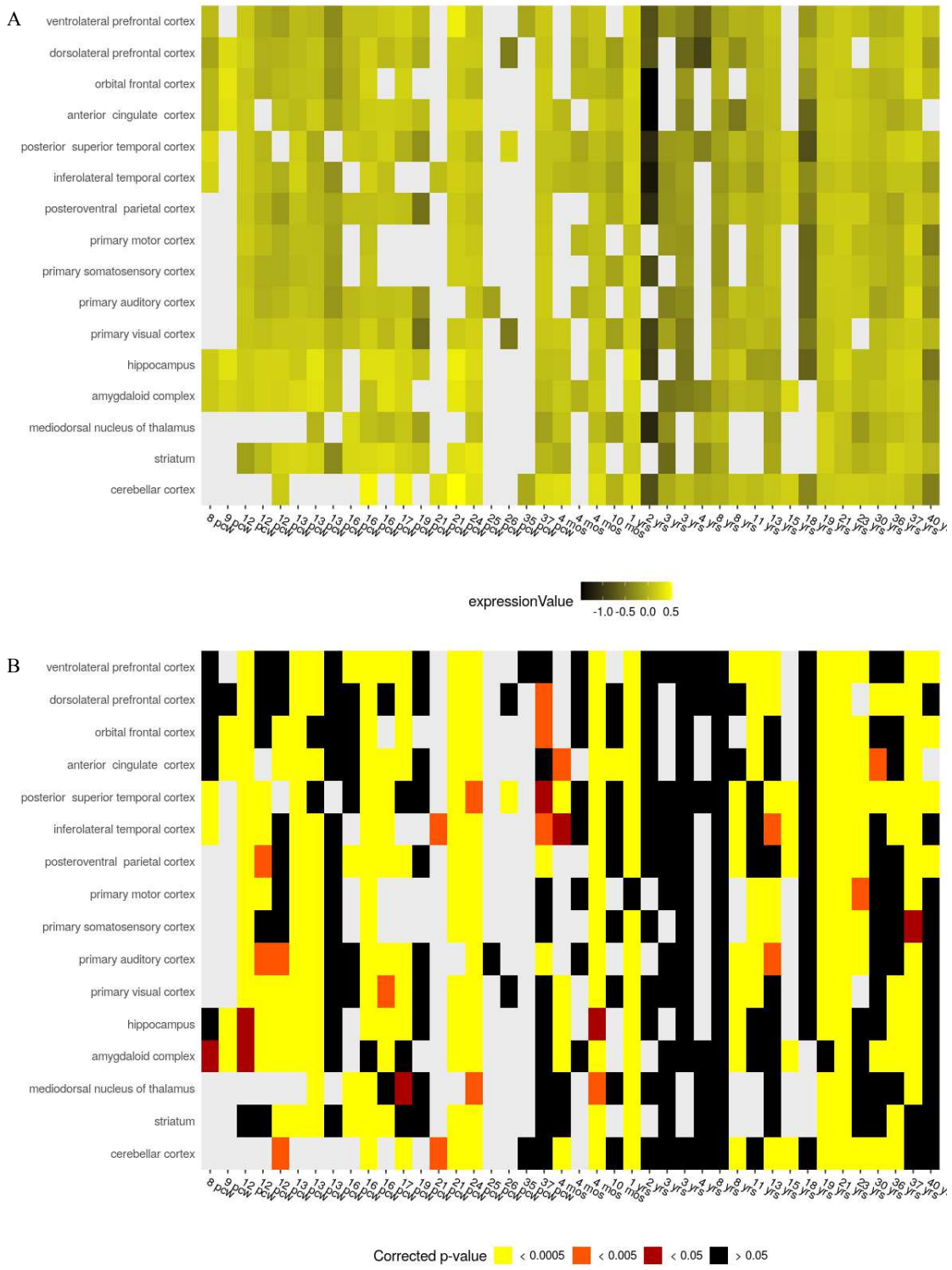


Figure 3.10: Spatiotemporal Expression of Cohort, ID and ASD genes (n = 2227) using the BrainSpan Atlas of the Developing Human Brain RNAseq data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

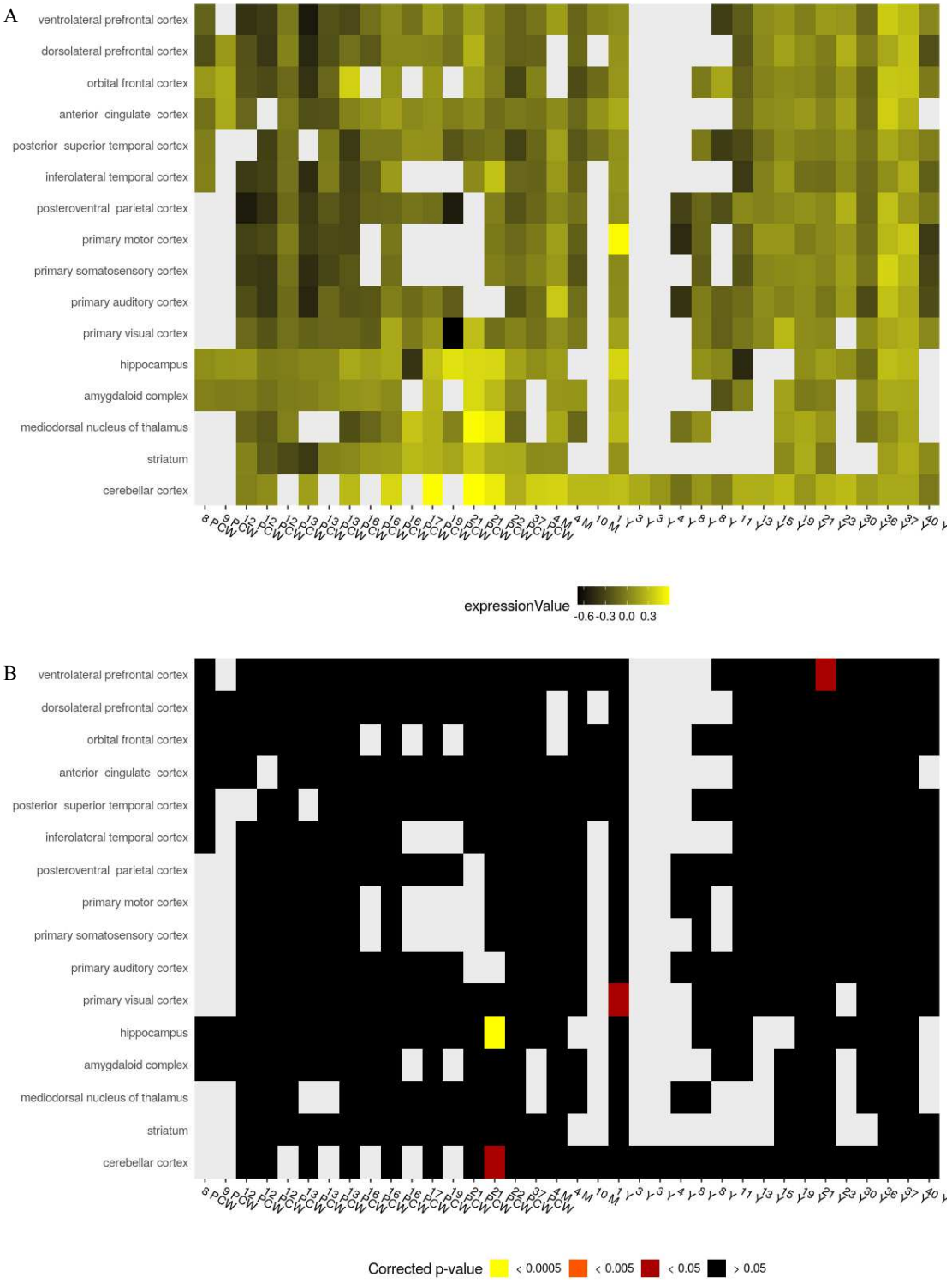


Figure 3.11: Spatiotemporal Expression of genes identified in this cohort (n = 148) using psychENCODE RNAseq data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

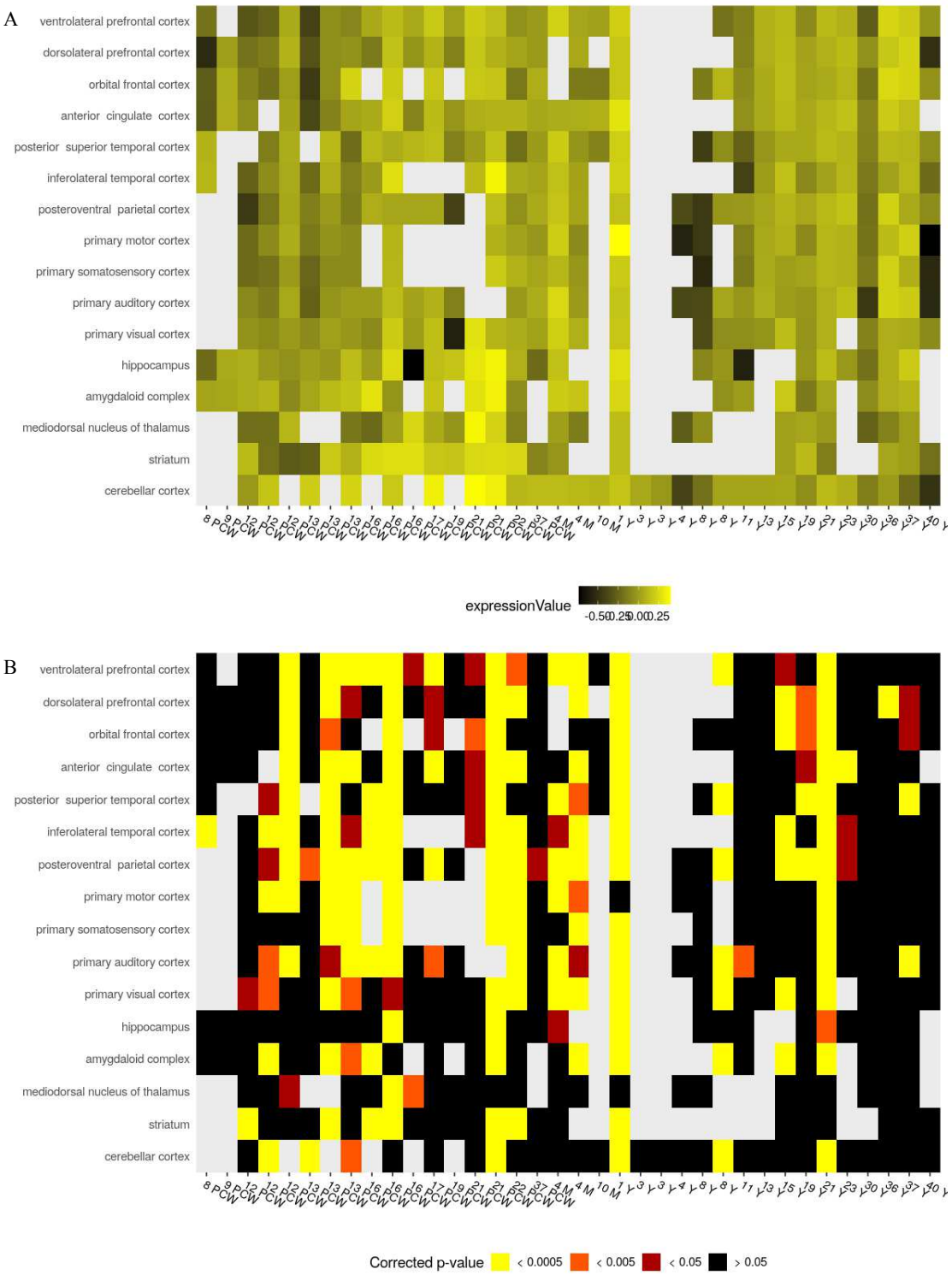


Figure 3.12: Spatiotemporal Expression of ASD genes (n =1068) using psychENCODE RNAseq data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

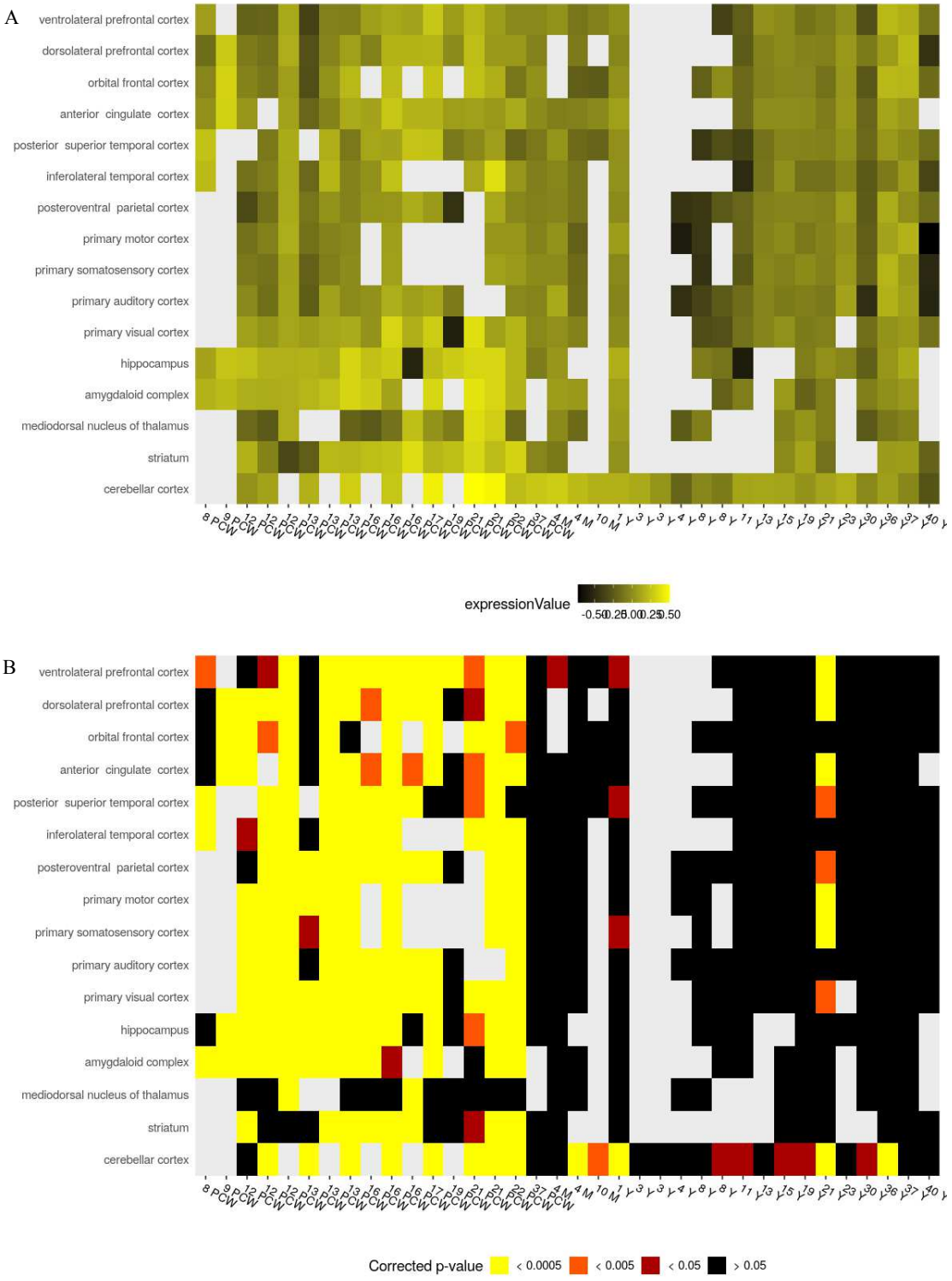


Figure 3.13: Spatiotemporal Expression of ID genes (n = 1346) using psychENCODE RNAseq data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

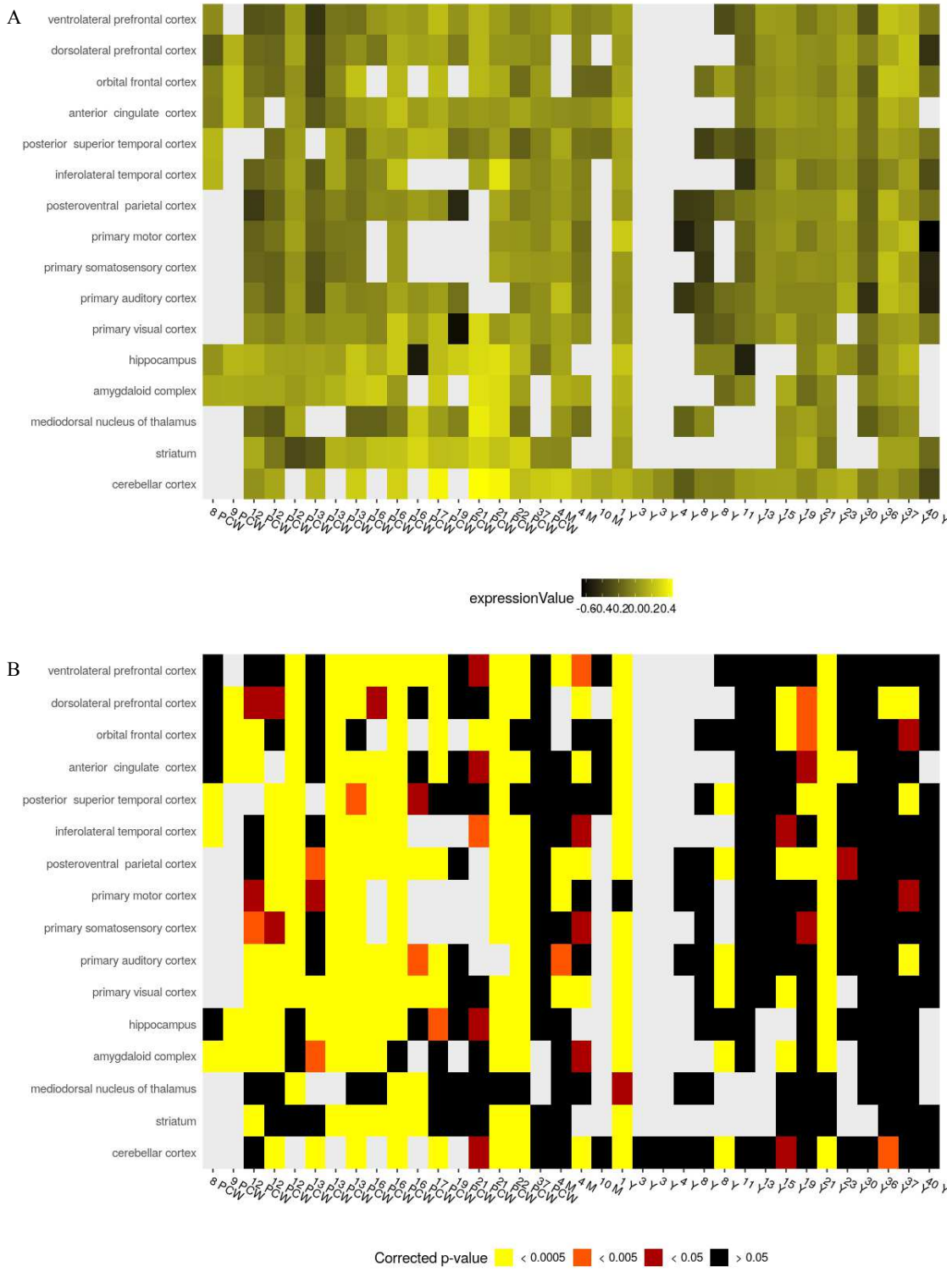


Figure 3.14: Spatiotemporal Expression of ID, ASD and gene identified in this cohort (n = 2227) using psychENCODE RNAseq data A) Expression Values as normalized RPKM heatmap and B) multiple testing corrected p-values heatmap

RNA-seq and microarray datasets were used to quantify the expression of gene set lists at different developmental timepoints and regions of the brain. RNA-seq and microarray datasets were used from PsychEncode and BrainSpan to delineate the relationship between the genes identified in this study and genes identified through SFARI, OMIM and literature searches for ASD and ID.

For the genes found in this study, the amygdaloid complex, hippocampus, striatum and cerebellar cortex showed significantly elevated expression from 16 to 24 weeks post conception compared to other regions of the brain, after multiple testing correction. Increased expression was observed for all genes identified in this cohort, ASD, ID genes and a composite expression of all genes identified in all lists and most strongly with the ID gene list across developmentally critical points before the age of 26 weeks post conception (Figure 3.3, Figure 3.4, Figure 3.5, Figure 3.6, Figure 3.7, Figure 3.8, Figure 3.9, Figure 3.10, Figure 3.11 and Figure 3.12) . The regions that have elevated expression were also seen in both the BrainSpan and PsychEncode datasets for both RNAseq and microarray expression experiments. The RNA-seq data seem to have more artifacts, likely related to the normalization, that differ from the microarray gene expression data, however, the trend of the cerebral cortex, amygdaloid complex, striatum, hippocampus and cerebellar cortex having high expression holds true (Figure 3.4, Figure 3.6, Figure 3.8 and Figure 3.10). The PsychEncode at some points does not have the same power as the BrainSpan data and this may be due to the number of the genes involved and the numbers of specimens in PsychEncode to power the analysis (Figure 3.11, Figure 3.12, Figure 3.13 and Figure 3.14). Taken together with brain region mapping and cell type mapping, the striatum showed elevated expression from the enrichment of genes that were identified in this cohort. The striatum is a a major intregration centre for information processing ((Reig & Silberberg, 2014)) and developmental changes been associated with ASD previously in the striatum ((Langen et al., 2014)).

3.4.7 Discussion

Many recent studies involving NGS in ASD have involved large cohorts, focusing predominantly on dominant/*de novo* inheritance. This is largely driven by the fact that autosomal recessive variants in novel genes are more difficult to identify in the outbred population, and informatic pipelines are not sufficiently tuned to identify compound

heterozygote variants. Compound heterozygous variants would need to have variants on opposite alleles, and in order to identify them would require informatic pipelines that focus on multiple damaging heterozygous variants in the same gene in a proband, and comparing with genotypes from the maternal and paternal WES or WGS data. This requires extra care when analyzing variants. The scientific method of using GWAS to identify variants is also not adequate to find recessive variants because the sample size required would be significantly larger than current Psychiatric Genomics Consortium studies in order to have sufficient power to find recessive or compound heterozygous variants. This is due to fact that most ASD genetics research emanates from North American and European countries, where larger families that allow for recessive variant mapping are not as readily available.

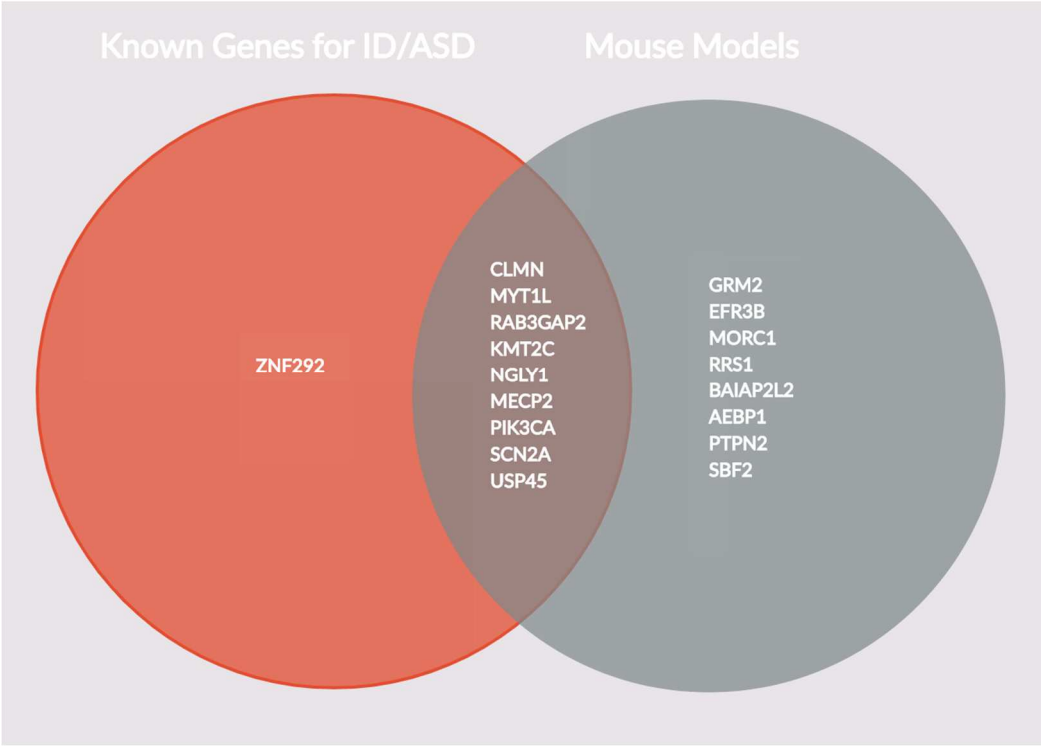


Figure 3.15 *De novo* Known ID/ASD Variants with Functional Models from Genes Identified in Pakistani and Iranian Trio Families

Using whole exome trio analysis in consanguineous families, we have enriched for recessive variants to assess the role these variants play in ASD in populations where endogamy is common. This study of 115 trios has identified 38 known genes associated with ASD or ID, resulting in a yield of 31.3 % compared to the other variants in the cohort Table 3.2: Percentage of Variants Identified by Inheritance Pattern for All Genes and Known Genes in this Study. In

our ASD cohort, 10 were *de novo* and 28 were recessive known ID/ASD variants. The diagnostic yield, which is being defined as previously associated genes for ID/ASD, of previously known *de novo* variants was ~ 6.7%, whereas the previously known recessive variants represent ~ 19 % of the genes in the cohort.

ZNF292 is the single known ID/ASD gene with no current mouse functional model identified through this study.

There were many other variants identified that met filtering criteria which could potentially represent novel ASD genes or targets. Cumulatively, candidate *de novo* variants were identified in 34 of the 115 trios (29.6 % of families) and 50 trios had candidate recessive variants identified in this cohort (43.5 %). In this cohort, there were genes that were associated with known inborn errors of metabolism (5.2 % of the diagnosis cohort). Genes such as *AGA, ASL, ASPA, HTRA2, MTHFR, NGLY1*, and *SLC36A1* may represent better clinical management opportunities for patients and potentially better therapies.

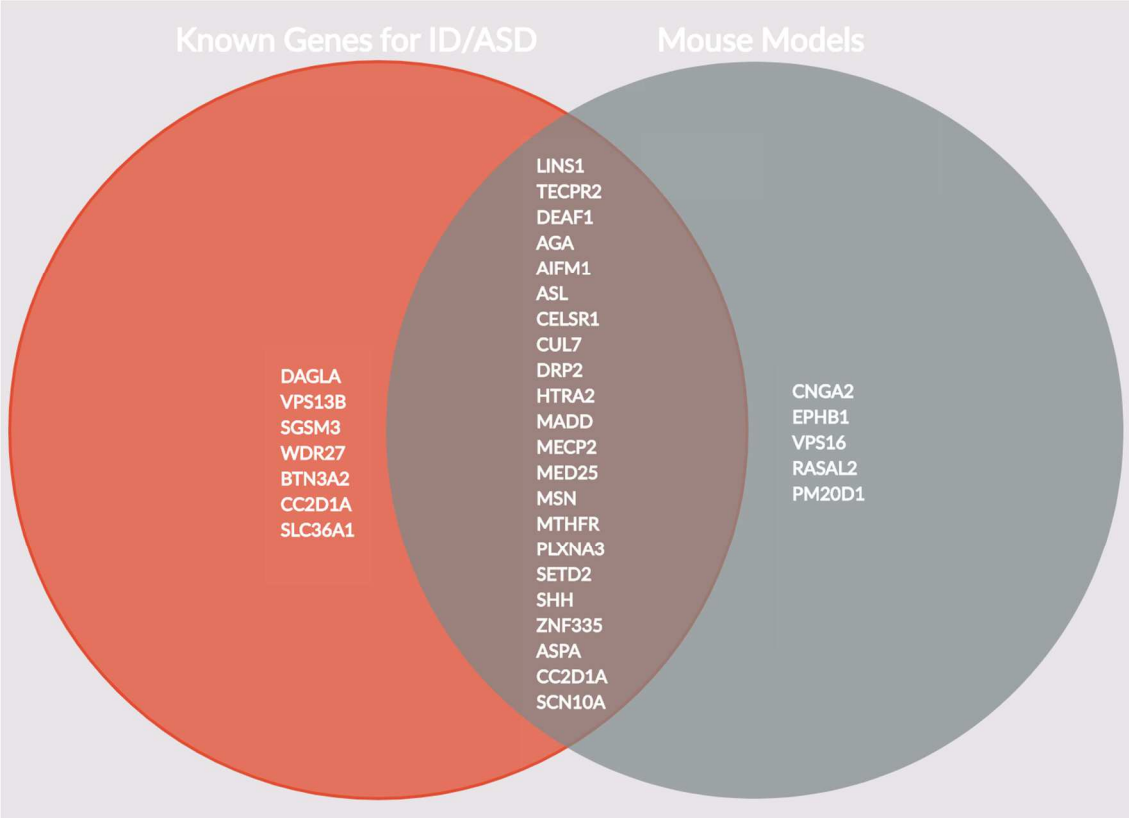


Figure 3.16 Known Recessive ID/ASD Variants with Functional Mouse Models from Genes Identified in Pakistani and Iranian Trio Families

Seven recessive variants in known ID/ASD genes, but without known mouse models have been identified in this study, however, an additional 22 known ID/ASD variants do overlap with a functional model. Additionally, five variants identified in this study have, to date, not been associated with ID/ASD but do have functional mouse models with relevant phenotypes (Figure 3.16).

Recently, it was estimated in the DDD cohort that recessive variants make up 3.6% of diagnoses, whereas, *de novo* variants would contribute 48.6% of variants. Martin et al., also looked at homozygous variants in a narrowly defined group of Pakistani Ancestry in the British Isles (PABI) which is most like our cohort (Martin et al., 2018). The PABI cohort had 333 undiagnosed probands from Pakistani origin and was used to estimate recessive variant contribution in consanguineous populations and only 5.2 % of cases would be explained by homozygous recessive variants (Martin et al., 2018).

Our study has reported a recessive variant diagnostic yield of 31.3 % of known recessive variants in contrast to the expected role of 5.2% from this previous study (Martin et al., 2018). This may be due to a lack of understanding about the degree of consanguinity and population substructure in different parts of the world. Our cohort differed from the DDD cohort by having a third of the study group from Iran which could result in different haplotype structures and variance in the cohort consanguinity from the PABI cohort. The cohort featured in this study is ~ 36 % Iranian where consanguinity is estimated to be lower for the Iranian population ((Saadat, Ansari-Lari, & Farhud, 2004);(Ullah, Husseni, & Mahmood, 2017)) . CNVs and structural variants are well known to be associated with ASD, however, due to sequencing depth and poor-quality DNA calling CNVs from NGS exome technology has proven to create many false positives (Samarakoon et al., 2016), (Sadedin, Ellis, Masters, & Oshlack, 2018)). In previously studies, a smaller subset of the families (n = 9) studied had disease causing CNVs that would explain ID or ASD (Harripaul et al., 2016; Harripaul et al., 2018).

Loss of function (LoF) variants were quite prominent in our cohort, with 81 of the 148 variants being LoF (54.7 % of variants identified). 16 LoF variants were recessive, biallelic variants (~11 % of variants identified), and 64 variants being *de novo* LoF variants (45 % of variants identified). The robustness of the LoF variants seen in our cohort exemplifies the role of highly

penetrant, disruptive LoF variants that play a major role in neurodevelopmental disorders and ASD in consanguineous families.

This cohort has identified mutations in both known and novel genes, with 38 genes being previously associated with ID or ASD spread across both autosomal recessive and *de novo*, and X-linked genetic models. Notably, genes *MYT1L* (Stevens et al., 2011), *ZNF335* (Hao Hu et al., 2019), *ZNF292* (Mirzaa et al., 2020), *MECP2* (Amir et al., 1999) and *SCN2A* (Weiss et al., 2003) have been previously implicated in neurodevelopmental disorder. However, it is difficult to categorize gene phenotypes as ASD and ID since there is a significant overlap between ASD and other comorbid disorders, particularly, ID. It has been reported that ID is present in ASD cases with estimates ranging from 40 to 70 % ((Mefford, Batshaw, & Hoffman, 2012)).

Reviewing literature and functional models in mice and zebrafish, 11 genes have relevant functional models in mice. Many of the candidate genes with functional models have deficits in the formation of the forebrain or nervous system. Examplex of *de novo* variants with functional models include genes *Efr3b , Morc1, Ptpn2, Rrs1, , Sbf2,* and. *Rrs1* LoF variants have abnormal behavioral characteristics and exhibit a hyperactivity phenotype (MGI:1929721). ASD has other comorbid phenotypes includinghyperactivity and attention deficit which has been linked to ASD in both genetic and brain structure studies (Aoki et al., 2017; Craig et al., 2016; Karalunas et al., 2018). *Efr3b* mouse models have decreased prepulse inhibition (MGI:2444851) which is a phenotype that has been observed and phenotypically reversed in other ASD mouse models (Teng et al., 2016). *Sbf2* mutant mice (MGI:1921831) have abnormal behavioral and neurological phenotypes such as abnormal limb grasping, impaired coordination and posture but more importantly, a plethora of nervous system phenotypes which include abnormal axon morphology, myelin sheath morphology, action potentials, nerve conduction and neuron degeneration.

In general, the recessive variants with functional models typically have more developmental and structural brain phenotypes. *Ephb1* (MGI:1096337) encodes an ephrin receptor. These receptors and their ligands are important for developmental processes, and most notably in the nervous system. Mouse models of *Ephb1* mutations are associated with abnormal axon guidance and morphology.

The main cortical developmental zones that were associated with the ASD genes from this cohort were the cortical plate, marginal zone, and subplate where the Cajal-Retzius cells transition into the marginal zone. It is believed that it is within this transition of cells that malformations arise (Folsom & Fatemi, 2013; Wasser & Herz, 2017). During development, as the cortical plate matures, connections are established between the cells in the subplate (Myakhar, Unichenko, & Kirischuk, 2011). Pioneer cells, such as the Cajal-Retzius, cells will create fibers and synapses within the cortical plate (Radnikow, Feldmeyer, & Lübke, 2002). The ventricular, subventricular and subplate zones were also associated with the ASD gene list (Radnikow et al., 2002). The neurons and white matter in the ventricular and subventricular zone migrate to the cortical plate and eventually becomes the white matter of the adult brain (Maki, Liang, Miyamoto, Lo, & Arai, 2013). The Cajal-Retzius cells are critically implicated in the migration of neurons in these cortical developmental zones (Frotscher, 1998). It has been postulated that the Cajal-Retzius cells are implicated in neurodevelopmental disorders such as schizophrenia (Bradshaw et al., 2017), bipolar disorder (Fatemi, Earle, & McMenomy, 2000), ASD (Cuchillo-Ibáñez, Andreo-Lillo, Pastor-Ferrándiz, Carratalá-Marco, & Sáez-Valero, 2020), lissencephaly (Wynshaw Boris, 2007) and temporal lobe epilepsy (Wynshaw Boris, 2007). Increased levels of Reelin (RELN), a key glycoprotein secreted by the Cajal-Retzius cells, have been observed in ASD patients (Cuchillo-Ibáñez et al., 2020).

Gene expression analysis was also performed on the gene lists for spatiotemporally developmental signals. The genes found in this cohort are significantly expressed in early development after conception and until birth. Expression of the ASD/ID gene from the constructed gene lists after birth is lower. This expression pattern consolidates the important role these genes play in development. Subregions of the brain were also tested, and the striatum consistently shows overexpression during development among the genes discovered in this cohort. The bulk of the connections of the striatum come from the cerebral cortex, which shows significant regions of the cortical plate and intermediate zone in the cortical layer mapping.

The validation and analysis of the trios is currently ongoing with Sanger sequencing for high quality, known variants. Currently, we see a 93% validation rate for recessive variants, with *de novo* variants pending. Fragile X Syndrome screen will be performed using PCR to test for Fragile X in the remaining Pakistani Autism Trio samples. Several different CNV calling algorithms were used to call CNVs in exome data but all resulted in a high false positive rate

with no validated CNVs. Due to the lack of comprehensive CNV calling all trio samples will be genotyped using Illumina Infinium CoreExome microarrays. The microarray data will provide a few benefits including reliable CNV calling and genotype calls that can be used to calculate F-coefficients for relatedness scores. The CNV analysis will also be bolstered by using the CLAMMS CNV algorithm that has been designed partition samples by similarity and use them as references for calling CNVs in exome data which has proven to yield more reliable CNV calls ((C. L. Martin et al., 2020),(Packer et al., 2016)). The microarray and exome CNV calling will add additional methods for detecting both large and small pathogenic CNVs.

Our study suggests the role of recessive variants may have previously been underestimated, as this study shows a ~ 19% diagnostic yield for known recessive variants in the consanguineous ASD cohort (Table 3.2). Furthermore, research in outbred populations provides clear support for autosomal recessive inheritance; biallelic LoF mutations have been estimated to contribute to 3% of ASD genetics in two US based case control studies (E. T. Lim et al., 2013), and such findings are not limited to population isolates or ethnic subgroups (Howrigan et al., 2016; Kaiser et al., 2015). The genes identified in our study will inform molecular diagnostics, increasing the clinical diagnostic yield. With further research, it is possible that these findings will lead to new therapeutic strategies and clinical management strategies, while also dissecting the biological complexities of the human brain such as with the *MeCP2* readthrough drugs which have recently had success in mouse models ((Merritt, Collins, Erickson, Dong, & Neul, 2020))

4 *De novo* and inherited variants in *ZNF292* underlie a neurodevelopmental disorder with features of autism spectrum disorder

This chapter is modified from the following published manuscript:

Mirzaa, G., Chong, J.X., Piton, A., Foss, K., Popp, B., Guo, H., **Harripaul, R.,** Xia, K., Scheck, J., Aldinger, K.A., Sajan, S., Rodan, L., Sunderland, W., Bernstein, J.A, Undiagnosed Disease Network, Martinez-Agosto, J., Vincent, J.B, Bamshad, M., The University of Washington Center for Mendelian Genomics (UW-CMG). (2019). *De novo* and inherited variants in ZNF292 underlie a neurodevelopmental disorder with Features of Autism Spectrum Disorder. Genetics in Medicine

*De novo*Below is the link to the published manuscript:

https://www.nature.com/articles/s41436-019-0693-9

Reprinted with permission from Nature Publishing Group

I was a key contrubutor in the early discussions, experimental design, and experimental validation efforts for the study. I contributed sample quality control, workflow improvements for DNA quality and extraction, data generation, data processing, data analysis, data interpretation, manuscript drafting, manuscript editing of text concerning our two trio families with the same variant. .

4.1 Abstract

Purpose: Intellectual disability (ID), developmental delay (DD) and autism spectrum disorder (ASD) are genetically highly heterogeneous neurodevelopmental (NDV) disorders, each with a complex genetic architecture that is rapidly being uncovered with the widespread application of

next generation sequencing methods including targeted, exome and whole genome sequencing. Herein, we sought to delineate the clinical, molecular and neuroimaging spectrum of a novel neurodevelopmental disorder caused by variants in the Zinc Finger Protein 292 (ZNF292).

Methods: We ascertained a cohort of 28 families with ID due to putatively pathogenic ZNF292 variants that were identified via targeted and exome sequencing. Clinical and molecular data on this cohort were shared through MatchMaker Exchange (MME) network nodes and by querying investigators with large cohorts of individuals with ID and/or ASD. All available phenotypic, neuroimaging and molecular data were analyzed to characterize the canonical phenotype and examine genotype-phenotype relationships.

Results: All probands with variants in ZNF292 presented with ID as well as a spectrum of neurodevelopmental features including ASD, Attention-Deficit/Hyperactivity Disorder (ADHD), among others. Several individuals had structural brain abnormalities and minor dysmorphic features. All ZNF292 variants were *de novo*, except for one family with a dominantly inherited variant from a clinically affected mother. ZNF292 encodes a highly conserved zinc finger protein that acts as a transcription factor and is highly expressed in the developing human brain supporting its' likely critical role in neurodevelopment.

Conclusion: *De novo* and dominantly inherited variants in ZNF292 are associated with a range of neurodevelopmental features including ID, ASD, ADHD. The clinical spectrum of individuals with ZNF292 variants is quite broad, and most individuals present with mild to moderate ID with or without other syndromic features. Our cohort suggests that variants in ZNF292 are likely a relatively common cause of a neurodevelopmental disorder manifesting as ID with or without ASD/ADHD

4.2 Introduction

Knowledge about the genetic architecture of intellectual disability (ID), developmental delay (DD) and autism spectrum disorder (ASD) has increased dramatically over the past decade with the wide application of exome and whole genome sequencing (ES/WGS) methods. As these genetic tools are becoming increasingly available in both the clinical diagnostic and research settings, a growing number of children with neurodevelopmental disorders (NDV) are now identified to have genetic variants that arise either *de novo*, or inherited as autosomal dominant,

X-linked, or, less commonly, autosomal recessive traits. Discovery of more than 400 genes underlying ID and / or ASD, to date, has markedly informed the diagnosis for families with ID/ASD and, furthermore, has also led to the identification and characterization of multiple cellular pathways involved in human brain development, behavior, learning, and memory.1-3 Such gene discovery efforts are important as the developmental roles of many of these pathways would not have otherwise been predicted from in vitro and model organism studies. In this study, we describe the clinical, neuroimaging and molecular features of 28 individuals with ID and / or ASD due to *de novo* or inherited variants in Zinc Finger Protein 292 (ZNF292; MIM: 616213).

ZNF292 encodes a highly conserved zinc finger protein that acts as a transcription factor. ZNF292 is composed of eight exons, the last of which is the largest and encodes all 16 highly conserved zinc fingers of the predicted 2723-residue protein (canonical transcript in GenBank: NM_015021.1). Three of these zinc fingers (10-12) bind DNA at the promoter of growth hormone where it cooperates with POU1F1, a member of the POU family of transcription factors known to activate transcription in somatotrophs.4 Accordingly, the ZNF292 protein was originally described as an enhancer of growth hormone (GH) expression in the pituitary gland of a rat animal model.4 Its' role was further delineated as a tumor suppressor with critical roles in tumor development and progression.5 However, the role of ZNF292 in neurodevelopment is virtually unknown.

4.3 Materials and Methods

Cohort ascertainment

We identified 28 families with *de novo* (N = 27) or inherited (N =1) pathogenic variants in ZNF292 using a combination of trio-based ES (20 families) and multi-gene panels (8 families), in both clinical diagnostic and research settings. Families were identified across twenty institutions in six countries with data shared via nodes of the MatchMaker Exchange (MME) network, including MyGene2, Gene Matcher, Phenome Central, and by querying investigators with large cohorts of patients with ID and / or ASD.6 We identified 12 additional families (15 affected persons total) with variants in ZNF292 that were considered likely pathogenic in a diagnostic setting, or "suspected" pathogenic in a research setting but for which our confidence in the pathogenicity of these variants was limited due to either incomplete parental testing or the identification of a missense ZNF292 variant of unclear significance. Therefore, we excluded

these families from the primary cohort to avoid confounding description of both the canonical phenotype and genotype-phenotype relationships (Supplemental Subjects and Methods). We collected and reviewed detailed clinical data including medical records, facial photographs, and magnetic resonance images (MRI) from all affected individuals (summarized in Table 4.1). The institutional review board of the University of Washington approved this study.

Table 4.1 Summary of the clinical features of ZNF292 mutation-positive individuals (N=24)

Feature	**Present**	**Absent**	**ND**
Neurodevelopmental features			
ID/DD	22	2	0
Speech delays	21	2	1
ASD	15	8	1
ADHD	6	16	2
Tone abnormalities	13	6	5
Brain MRI abnormalities	9	6	9
Epilepsy	4	17	2
Other features			
Dysmorphic features	14	8	2
Eye anomalies	11	11	2
Growth failure Weight/height <2 SD	9	12	3
Microcephaly OFC <2SD	4	20	0
Feeding issues	8	13	3

Constipation	2	18	4
Skeletal problems	4	16	4
Hypermobility	4	17	3
Cardiac problems	4	16	4
Abbreviations: ADHD, attention-deficit hyperactivity disorder; ASD, autism spectrum disorder; ID, intellectual disability; ND, no data; OFC, occipitofrontal circumference; SD, standard deviations.			

Molecular Methods

In the primary cohort, 20 families underwent exome sequencing (ES) in either a clinical or research setting, and eight families underwent targeted sequencing of multi-gene panels. Of the cohort who underwent targeted sequencing, five underwent targeted capture of a multi-gene panel for ID and three underwent testing via a single molecule molecular inversion probe (summit) based panel of more than 100 genes associated with ID/ASD.7 Exome sequencing methods are further provided in the

Whole exome sequencing methods for 17-014

Institutional Research Ethics Board approval was received through the Centre for Addiction and Mental Health (CAMH) and other institutional recruiting sites. All parents from Iranian cohorts filled out the Social Communication Questionnaire (SCQ) and children who met the specified cut-off were then assessed using the Autism Diagnostic Interview-Revised (ADI-R) and some with the Autism Diagnostic Observation Schedule (ADOS). Pakistani ASD trios were assessed using the Childhood Autism Rating Scale (CARS). The diagnosis of ASD was confirmed in all subjects using the ADI-R and ADOS and 3DI. **Whole Exome Sequencing.** WES was performed using the Thruplex DNA-Seq (Rubicon Genomics) Library Preparation Kit with the Agilent SureSelect V5 Exome Capture system for the affected child and both parents. Sequencing was performed on the Illumina Hiseq 2500. **Variant Annotation and Prioritization.** Variants were initially prioritized based on their read depth (20X). In-house control exome data was also used

to assess the likelihood of variants being common in Iranian and Pakistani populations as well as gnomAD and the Greater Middle Eastern Variome Server. Classic algorithms will be used to predict the functional impact of variants on pathogenicity. Cross-referencing other ASD genetics studies was also used to support a decision whether a variant is implicated in ASD. All variants were Sanger-confirmed.

4.4 Results

We identified 23 *de novo* protein truncating variants (PTVs; nonsense, splice, or frameshift) in 27 families and one transmitted (i.e., inherited from an affected mother) PTV in one family (18-003) for a total of 24 putatively pathogenic variants in ZNF292 (, Figure 4.1) in 28 families. Two variants were observed in multiple families: c.6160_6161delGA (p. Glu2054Lysfs*14) found in four unrelated individuals, and c.3066_3069delAAGA (p. Glu1022Aspfs*3) found in two unrelated individuals. Two individuals in our cohort were recently reported in a large cohort of individuals with ASD/ID: 17-022 (p.Ser832Ilefs*28) and 17-023 (p. Ser1473Phefs*5).7 All ZNF292 variants identified were absent in population controls (gnomAD release v2.1), with the exception of one variant that was present at a very low frequency: c.1360C>T (p.Arg454*) present in 1 out of 248786 alleles (mean allele frequency 0.00000402). CADD (v1.4) scores (Phred-scaled) for the six nonsense variants in our series ranged from 35 to 42 with a median of 38.8 Most variants (22/24) are located in exon 8, the last and largest exon of ZNF292, which encodes a large DNA binding domain of the protein (Figure 4.1). The majority of families (20/28, 71%) had PTVs that were either insertions or deletions. Accordingly, we sought to explore whether local sequence context contributed to regional instability of this gene. We identified at least seven of these 20 insertion/deletion events appear likely to have been influenced by sequence context, including five events within palindromic repeat sequences either flanking or directly adjacent to the breakpoints and two in which the variant occurred within a mononucleotide repeat sequence (**Error! Reference source not found.**). This local sequence complexity of ZNF292 may partially explain the high frequency of somatic mutations observed in ZNF292 in tumor tissues as well.

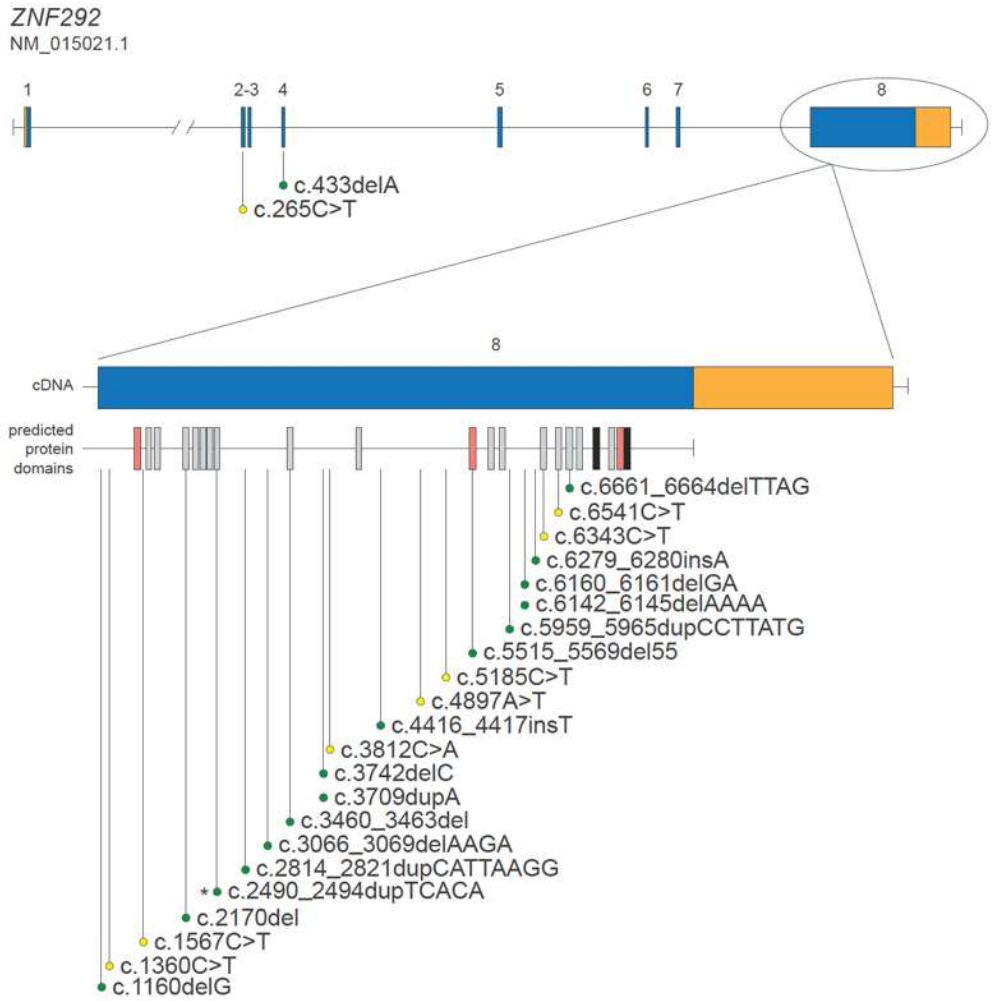


Figure 4.1 Genomic structure and distribution of variants in ZNF292. Most of the identified variants in *ZNF292* are truncating (frameshift, nonsense) located within the largest and most terminal exon (8) of the gene that encodes a ZNF292 DNA binding domain. The cDNA panel shows the coding and non-coding regions of the gene (in blue and yellow, respectively). The bottom panel shows the predicted protein domains including the zinc finger (C2H2 type) regions (shown in gray), the coiled coil domain (in pink) and the nuclear localization signal (in black). ZNF292 variants in the main cohort are shown, color-coded by type with nonsense variants shown in yellow, and frameshift variants in green

Ten rare PTVs in ZNF292 are present in the "controls-only" subset of gnomAD (release v2.1), with six frameshifts and four nonsense variants that are predicted to affect the canonical transcript. However, manual review of many of these PTVs suggest that they may be false positive calls, consistent with our observation of multiple palindromic sequences in ZNF292 that complicates read alignment. For example, manual review of the BAM files available in the

gnomAD browser reveals that two frameshifts, c.2574_2575delTC and c.2576_2577insAG, are observed only once and adjacent to one another suggesting they represent a single miscalled variant. Another variant that was annotated as nonsense, c.2690C>A, actually consists of two adjacent single-base substitutions and should have been annotated as a missense variant rather than a PTV. A third variant, c.4592delC, was listed in the gene-overview gnomAD interface but was not actually called in the heterozygous or homozygous state in any individuals (Figure A- 6). Overall, only half (i.e. two frameshifts and three nonsense variants) in gnomAD controls appear to be of high quality. These findings are consistent with other reports, including recent guidance from the gnomAD consortium, on the need for manual curation and review of PTVs called in gnomAD. 12 Furthermore, none of the high-quality variants in gnomAD are located between AA1588-2649, which correspond to zinc fingers 10-16 plus a putative coiled coil region and final nuclear localization signal, contrasting with variants in our cohort most of which are within these critical domains.13,14

ZNF292 has a pLI score of 1.0 suggesting it is highly intolerant to loss of function (LoF) variants.15 Assessing the statistical significance of observing 28 families with PTVs in ZNF292 (23/24 variants being *de novo*) is challenging due to ascertainment bias as families were ascertained via matchmaking, rather than unbiased sequencing of a single cohort. Accordingly, the true denominator (i.e. number of cases sequenced worldwide and either available via MME or in large ID/DD/autism cohorts) is unknown. This is a long-standing challenge for all studies of rare Mendelian disorders in general in which cases are ascertained and studied via matchmaking. Nevertheless, to estimate the probability of ascertaining 23 families with *de novo* PTVs in ZNF292, we tested for enrichment of *de novo* variants in this gene.17 Specifically, we approximated the number of families with ID/DD/ASD that have been sequenced worldwide and assumed that candidate genes for each family were either published or available for matchmaking via MME. The lower bound (of 100,000 individuals) is three-fold larger than the number of families with ID/DD/ASD who have reportedly undergone ES by GeneDx (personal communication, K. Retterer, GeneDx, February 4, 2019) and is the sum of families who underwent clinical sequencing from large diagnostic laboratories in the U.S. and Europe as well as those sequenced via research studies of large ID/DD/ASD cohorts. For an upper bound (300,000 individuals), we assumed that the ~800,000 persons with rare conditions estimated to have been sequenced worldwide18 likely represent ~350,000 families (~2 exomes per family)

and the primary indication for ~70% of those families was ID/DD/ASD. This is likely an overestimate, but it therefore serves as a conservative upper bound. Under these assumptions, the identification of *de novo* PTVs in 23 independent families yields a significant enrichment of between 7.2-fold (p=7.7x10-13 if N=300,000) to 21.5-fold (p=6.3x10-23 if N=100,000) compared to a exome-wide significance cutoff of p<2.7x10-6 under a Bonferroni adjustment for ~18,500 tests/genes. These data suggest that ZNF292 variants are possibly a common cause of ID and / or ASD.

All individuals in this cohort had ID with or without ASD and ADHD, with the exception of only one individual (17-023) who didn't have evidence of ID but had ASD and speech delays at age 6 years. Of the individuals with ID, delays were mild in 11/27 (40%), moderate in 6/27 (22%) and severe in 3/27 (11%). A confirmed or suspected diagnosis of ASD was present in 17/27 (62%) individuals and ADHD in 9/27 (33%) (Figure A- 7). Speech delays were prominent in this cohort, seen in 26/27 (96%) individuals overall. One proband (17-027) had severe expressive language delay at age seven years, another (17-003) was minimally verbal at age five years and two children (17-013, 17-016) were non-verbal at the ages of four years and 18 years, respectively. Two children (17-003, 17-015) also had regression of speech and language development at ages six years and two years, respectively. Another individual (17-007), a 24-year-old male, had progressive developmental issues including memory problems with a suspicion for regression overall. Most affected children walked prior to age two years with the exception of one child who remained non-ambulatory at age four years. Notably, none of the individuals in this series had isolated behavioral issues without ID or ASD.

Growth abnormalities, including short stature, were diagnosed in 11/27 individuals (Figure 4.2). Tone abnormalities were observed in 13/27 individuals including hypotonia (N = 10), hypertonia or mixed tone (N = 3). Dysmorphic facial features, most notably micrognathia and hypertelorism, were observed in thirteen individuals (Figure 4.2). Less common facial features included prominent incisors, protruding ears, and prominent nasal bridge. Overall, these facial features are not characteristic. Ocular abnormalities including nystagmus, esotropia, and strabismus were found in nine individuals. Four individuals had mild microcephaly with a head occipitofrontal circumference (OFC) of 2–3 standard deviations (SD) below the mean, and one child had an OFC of 4 SD below the mean at age four years. Overall, the observed dysmorphic

features were non-specific leading to likely low clinical recognizability of individuals with pathogenic ZNF292 variants.

Notable brain abnormalities were detected in three of twelve probands who underwent brain magnetic resonance imaging (Figure 4.3, Table A- 11). One child (17-009) had complex cerebellar abnormalities with hypoplasia of the cerebellar vermis and hemispheres, with marked cerebellar asymmetry and possible clefting within the cerebellum. There was evidence of asymmetric hemosiderin deposition on imaging suggestive of a previous vascular injury. However, there was no documented history of an in utero vascular insult or injury as pregnancy was uneventful and delivery was at term without complications. The other two children also had evidence of vascular injury on brain imaging: 18-003 was delivered premature at 25 weeks of gestation and had post hemorrhagic hydrocephalus and focal cystic encephalomalacia attributed to prematurity. The second child (17-008) had a lacunar insult in the subcortical white matter with white matter injury. This child was delivered at term without notable complications during pregnancy. Other brain MRI findings present in this cohort include ventriculomegaly, callosal abnormalities, and periventricular nodular heterotopia (seen in one individual each).

Finally, we identified 12 additional families (15 individuals total) with ZNF292 variants but for which pathogenicity was suspected but less certain because the affected person / family did not meet one or more of the following criteria: (a) lack of sufficient phenotype data from an individual with the candidate variant to determine affectation status; (b) inability to determine whether a candidate variant is inherited or *de novo* due to lack of parental genotype data (i.e. incomplete or absent parental testing information); and (c) the candidate variant was missense with no functional data to support pathogenicity.15 Nine of these families were identified by ES and six by multi-gene panel testing. The clinical and molecular data on this additional cohort are provided in Table A- 11, and Figure A- 10 and S6. Notably, these variants were also rare or absent in gnomAD controls (N = 114, 704) and had high CADD scores. One family in this cohort harbored a variant that was also identified as a *de novo* variant in our primary cohort (p. Leu2221Serfs*10). However, this family was tested via a targeted multi-gene panel and there was insufficient data to determine whether the genotype segregated with affected individuals. Therefore, we conservatively assessed this family to be of uncertain clinical affection status. Nevertheless, it is likely that this secondary cohort is enriched for additional pathogenic variants.

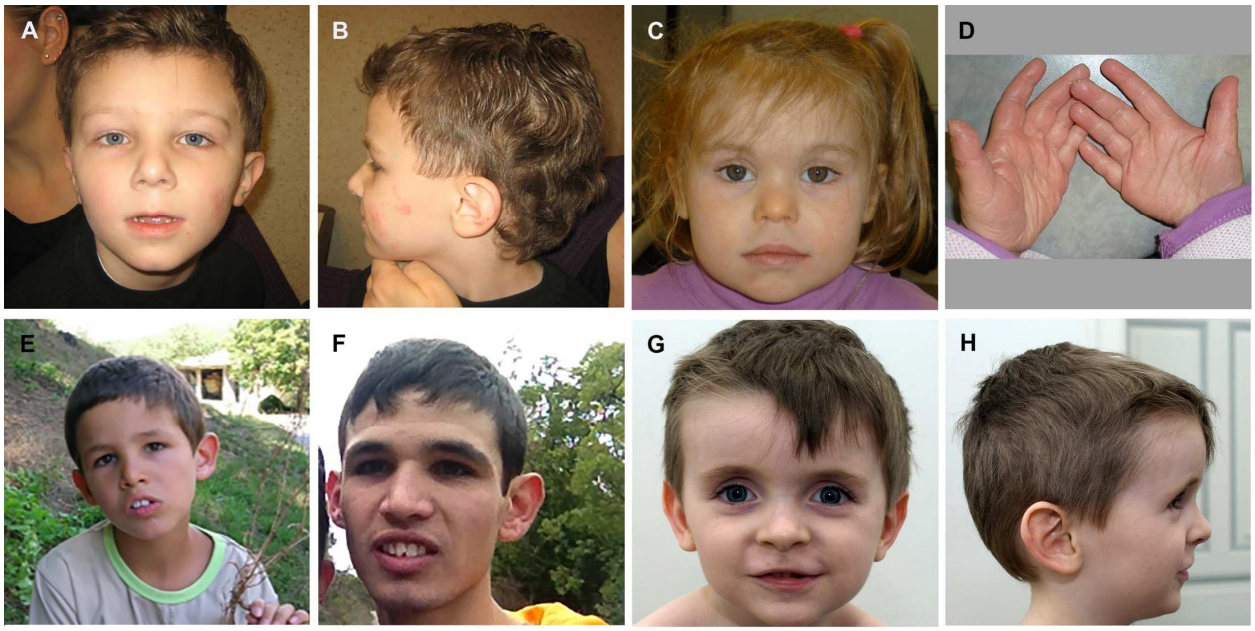


Figure 4.2 Photographs of individuals with pathogenic ZNF292 variants. (A, B) Photos of 17-027 showing a thin upper lip, smooth philtral folds, upturned nasal tip, sparse but long eyebrows with synophrys; (C-D) Photos of 18-007 at age 3.5 years showing epicanthal folds, mildly upslanted palpebral fissures, prominent forehead, and bulbous nose. D, hand photographs of the child showing ichthyosis; (E, F) Photo of 17-013 as a child (E), and as a teenager (F) showing laterally prominent ears, thick lips with a tented upper lip, short philtrum, prominent eyebrows with very prominent brow ridge and deep set eyes; (G, H) Frontal and lateral facial photograph of 17-005 at age 4 years 1 month showing mild micrognathia, short philtrum, and mildly down slanting palpebral fissures. All affected individuals have a prominent chin.

4.5 Discussion

We discovered variants in *ZNF292* that are likely pathogenic for a neurodevelopmental disorder variably accompanied by ASD and minor dysmorphic facial features that together delineate a novel condition with low clinical recognizability. Indeed, in none of the families with pathogenic *ZNF292* variants was the diagnosis of a syndrome with high clinical recognizability even considered. Accordingly, we anticipate that the number of persons with neurodevelopmental delay caused by pathogenic variants in *ZNF292* identified via multigene panels, exome- or genome sequencing will continue to grow. If this prediction is correct, it seems reasonable to wonder why *ZNF292* has not been previously reported as a priority candidate gene for NDV in large cohorts of probands or trios with ID/DD/autism.

ZNF292 first appeared as a possible candidate gene for NDV in 2012 because it was included in a supplemental table of 77 genes in which *de novo* variants were found in a cohort of 100 trios with severe ID.(Joep de Ligt et al., 2012) Over the next seven years, seven additional probands with NDV with *de novo* variants in *ZNF292* were reported across five different ID/DD/ASD cohorts or a total of eight probands with *de novo* variants in *ZNF292* variants in ~8,800 families with ID/DD/ASD tested (De Rubeis et al., 2014; Guo et al., 2018; Study et al., 2017; T. Wang et al., 2016; Ware, Samocha, Homsy, & Daly, 2015) (). In only two of these cohorts was more than one proband identified with a *de novo* variant in *ZNF292* and the largest of such studies (DDD) included only one individual with a PTV; while the others had missense variants. Accordingly, none of these studies had adequate statistical power to detect a significant enrichment of *de novo* missense variants or PTVs in *ZNF292*. A recently reported analysis of 187 candidate genes including *ZNF292* by one of our authors (H.G.) detected a significant association (p =0.016) with ID/DD/ASD only after combining *de novo* variants identified across 2,926 families with ID/DD/ASD from a previously reported association study.(Guo et al., 2018) However, this result would not reach genome-wide significance and the lack of deep phenotyping limited conclusions about both the canonical phenotype and distribution of phenotypic effects in persons with pathogenic variants in *ZNF292*.

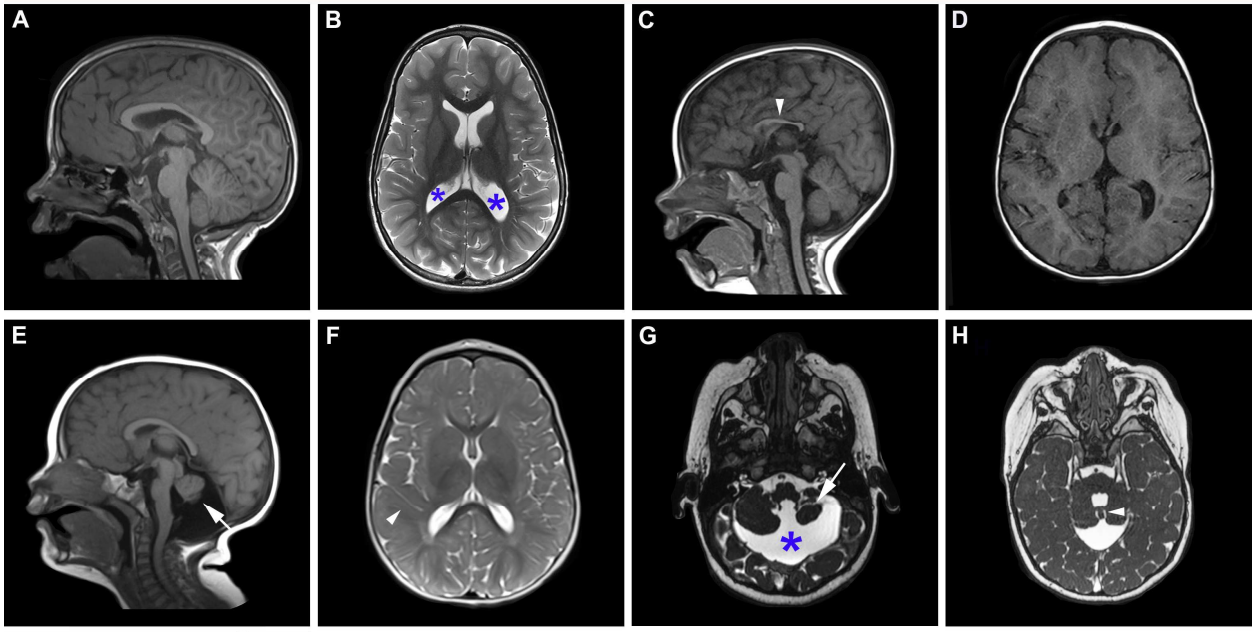


Figure 4.3 Brain MRI images of individuals with pathogenic variants in ZNF292. (A-B) T1-weighted and T2 weighted brain MR images of patient 17-003 showing mildly prominent ventricles

(asterisks, F); (C-D) T1 weighted sagittal and axial images of 17-008 showing paucity of the white matter due to an in utero vascular insult and a thin corpus callosum (arrowhead; G); (E-H) T1 weighted and Constructive Interference in Steady State (CISS) images of patient 17-009 showing multiple abnormalities including hypoplasia of the cerebellar vermis and hemispheres, with marked asymmetry (arrow, I; asterisks, G), with possible clefting OF the cerebellum (arrow, G), as well as a deep infold within the cortical surface (arrowhead, F). Patient also has evidence of possible hemosiderin deposition that is asymmetric, suggesting a previous vascular insult/injury.

Our study nicely illustrates that substantially greater statistical power can be achieved by testing a very large sample size (i.e., putatively all families with ID/DD/ASD tested to date) via effective data sharing and summing rare variants in the same candidate gene across families tested in research and clinical labs with different (i.e., gene discovery vs. diagnostic testing), albeit, complementary motivations. This strategy is expected to be particularly productive for identifying moderate to large effect alleles for genetically heterogeneous conditions with low clinical recognizability and to be much more efficient nowadays with the emergence over the past several years of multiple platforms that facilitate global sharing of candidate genes. Indeed, the process of matching to build confidence that a candidate gene is causal would likely happen much more quickly today than the seven years required to demonstrate beginning in 2012 that variants in *ZNF292* underlie an NDV disorder. However, it should also be noted that even most existing platforms for data sharing do not allow public data sharing, encourage sharing of all candidate genes identified in a family and their phenotypic data, or permit direct participation of families in matching.

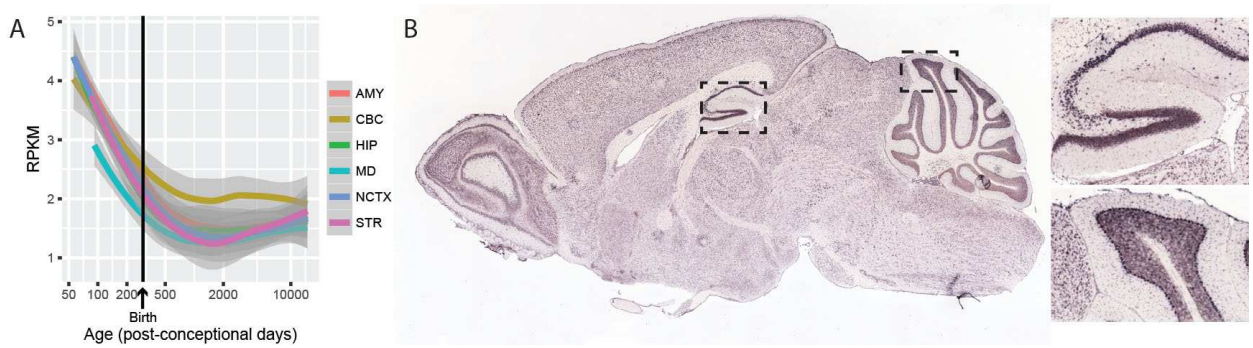


Figure 4.4 Expression of ZNF292 in developing human and mouse brains. A, *ZNF292* expression in the developing human brain (normalized RPKM data) showing high expression during early prenatal development that diminishes in the postnatal brain. *Data obtained from BrainSpan* http://www.brainspan.org . Codes: AMY, amygdala; CBC, cerebellum; HIP, hippocampus; MD, medial

dorsal nucleus of the thalamus; NCTX, neocortex; NCTX, neocortex. B, *Zpf292* expression in the adult mouse brain showing the highest expression (indicated by higher intensity staining) in hippocampus and Purkinje cells of the cerebellum.

ZNF292 is highly expressed in the developing human brain (and is among the most highly expressed ZNFs in the BrainSpan dataset), especially the cerebellum, with the highest expression identified during the prenatal period (Figure 4.4). However, the mechanism by which variants in *ZNF292* disrupt human brain development and behavior are unclear. The two most likely possibilities by which PTV might underlie disease are escape from nonsense-mediated decay (NMD), leading to expression of a truncated protein that has either a gain of function or dominant negative effect, or simple haploinsufficiency. Eighteen of the 20 unique PTVs in our cohort are predicted to escape NMD,(Coban-Akdemir et al., 2018) and in contrast to the remaining handful of high quality PTV calls in gnomAD, ten variants in our series overlap a residue between 1588-2649, which form zinc fingers 10-16, a putative coiled coil region, and the final nuclear localization signal. This is a potentially significant concentration of variants overlapping those residues (Fisher exact test p=0.004).(Kosugi, Hasebe, Tomita, & Yanagawa, 2009; Wolf, Kim, & Berger, 1997) However, the seeming concentration of PTVs in NMD-escape regions of the gene is consistent with the distribution expected by random chance (Fisher exact test p=1) and not necessarily indicative of the pathogenic mechanism as the last exon is very large (7,152 bp, or ~88% of total coding transcript length).(Hsu, Lin, & Chen, 2017) Additionally, *ZNF292* is not significantly depleted of NMD escape variants.(Coban-Akdemir et al., 2018) To examine this further, we performed RT-PCR on total RNA from two individuals, 17-005 with the p.Glu1022Aspfs*3 variant and 19-005 with the p.Arg454* variant, which showed bi-allelic expression of the normal and termination codon containing transcript, indicating that these transcripts are not degraded by nonsense□mediated mRNA decay (Figure A- 11 and Figure A- 12). Although the prematurely truncated transcript is expressed, it is still possible that the pathogenic mechanism is that of haploinsufficiency, depending what functions are retained by the truncated transcript. Haploinsufficiency is consistent with the prediction that *ZNF292* is highly intolerant to loss of function variation (pLI = 1.0) and is parsimonious given the preponderance of variants in both our primary and secondary cohorts are PTVs and not missense.

Finally, one affected parent in our cohort had mild ID that was diagnosed as an adult, suggesting that affected persons may go undiagnosed or be diagnosed later in life. This is consistent with the observation of five PTVs in the gnomAD "control" group that appear to be valid. These observations suggest that some pathogenic *ZNF292* genotypes are incompletely penetrant and/or they underlie mild ID/DD/ASD.

In summary, this study demonstrates that *de novo* and dominantly inherited variants in *ZNF292* are associated with a spectrum of neurodevelopmental features including ID, ASD, ADHD, among other features. The clinical spectrum of individuals with *ZNF292* variants is broad, with evidence of incomplete penetrance. This cohort suggests that variants in *ZNF292* are likely a relatively common cause of ID with or without ASD and other neurodevelopmental features.

5 Concluding Summary, General Discussion and Future Directions

5.1 Concluding Summary

ID affects 1 % of the population and is often comorbid with other neurodevelopmental disorders with significant overlap with ASD, complicating the clinical management/treatment options and increasing the socioeconomic burden on families, communities and healthcare systems. There are many genes that have been found for both syndromic and non-syndromic forms of ID but many cases remain a unresolved. The genetic overlap between ASD and ID remain prominent with a modern estimate 40% of ASD cases also having ID (Christensen et al., 2019; Fombonne, 2005). This strong genetic overlap has been documented previously in genes such as *SHANK3*, *PTCHD1*, *HNMT* (Heidari et al., 2015), *MeCP2* (Fieremans et al., 2014), *MYT1L* and *CC2D1A* (Durand et al., 2007; Johansen et al., 2016; Mirzaa et al., 2020; Noor et al., 2010). Novel genes in the chapters have been added to by confirmatory mutations in additional families, for example in *MBOAT7* (Johansen et al., 2016), *ZNF292* (Mirzaa et al., 2020), *PIDD1* (Harripaul et al., 2018; Hao Hu et al., 2019) Sheikh et al, submitted to Translational Psychiatry, under review. () , *DUOX2* (Bruellman et al., 2020; Peters et al., 2019), *NARS* (Manole et al, submitted to Nature Communications), *GPT2* (Ouyang et al., 2019), *ZBTB11* (Fattahi et al., 2018; Hao Hu et al., 2019), *TRAPPC6B* (Harripaul et al., 2018;; Marin-Valencia et al., 2018), *TBC1D23* (Harripaul et al., 2018;; Ivanova et al., 2017; Marin-Valencia et al., 2017), *SYNRG* (Saima Riazzuddin et al, in preparation), *CAPS2* (Hao Hu et al., 2019). Also, with the significant overlap between ID and ASD, sequencing studies for ASD have yielded many genes that appear to be shared in common with ID, however, the diagnostic yield for the ASD studies is much lower. The hypothesis that there is “not one type of ID/ASD” has come to the forefront and a more personalized approach to diagnosing and understanding the genetic heterogeneity present in ID/ASD is needed. Still many cases of ID/ASD remain undiagnosed. This sets the stage for further investigation into the mechanisms of cognitive development, and natural variation in such mechanisms that is heritable. This study also strives to identify novel genes and pathways in cognitive development through recessive gene mapping in both consanguineous and non-consanguineous populations. This thesis is composed of three main studies with the primary goal of determining the role autosomal recessive variants play in ID and ASD. The secondary goal is to determine the role

these autosomal recessive genes have in native and outbred populations. The tertiary goal is to show that using consanguineous populations for studying recessive disorders is a critical tool for understanding cognitive development and the ways in which ID and ASD and related variants can be inherited and maintained in the population. The quaternary goal is to further explore the genetic heterogeneity and pathways involved in constructing a complex organ such as the brain and ID.

The first study used 192 consanguineous multiplex families from Iran and Pakistan to perform homozygosity mapping and whole exome sequencing to identify novel genes for ID. This study has identified 26 novel genes for ID (an additional 11 novel genes were discovered from this cohort but had been reported initially in separate publications) with 72 identified genes identified in the 192 families. This cohort has achieved a 51 % diagnostic yield with a combined exome sequencing and microarray approach (10 pathogenic CNVs and 21 genes with multiple variants). Here, we have highlighted the important role recessive genes play in ID and with overlapping genetic variants in gene related to ASD, epilepsy and schizophrenia. Some of the genes identified here were found in outbred populations across the world. This study also provides a greater appreciation for the high degree of genetic heterogeneity in ID as well as within large families. The use of consanguineous families has been proven to be an excellent approach to mapping genetic disease.

The second study of this thesis focused on mapping recessive variants in ASD using parent-child trios from Iran, Pakistan and Saudi Arabia. These trios were all diagnosed ASD in their respective countries, many using ADI-R and ADOS research assessment tools. The study made use of consanguineous families, and the enrichment for recessive variants that would be found in these populations. The average degree of consanguinity in these countries ranges from 40% – 60% depending on the country. Our cohort of 115 trios had a diagnostic yield of 26% of patients. Examples of the functional model phenotypes include abnormal brain development, abnormal synaptic development, abnormal brain pruning, impaired learning and cognition or behavioral abnormalities. The developmental spatiotemporal RNA expression data suggest that these genes are important in early developmental in specific parts of the brain related to learning, memory and information integration. Using RNA-seq and microarray expression data, major brains regions were examined using constructed gene lists from known ASD and ID genes in addition to the genes identified in this cohort (Thesis n = 148 , ASD n = 1086, ID n =1346) (Banerjee-

Basu & Packer, 2010; Hamosh et al., 2005; Hao Hu et al., 2019; Jamra et al., 2011; Najmabadi et al., 2011; Reuter et al., 2017; Riazuddin et al., 2017). The striatum was identified in three separate datasets as significantly expressing the genes related to ID and ASD. The striatum has been postulated to be involved in ASD before but this is evidence for a molecular basis for this striatum being involved in neurodevelopment and expression of ASD.

The third study highlights the translational and clinical utility of the findings we presented. We have contributed a quarter of the families involved in deep phenotyping of the neurodevelopmental gene *ZNF292*. *ZNF292* is a Zinc Finger Protein, and is believed to be a transcription factor. Previously, *ZNF292* has been associated with ID and ASD but very little is known about its role and mechanism. This study made use of clinical information to assess the spectrum of disorders that have been found associated with *ZNF292*. In total, 28 families with mutations in *ZNF292* were studied, and a more nuanced clinical picture was able to be developed through this study. Most individuals with *ZNF292* variants present with mild to moderate ID and broad spectrum of developmental features, including ASD. All cases can be considered as syndromic except for a single case inherited from parent to child. It should also be noted that some variants do have an allele frequency in gnomAD and thus may indicate the presence in the population of "affected" individuals with undiagnosed mild ID. Our consanguineous families were used along with outbred families to identify the clinical features associated with *ZNF292*. This shows the value of consanguineous families and research into consanguineous families for disease gene identification.

5.2 Novelty to the Field

These three studies all have novel aspects contributing to the field of neurodevelopmental genetics. The first study was the largest study at a Canadian institute involving consanguineous multiplex families in ID and neurodevelopmental disorders. It was also the third largest study to date studying neurodevelopmental disorders using consanguineous families and it was the second largest ID study in the literature to date studying ID. It should also be noted, that this study identified 26 novel genes, with an additional 11 novel ARID genes from the cohort reported individually elsewhere (*MAN1B1*, *CC2D2A*, *TRAPPC9*, *TCTN2*, *HNMT*, *DCPS*, *MBOAT7*, *NSUN2*, *FMN2*, *METTL23* and *FBXO31*) These 26 novel genes are completely unique to our study, however, a growing list of corroborative findings have since emerged in

other studies (e.g. *PIDD1*:(Hao Hu et al., 2019); *TBC1D23*:(Marin-Valencia et al., 2018);(Ivanova et al., 2017)), that add support to a number of these genes.

The second study was the first study of its kind to use ASD consanguineous trios from to enrich and identify novel variants and genes associated with ASD. It is also one of the only studies to have looked at the number of recessive and dominant variants that come from a cohort of ASD families, and the first to do so using a rare variant approach to identify pathogenic variants for ASD. It was also the first cohort to actively determine that recessive variants are more common than previous estimates ((Martin et al., 2018)) among the Pakistani/Iranian ASD population, and potentially to other outbred populations. Previous estimates differ from our findings, however, the earlier studies were performed using a statistical approach that may potentially be insufficient to identify rare variants. This study also proposed 25 putative ASD candidate genes with 14 of them having functional models with related phenotypes.

The third study investigated the clinical symptoms and phenotypes of single nucleotide, heterozygous, rare variants in *ZNF292*, a gene that has been associated with ASD previously. The study found that the phenotypes most commonly reported in *ZNF292* patients were ID and then ASD, with a broad array of other comorbid symptoms and sometimes dysmorphic features are also possible. The inheritance was found to be *de novo*; however, one family has an inherited parent to child variant where the parent had mild ID. This last variant was of particular interest, as the variant was also found at a very low frequency in the gnomAD database. This would be consistent with the possibility that if associated with much milder/borderline ID or learning disorder such mutations may not be detected clinically until adulthood, if at all.

5.3 Contribution to the Field

In this thesis, 26 novel candidate genes have been added to the field of ID and 25 additional putative genes have been postulated for ASD. The studies in this thesis have also shown that recessive genes/variants have an appreciable role in both ASD and ID, genetics thought have tended to be underestimated in their contribution to neurodevelopmental disorders. Indeed, previously published estimates of recessive variant contributions to ID are much lower than in our study suggesting that recessive variants play a large role (Martin et al., 2018). The identification of novel variants can also contribute towards molecular diagnostic purposes, which will eventually allow more families to have answers and better clinical management options

(Page, Bell, & Watts, 2020; Scala et al., 2020; Turrini, Bourgain, & Medicine, 2020). It was also evident in our ARID study, that hormonal genes are a strong contributor to ID and particularly thyroid related genes. If these probands were identified earlier, interventions could have been put in place to ensure that their hormonal imbalance was treated earlier, which may have prevented their cognitive symptoms from getting worse. Our gene set for this cohort will also help identify converging pathways for ASD/ID that will further lead to more novel gene findings and a potential to target pathways for therapeutics.

The prospect of better diagnosis, treatment and prevention from this research could also lead to economic benefits. ASD is estimated to cost the US 236 billion dollars a year in financial costs (Griffiths et al., 2020). In other parts of the world where consanguinity is more common, involving approximately one billion people (almost one sixth of the global population), these results are even more directly applicable and the genes identified in this thesis will likely be useful for diagnostic purposes. It is estimated that 13-24% of the outbred population are affected by autosomal recessive ID (Musante & Ropers, 2014).

Some of the work published in these studies has been amongst the largest in the field; e.g. the Autosomal Recessive Intellectual Disability study is the second largest study using consanguineous families for ID and the third largest study using consanguineous family mapping from Iran and Pakistan for cognitive disabilities. The ID study using consanguineous multiplex families is also the largest study of its kind published by a Canadian group and the largest such study for ID using Pakistani families.

The results of these studies will hopefully be able to be returned clinically to the families involved, in accordance with Research Ethics Board-sanctioned procedures (i.e. involving confirmation in CLIA approved laboratories). Many of the families in these studies would not otherwise have access to the more advanced methods such as exome sequencing and the family-based counselling that our group can provide. It is also of relevance to the parents of the affected probands to understand the probability of passing disease causing genes onto future children and having a knowledge of the consequences of their carrier status. Thus improved genetic education among populations practicing endogamy would be an important future goal. The genes and variants identified in this study can also be used for diagnostic testing. There is a planned approach for returning genetic testing results to families through genetic counselling from local

genetic counsellors. The knowledge gained in this thesis will allow the scientific community to dissect the molecular underpinnings of the brain development. With more knowledge and using these genes as targets, drug screening experiments for novel therapeutics can be carried out which will lead to better clinical management.

5.4 Unifying Discussion

Throughout my thesis and investigation of ID and ASD, several themes have emerged/observations have emerged with the refinement of research on other neuropsychiatric disorders. The research demonstrates that recessive genes play a larger role in ID than was previously appreciated and that consanguineous families are an underappreciated resource for the study of common diseases, and that these findings can be translated to outbred populations (Mirzaa et al., 2020; Trost et al., 2020).

5.4.1 Recessive variants contribute to ASD and ID more than expected

In our special cohorts, there was a 51% diagnostic yield (SNVs and CNVs) for ID and a 31.3 % diagnostic yield for ASD. These disorders, specifically for ASD, had a much higher diagnostic yield, attributable to the larger number of recessive variants than previously published estimates discovered. A previous estimate, from the DDD cohort including a large number of Pakistani trios (Martin et al., 2018), estimated that, within the consanguineous cohort, 5.2 % of cases would be explained by an autosomal recessive variant. Our cohort showed a much larger, almost ~73 %, proportion of probands segregating damaging recessive variants (Table 3.3, Table 3.4 and Table 3.5). This increase in recessive variant identification ia due to the use of consanguineous families and fine-tuned filtering methods.Of these 21 % are in known neurodevelopmental disease genes, and 79% of these ID/ASD genes show supporting evidence in animal models (Figure 3.15 and Figure 3.16). These results show the underappreciated role of recessive variants in ID/ASD. The role recessive variants play may be under studied because North American cohorts do not typically have large enough sample sizes to accurately identify recessive or compound recessive mutations. The dominant model was more evident from North American cohorts as where the family sizes tend to be smaller.

5.4.2 Exome sequencing in addition to microarray is an excellent approach

The use of next generation sequencing (NGS) for disease is an incredible tool for the base level resolution of DNA changes and disease gene identification. Microarray genotyping was both cost effective and allowed a more reliable detection of CNVs in the cohort compared to WES with low quality of variable depth sequencing (Marchuk et al., 2018). However, the main use of these tools in our study was to find regions of shared homozygosity for the identification of regions of interest. The microarray was usually performed first and then the data for each family were analyzed for homozygous-by-descent regions where a disease gene is most likely to reside. This provided us with coordinates that were quickly investigated for pathogenic variants using WES data. In events where the family had no regions of shared homozygosity then we could exclude these from homozygosity mapping or move to examining other inheritance models.

5.4.3 Neurodevelopmental disorders represent a spectrum

In this thesis, and specifically, chapter 2, the genes where pathogenic variants were found were compared to other disorders from outbred population databases. We observed that 38 genes found in the ID cohort overlapped with other neuropsychiatric disorders including ASD, epilepsy and schizophrenia. These findings echoed what was observed in the field during and after this paper was published. In light of these findings, it can be postulated that neurodevelopmental disorders share common pathways associated with psychiatric disorders (Amir et al., 1999; Qiao et al., 2014). Genes such as *ATRX* and *MECP2* have many clinical features including epilepsy, schizophrenia and ASD symptoms (Amir et al., 1999; Qiao et al., 2014). Several families in our cohort, have *MECP2* gene mutations that are already deemed to be pathogenic or likely pathogenic. The variants involved may also play a role in the manifestation of the disorder. For example, Strong more truncating biallelic variants are more likely to be ID related while, missense variants are more likely to be ASD or Schizophrenia related (Singh et al., 2017). It should be noted that even the same gene can have a spectrum of clinical manifestations as seen in chapter 4. For *ZNF292*, 28 families have been listed with a broad range of clinical symptoms demonstrating that different variants in the same gene can have different phenotypes and nuances to their manifestations. The significant overlap that has been seen in other neuropsychiatric/neurodevelopmental disorders also lends support to the possible

underappreciated role recessive variants may play in common diseases and the potential role recessive variants have in explaining more cases in common disease genetics.

A major aspect of ID and ASD from chapter 2 and 3, is the extreme genetic heterogeneity observed in the families represented. There were some instances where multiple families had the same genes mutated for ID/ASD (*MECP2, AFF3, SPATA13, TBC1D23, DUOX2, VPS13B*), however, for a vast majority there was only one instance of a gene being mutated in our cohort which has been observed in other recessive variant gene identification studies (Jamra et al., 2011; Najmabadi et al., 2011; Reuter et al., 2017; Riazuddin et al., 2016, 2017).

In chapter 3, a major message from this is that rare variants contribute significantly to ID/ASD and neurodevelopmental disorders as a whole as observed previously (Satterstrom et al., 2019; Singh et al., 2017). This thesis strengthens the rare variant hypothesis for Mendelian disease but also that rare variants contribute significantly for common diseases.

As observed in the third chapter, *ZNF292* is a gene that had clinical phenotyping performed for all families and this showed a very broad spectrum of clinical symptoms associated with this gene across the families identified in the study. This spectrum includes ID, ASD and many other clinical features (including brain MRI abnormalities, ADHD, speech delays, tone abnormalities and epilepsy to name a few), which could be due to variable penetrance showing genetic pleiotropy.

In light of extreme genetic heterogeneity and the rarity of the variants that can be localized to a family or geographic region, it is necessary to treat every case as a unique event. Another major lesson learned is that the frequency of the variants discovered are so rare that clinical information is critical in understanding all the symptoms that can be the result of a gene alteration. The samples sizes that are needed to find multiple hits for a gene are so large that information has to be well organized and accurate identify meaningful phenotypes when they arise. This will feed well into the era of precision medicine an allow for precision psychiatry to be applied to patients suffering from these debilitating disorders. As more NGS tools become available and our ability to find and interpret minute genetic changes improves over time, individualized care may become available for many people suffering from neurodevelopmental disorders such as ASD and ID as seen with the work from the Deciphering Developmental Disorder group and the UK

Biobank (Bragin et al., 2013; H. C. Martin et al., 2020; Martin et al., 2018; Niemi et al., 2018; Turro et al., 2020).

5.5 Limitations

The major limitations are in clinical ascertainment and CNV calling methods in exome data. The IQ scoring for the ID and ASD cohort is non-standard due to the lack of an available standardized IQ scoring test. There is currently a lack of standardized methods for IQ assessment that have been translated to Urdu and normalized in the local population. The gold standard for ASD assessment, ADI-R and ADOS, do not currently have translated versions for use in Pakistan (Gotham, Pickles, Lord, & disorders, 2009; Lord, Rutter, Dilavore, & Risi, 2008; Lord, Storoschuk, Rutter, & Pickles, 1993), although these are being developed by our collaborative group. The clinical ascertainment was also lacking, particularily, in chapter 3 for the ASD study. The ASD study would have greatly benefitted from having both IQ scores and ADI-R and ADOS assessments to provide information on whether there was comorbid ID. Having deeper phenotyping would have also been useful as some genes identified in chapters 2 and 3 showed syndromic forms of ID or ASD. The molecular findings found for families in chapters 2 and 3 would be further validated with additional clinical information or more nuanced clinical features can lead to syndromic diagnosis of ID (Kaufman et al., 2010).

Another major limitation is the lack of CNV calls in the chapter 2 ASD study. The CNV calling was attempted using the most cited CNV calling algorithms (Fromer & Purcell, 2014; Krumm et al., 2012). Both XHMM and CoNIFER resulted in a false positive call with our ASD cohort that did not validate despite attempting to take the union of both calls. CNVs are a major contributor of pathogenic variation to both ID and ASD (Cook Jr & Scherer, 2008; Glessner et al., 2009; Levy et al., 2011; Marshall et al., 2008). The poor validation rate of our CNV calling in the ASD cohort was due to the variable read depth of individual samples sequenced and the poor quality of the received DNA. Additional methods for solving poor CNV detection are being implemented such as using microarrays for CNV detection, as well as using the CLAMMS software which has proven to be promising and effective in calling CNVs in exome data with variation in sample quality and depth (C. L. Martin et al., 2020; Packer et al., 2016).

5.6 Future Directions

Many cases remain undiagnosed as our diagnostic yields are at 51% for ID and 26 % for ASD (Harripaul et al., 2018). These yields show that there are many more cases that remain unresolved. There are other techniques and approaches that could improve and resolve the number of cases. As more and more sequencing studies are performed, with larger and larger cohorts, databases and resources like OMIM, SFARI, GeneMatcher and gnoMAD are becoming available to find novel genes and associations for ASD and ID (Banerjee-Basu & Packer, 2010; Eden et al., 2009; Karczewski & Francioli, 2017; Sobreira, Schiettecatte, Valle, & Hamosh, 2015). These resources necessitate a continual reanalysis of data as it becomes available. In these cohorts for ID and ASD, our WES data and HBD regions can be reevaluated with information from updated databases as mentioned above or contemporary literature. There is also the possibility in some families with multiple affected siblings to assume there is a high degree of genetic heterogeneity, even within a multiplex family, and to genotype the additional family members in order to perform a re-iterative HBD mapping and sequencing. There is a known contribution of CNVs in ID/ASD that can be further explored using microarray genotyping or WGS to uncover single exon CNVs and non-exonic mutations (Marshall et al., 2008). Genoptying samples on microarray platforms and additionally performing CNV detection on exome data are approaches that are also available with the data currently available from these cohort studies. Additional longer-term ventures are possible but refining the clinical data currently in these cohorts is a necessary first step. The more detailed and refined the clinical descriptions are is also an excellent way of finding more nuanced phenotyping to identify novel genes

5.6.1 More detailed clinical descriptions

One of the major lessons learned is that detailed clinical information is imperative for identifying syndromic and clinical features associated with a specific gene mutation. Due to the rarity of the mutations, as is the nature of the field, detailed clinical information is necessary to link phenotypes with genes and identify the phenotypes in other individuals. Ideally, detailed clinical information and cognitive, and adaptive behavior tests should be performed on all probands especially related to ID and ASD since there is such a large overlap in their genetic profiles. Another parameter that would be useful is to have the IQ score/educational level and some basic

information about the parents such as age, cognitive testing, other relatives that affected. It is also imperative that standardized scales IQ and diagnosis methods such as ADI-R and ADOS are used where possible which presented as a challenge to this study (Gotham et al., 2009; Lord et al., 2008; Lord et al., 1993). IQ testing and the ADI-R/ADOS will need to be translated to their respective languages and cultures and the methods of clinical data collection standardized (Gotham et al., 2009; Lord, DiLavore, & Gotham, 2012). For instance, currently, there is no Urdu version of the ADI-R available for studies.

5.6.2 Whole Genome Sequencing Potential for ASD and ID

Next Generation Sequencing has no doubt revolutionized modern genetics, and, while it has become increasingly more cost effective the technology has also been further refined. The use of exome sequencing yielded a tremendous number of variants and novel genes in the coding regions of the genome. The coding region of the genome, however, only accounts for ~ 1.5 % of the human genome.

The use of WGS holds an enormous potential to understand biology. The use of WGS can be used for better structural variant detection as well as CNVs, which has been long associated with ASD and ID for as a major contributor to disease (Marshall et al., 2008; Tammimies et al., 2015).. As the price of sequencing falls, WGS is becoming a more affordable and feasible option for larger scale studies. The use of WGS data would enable a broader range of variants to be detected including more difficult to assess variants such as structural variants, expansion repeats as well as being able to detect point mutations, CNVs, indels in a single assay. Due to the high variability of the sequencing depth in our WES data, CNV calling was extremely difficult resulting in a high rate of false positives. This resulted in difficulty in resolving CNVs and impaired our ability to find CNVs. WES data can be used for CNV calling; however, this necessitates a much higher read depth. Use of WGS, if applied to these ASD/ID probands would yield more variants of interest. The use of WGS allow for the standard indels and SNVs to be detected, however, there would be greater power to detect intronic, tandem repeats and splicing variants (Gardner et al., 2019; Jaganathan et al., 2019; Trost et al., 2020; Turro et al., 2020). A recent study was able to detect 10 % more pathogenic cryptic splicing variants with WGS data (Jaganathan et al., 2019). Often times, splicing variants are mislabeled as synonymous variants (Jaganathan et al., 2019). WGS also allows for non-coding regions of the genome to be

interrogated. Interpreting non-coding regions of the genome is exceptionally difficult even with the most modern sequencing technologies, however, this is not an impossible task as recent genetics studies have identified *de novo* LoF variants in non-coding regions(Xiong et al., 2015; Yuen et al., 2017). As annotation of the genome becomes better and machine-learning/pattern recognition algorithms are developed, the non-coding genome will become more accessible and interpretable for scientific study.

5.6.3 Advances Utilizing RNA Sequencing

Through large cohorts and databases, this thesis has used RNA-sequencing to probe the spatiotemporal expression of genes identified to be significantly associated with ASD and ID. This has provided many insights into the genes, timing and tissue specific expression with genes involved in ASD and ID. Studies have shown that long non-coding RNAs are differentially expressed in ASD, and this topic has not been attempted at a large scale for ID (Akbarian et al., 2015; Gandal et al., 2018; Short et al., 2018). The use of RNA-sequencing for ID may be more difficult due to the extreme genetic and clinical heterogeneity observed in probands. Nonetheless, RNA-sequencing can still offer novel findings for ID and confirmation of variants that have been identified. RNA can also be used to validate and identify novel splicing variants that are often times mislabeled in standard annotation packages like MutationTaster, MaxEntScan and Polyphen2 (Jaganathan et al., 2019). Using the genome sequencing approach, 10% more pathogenic variants can be identified (Jaganathan et al., 2019) and change in splicing can alter the splicing of transcript in the human brain (Walker et al., 2019). Recent studies have also used RNA sequencing to recognize expression hubs or genes that are most expressed in ASD (Gandal et al., 2018). RNA-sequencing can also be scaled to single cell resolution that will disentangle the cellular mechanisms of ASD or ID by showing tissue specific and cell specific expression patterns for the developing brain (Welch et al., 2019). The insights gained in single cell studies in control populations can allow a baseline expression to be established and understand individual cell type expression per tissue type (Welch et al., 2019). In chapter 3, cell type expression was interrogated to identify cell types that are most associated with ASD and ID genes in tissue types and cell types that have major implications for brain development and cognitive disorder etiology. As a note, the development of computational biology tools and particularly machine learning algorithms to recognize the patterns inherent in complex brain

development will be an active area of research and influential in our understanding of the brain and brain disorder.

5.6.4 Translation and Functional Validation of Intellectual Disability and Autism Spectrum Disorder Genes

Many genes have been identified for both ID and ASD, however, few have functional and causative models developed. The rarity of these mutations in the population or even cohorts where they are identified means that there are few cases that will have the same gene identified and this will invariably reduce the statistical power that is needed for a true association. Finding multiple families, is a first step and groups such as GeneMatcher are useful in identifying other probands with mutations in the same genes (Sobreira et al., 2015). For rare variants, functional models are an excellent way to demonstrate a genetic association to disease, and particularly if there are multiple affected families with variants in the same gene (Eppig et al., 2015). Typically, it is useful to start on induced pluripotent stem cells (iPSCs) that can be induced to the cell type of desire, neuronal progenitor cells in the case of ID (B.-Y. Hu et al., 2010). The method of choice for inducing mutations into cells is now CRISPR/cas9 (Ran et al., 2013). Induced iPSCs can then be used for cell imaging and immunoprecipitation assays to determine the likely cell localization and accessory proteins that are involved. The gene itself is also capable of being cloned with a GFP to aid in these studies as well as having mutations induced in neuronal stem cells (Feng et al., 2000). Induced iPSCs techniques will greatly, enhance the study of neuronal function and morphology. These are useful future directions to gain a grasp of the molecular mechanisms involved with this gene and disease. There are limitations to using cells, as social dysfunction and intellect are difficult to assess in a cell, which is why cell work will be a stepping stone towards functional models. Chapter 2 and 3 have reported known mouse and zebrafish functional models that exhibit the appropriate phenotypes such as aberrant dendritic spines, aberrant axon direction/growth, lower water maze scores, abnormal behavioural/neurological phenotypes. Transgenic mouse and zebrafish models can be used for further studies. In the absence of a functional model for a gene they can be created using CRISPR-Cas9 gene editing (Ran et al., 2013). All models developed, from iPSCs to zebrafish, can be useful for high-throughput drug discovery and then translated to mouse models (Pott, Rottbauer, & Just, 2020). Drug screening against a large chemical large or library of CRISPR-Cas9 probes can yield many therapeutic compounds and potentially cures for ASD/ID that is

targeted and personalized for every gene identified (Fellmann, Gowen, Lin, Doudna, & Corn, 2017). Due to the fact that there are many genes, it will be sensible to look for the most frequently effected genes.

6 References

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013a). Predicting functional effect of human missense mutations using PolyPhen□2. *Current protocols in human genetics*, 7.20. 21-27.20. 41.

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013b). Predicting functional effect of human missense mutations using PolyPhen□2. *Current protocols in human genetics, 76*(1), 7.20. 21-27.20. 41.

Adzhubei, I., Schmidt, S., Peshkin, L., Ramensky, V., Gerasimova, A., & Bork, P. (2010). Catalog of Somatic Mutations in Cancer and bioinformatics mutation-prediction tools PolyPhen2. *Nat Methods, 7*, 248-249.

Ahmed, I., Buchert, R., Zhou, M., Jiao, X., Mittal, K., Sheikh, T. I., . . . Brohi, M. Q. (2015). Mutations in DCPS and EDC3 in autosomal recessive intellectual disability indicate a crucial role for mRNA decapping in neurodevelopment. *Human molecular genetics, 24*(11), 3172-3180.

Akbarian, S., Liu, C., Knowles, J. A., Vaccarino, F. M., Farnham, P. J., Crawford, G. E., . . . Geschwind, D. H. (2015). The psychencode project. *Nature neuroscience, 18*(12), 1707.

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature genetics, 23*(2), 185-188.

Anazi, S., Maddirevula, S., Faqeih, E., Alsedairy, H., Alzahrani, F., Shamseldin, H., . . . Abdulwahab, F. (2017). Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. *Molecular Psychiatry, 22*(4), 615-624.

Aoki, Y., Yoncheva, Y. N., Chen, B., Nath, T., Sharp, D., Lazar, M., . . . Di Martino, A. (2017). Association of white matter structure with autism spectrum disorder and attention-deficit/hyperactivity disorder. *JAMA psychiatry, 74*(11), 1120-1128.

Arloth, J., Bader, D. M., Röh, S., & Altmann, A. (2015). Re-Annotator: Annotation pipeline for microarray probe sequences. *PLoS One, 10*(10).

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Eppig, J. T. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics, 25*(1), 25-29.

Autry, A. E., & Monteggia, L. M. (2012). Brain-derived neurotrophic factor and neuropsychiatric disorders. *Pharmacological reviews, 64*(2), 238-258.

Baio, J. (2014). Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010.

Banerjee-Basu, S., & Packer, A. (2010). SFARI Gene: an evolving database for the autism research community. *Disease Models and Mechanisms, 3*(3-4), 133-135.

Barber, J. C., Maloney, V. K., Bewes, B., & Wakeling, E. (2006). Deletions of 2q14 that include the homeobox engrailed 1 (EN1) transcription factor are compatible with a normal phenotype. *European journal of human genetics, 14*(6), 739-743.

Bastian, F., Parmentier, G., Roux, J., Moretti, S., Laudet, V., & Robinson-Rechavi, M. (2008). *Bgee: integrating and comparing heterogeneous transcriptome data among species.* Paper presented at the Data Integration in the Life Sciences.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological), 57*(1), 289-300.

Bernkopf, M., Webersinke, G., Tongsook, C., Koyani, C. N., Rafiq, M. A., Ayaz, M., . . . Naeem, F. (2014). Disruption of the methyltransferase-like 23 gene METTL23 causes mild autosomal recessive intellectual disability. *Human molecular genetics, 23*(15), 4015-4023.

Bolger, A. M., Lohse, M., & Usadel, B. J. B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *30*(15), 2114-2120.

Bradford, Y., Conlin, T., Dunn, N., Fashena, D., Frazer, K., Howe, D. G., . . . Moxon, S. A. (2011). ZFIN: enhancements and updates to the Zebrafish Model Organism Database. *Nucleic acids research, 39*(suppl 1), D822-D829.

Bradshaw, N. J., Trossbach, S. V., Köber, S., Walter, S., Prikulis, I., Weggen, S., & Korth, C. (2017). Disrupted in Schizophrenia 1 regulates the processing of reelin in the perinatal cortex. *Schizophrenia research*.

Buescher, A. V., Cidav, Z., Knapp, M., & Mandell, D. S. (2014). Costs of autism spectrum disorders in the United Kingdom and the United States. *JAMA pediatrics, 168*(8), 721-728.

Carlson, M. GO. db: A set of annotation maps describing the entire Gene Ontology [Internet]. 2016. In.

Clarke, T., Lupton, M., Fernandez-Pujals, A., Starr, J., Davies, G., Cox, S., . . . MacIntyre, D. (2015). Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Molecular Psychiatry*.

Coban-Akdemir, Z., White, J. J., Song, X., Jhangiani, S. N., Fatih, J. M., Gambin, T., . . . Punetha, J. J. T. A. J. o. H. G. (2018). Identifying genes whose mutant transcripts cause dominant disease traits by potential gain-of-function alleles. *103*(2), 171-187.

Colantuoni, C., Lipska, B. K., Ye, T., Hyde, T. M., Tao, R., Leek, J. T., . . . Weinberger, D. R. (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature, 478*(7370), 519-523.

Craig, F., Margari, F., Legrottaglie, A. R., Palumbi, R., De Giambattista, C., & Margari, L. (2016). A review of executive function deficits in autism spectrum disorder and attention-deficit/hyperactivity disorder. *Neuropsychiatric disease and treatment, 12*, 1191.

Cuchillo-Ibáñez, I., Andreo-Lillo, P., Pastor-Ferrándiz, L., Carratalá-Marco, F., & Sáez-Valero, J. (2020). Elevated Plasma Reelin Levels in Children With Autism. *Frontiers in Psychiatry, 11*, 242.

Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., . . . Quake, S. R. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences, 112*(23), 7285-7290.

Darnell, J. C., Van Driesche, S. J., Zhang, C., Hung, K. Y. S., Mele, A., Fraser, C. E., . . . Chi, S. W. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell, 146*(2), 247-261.

David, M., Dzamba, M., Lister, D., Ilie, L., & Brudno, M. (2011). SHRiMP2: sensitive yet practical short read mapping. *Bioinformatics, 27*(7), 1011-1012.

Davies, G., Marioni, R. E., Liewald, D. C., Hill, W. D., Hagenaars, S. P., Harris, S. E., . . . Lyall, D. (2016). Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N= 112 151). *Molecular Psychiatry*.

Dawson, M., Mottron, L., & Gernsbacher, M. A. (2008). Learning in autism. *Learning and memory: A comprehensive reference, 2*, 759-772.

de Ligt, J., Willemsen, M. H., van Bon, B. W., Kleefstra, T., Yntema, H. G., Kroes, T., . . . Gilissen, C. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *New England Journal of Medicine, 367*(20), 1921-1929.

de Ligt, J., Willemsen, M. H., van Bon, B. W., Kleefstra, T., Yntema, H. G., Kroes, T., . . . Vissers, L. E. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med, 367*(20), 1921-1929. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23033978. doi:10.1056/NEJMoa1206524

De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., . . . Walker, S. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature, 515*(7526), 209-215.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., . . . Hanna, M. J. N. g. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *43*(5), 491.

Dong, E., Gavin, D. P., Chen, Y., & Davis, J. (2012). Upregulation of TET1 and downregulation of APOBEC3A and APOBEC3C in the parietal cortex of psychotic patients. *Translational psychiatry, 2*(9), e159.

Dong, E., Ruzicka, W., Grayson, D., & Guidotti, A. (2015). DNA-methyltransferase1 (DNMT1) binding to CpG rich GABAergic and BDNF promoters is increased in the brain of schizophrenia and bipolar disorder patients. *Schizophrenia research, 167*(1), 35-41.

Durand, C. M., Betancur, C., Boeckers, T. M., Bockmann, J., Chaste, P., Fauchereau, F., . . . Anckarsäter, H. (2007). Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature genetics, 39*(1), 25-27.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC bioinformatics, 10*(1), 48.

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), 57-74.

Fatemi, S., Earle, J., & McMenomy, T. (2000). Reduction in Reelin immunoreactivity in hippocampus of subjects with schizophrenia, bipolar disorder and major depression. *Molecular Psychiatry, 5*(6), 654-663.

Figueiredo, T., Melo, U., Pessoa, A., Nobrega, P., Kitajima, J., Rusch, H., . . . Kok, F. (2015). A homozygous loss-of-function mutation in inositol monophosphatase 1 (IMPA1) causes severe intellectual disability. *Molecular Psychiatry*.

Folsom, T. D., & Fatemi, S. H. (2013). The involvement of Reelin in neurodevelopmental disorders. *Neuropharmacology, 68*, 122-135.

Franić, S., Groen-Blokhuis, M. M., Dolan, C. V., Kattenberg, M. V., Pool, R., Xiao, X., . . . van der Sluis, S. (2015). Intelligence: shared genetic basis between Mendelian disorders and a polygenic trait. *European Journal of Human Genetics*.

Fromer, M., Pocklington, A. J., Kavanagh, D. H., Williams, H. J., Dwyer, S., Gormley, P., . . . Ruderfer, D. M. (2014). *De novo* mutations in schizophrenia implicate synaptic networks. *Nature, 506*(7487), 179-184.

Frotscher, M. (1998). Cajal—Retzius cells, Reelin, and the formation of layers. *Current opinion in neurobiology, 8*(5), 570-575.

Gazal, S., Sahbatou, M., Babron, M.-C., Génin, E., & Leutenegger, A.-L. (2014). FSuite: exploiting inbreeding in dense SNP chip and exome data. *Bioinformatics, 30*(13), 1940-1941.

Gazal, S., Sahbatou, M., Babron, M. C., Genin, E., & Leutenegger, A. L. (2014). FSuite: exploiting inbreeding in dense SNP chip and exome data. *Bioinformatics, 30*(13), 1940-

1941. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24632498. doi:10.1093/bioinformatics/btu149

Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W., Willemsen, M. H., . . . Schenck, A. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature*.

Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W., Willemsen, M. H., . . . Veltman, J. A. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature, 511*(7509), 344-347. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24896178. doi:10.1038/nature13394

Glessner, J. T., Wang, K., Cai, G., Korvatska, O., Kim, C. E., Wood, S., . . . Bradfield, J. P. (2009). Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature, 459*(7246), 569-573.

Gogolla, N. (2017). The insular cortex. *Current Biology, 27*(12), R580-R586.

Gordon, A., & Hannon, G. J. F. A. s.-r. p. t. h. h. c. e. f. t. (2010). Fastx-toolkit. *5*.

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., . . . Anney, R. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics, 51*(3), 431-444.

Guo, H., Wang, T., Wu, H., Long, M., Coe, B. P., Li, H., . . . Duan, G. J. M. a. (2018). Inherited and multiple *de novo* mutations in autism/developmental delay risk genes suggest a multifactorial model. *9*(1), 64.

Haddad, L. A., Mingroni-Netto, R. C., Vianna-Morgante, A. M., & Pena, S. D. (1996). A PCR-based test suitable for screening for fragile X syndrome among mentally retarded males. *Human genetics, 97*(6), 808-812.

Hall, J., & Owen, M. J. (2015). Psychiatric classification–a developmental perspective. *The British Journal of Psychiatry, 207*(4), 281-282.

Hamamy, H. (2012). Consanguineous marriages. *Journal of community genetics, 3*(3), 185-192.

Hamdan, F. F., Gauthier, J., Araki, Y., Lin, D.-T., Yoshizawa, Y., Higashi, K., . . . Piton, A. (2011). Excess of *de novo* deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *The American Journal of Human Genetics, 88*(3), 306-316.

Hamdan, F. F., Srour, M., Capo-Chichi, J. M., Daoud, H., Nassif, C., Patry, L., . . . Michaud, J. L. (2014). *De novo* mutations in moderate or severe intellectual disability. *PLoS Genet, 10*(10), e1004772. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25356899. doi:10.1371/journal.pgen.1004772

Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., & McKusick, V. A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic acids research, 33*(suppl 1), D514-D517.

Harripaul, R., Vasli, N., Mikhailov, A., Rafiq, M. A., Mittal, K., Windpassinger, C., . . . Downey, S. (2018). Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. *Molecular Psychiatry, 23*(4), 973-984.

Harripaul, R., Vasli, N., Mikhailov, A., Rafiq, M. A., Mittal, K., Windpassinger, C., . . . Vincent, J. B. (2017). Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. *Molecular Psychiatry*. Retrieved from http://dx.doi.org/10.1038/mp.2017.60. doi:10.1038/mp.2017.60

https://www.nature.com/articles/mp201760#supplementary-information

Heidari, A., Tongsook, C., Najafipour, R., Musante, L., Vasli, N., Garshasbi, M., . . . Sritharan, K. (2015). Mutations in the histamine N-methyltransferase gene, HNMT, are associated with nonsyndromic autosomal recessive intellectual disability. *Human molecular genetics, 24*(20), 5697-5710.

Heslop, P., Blair, P. S., Fleming, P., Hoghton, M., Marriott, A., & Russ, L. (2014). The Confidential Inquiry into premature deaths of people with intellectual disabilities in the UK: a population-based study. *The Lancet, 383*(9920), 889-895.

Howard, D., Negraes, P., Voineskos, A. N., Kaplan, A. S., Muotri, A., Duvvuri, V., & French, L. (2018). Molecular neuroanatomy of anorexia nervosa. *bioRxiv*, 440313.

Howard, H. C., Knoppers, B. M., Cornel, M. C., Clayton, E. W., Sénécal, K., & Borry, P. (2015). Whole-genome sequencing in newborn screening? A statement on the continued importance of targeted approaches in newborn screening programmes. *European Journal of Human Genetics*.

Hsu, M.-K., Lin, H.-Y., & Chen, F.-C. J. P. o. (2017). NMD Classifier: A reliable and systematic classification tool for nonsense-mediated decay events. *12*(4).

Hu, H., Haas, S., Chelly, J., Van Esch, H., Raynaud, M., de Brouwer, A., . . . Laumonnier, F. (2016). X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. *Molecular Psychiatry, 21*(1), 133-148.

Hu, H., Haas, S. A., Chelly, J., Van Esch, H., Raynaud, M., de Brouwer, A. P., . . . Kalscheuer, V. M. (2015). X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. *Mol Psychiatry*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25644381. doi:10.1038/mp.2014.193

Hu, H., Kahrizi, K., Musante, L., Fattahi, Z., Herwig, R., Hosseini, M., . . . Larti, F. (2019). Genetics of intellectual disability in consanguineous families. *Molecular Psychiatry, 24*(7), 1027-1039.

Huang, J., Zhu, T., Qu, Y., & Mu, D. (2016). Prenatal, Perinatal and Neonatal Risk Factors for Intellectual Disability: A Systemic Review and Meta-Analysis. *PLoS One, 11*(4), e0153655.

Huang, L., Jolly, L. A., Willis-Owen, S., Gardner, A., Kumar, R., Douglas, E., . . . Cohen, M. (2012). A noncoding, regulatory mutation implicates HCFC1 in nonsyndromic intellectual disability. *The American Journal of Human Genetics, 91*(4), 694-702.

Hussain, R., & Bittles, A. (1998). The prevalence and demographic characteristics of consanguineous marriages in Pakistan. *Journal of biosocial science, 30*(02), 261-275.

Iqbal, Z., Neveling, K., Razzaq, A., Shahzad, M., Zahoor, M. Y., Qasim, M., . . . van Bokhoven, H. (2012). Targeted next generation sequencing reveals a novel intragenic deletion of the TPO gene in a family with intellectual disability. *Arch Med Res, 43*(4), 312-316. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22387573. doi:10.1016/j.arcmed.2012.01.011

Jaffe, A. E., Gao, Y., Deep-Soboslay, A., Tao, R., Hyde, T. M., Weinberger, D. R., & Kleinman, J. E. (2015). Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nature neuroscience*.

Jagadeesh, K. A., Wenger, A. M., Berger, M. J., Guturu, H., Stenson, P. D., Cooper, D. N., . . . Bejerano, G. (2016). *M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity* (1061-4036). Retrieved from

Jamra, R. A., Wohlfart, S., Zweier, M., Uebe, S., Priebe, L., Ekici, A., . . . Fakher, M. (2011). Homozygosity mapping in 64 Syrian consanguineous families with non-specific intellectual disability reveals 11 novel loci and high heterogeneity. *European Journal of Human Genetics, 19*(11), 1161-1166.

Jensen, L. R., Lenzner, S., Moser, B., Freude, K., Tzschach, A., Wei, C., . . . Moraine, C. (2007). X-linked mental retardation: a comprehensive molecular screen of 47 candidate genes from a 7.4 Mb interval in Xp11. *European Journal of Human Genetics, 15*(1), 68-75.

Jian, X., Boerwinkle, E., & Liu, X. (2014). In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic acids research, 42*(22), 13534-13544.

Johansen, A., Rosti, R. O., Musaev, D., Sticca, E., Harripaul, R., Zaki, M., . . . Froukh, T. (2016). Mutations in MBOAT7, Encoding Lysophosphatidylinositol Acyltransferase I, Lead to Intellectual Disability Accompanied by Epilepsy and Autistic Features. *The American Journal of Human Genetics*.

Kakar, N., Goebel, I., Daud, S., Nürnberg, G., Agha, N., Ahmad, A., . . . Borck, G. (2012). A homozygous splice site mutation in TRAPPC9 causes intellectual disability and microcephaly. *European journal of medical genetics, 55*(12), 727-731.

Karalunas, S. L., Hawkey, E., Gustafsson, H., Miller, M., Langhorst, M., Cordova, M., . . . Nigg, J. T. (2018). Overlapping and distinct cognitive impairments in attention-deficit/hyperactivity and autism spectrum disorder without intellectual disability. *Journal of abnormal child psychology, 46*(8), 1705-1716.

Karczewski, K., & Francioli, L. (2017). The Genome Aggregation Database (gnomAD). *MacArthur Lab*.

Kaufman, L., Ayub, M., & Vincent, J. B. (2010). The genetic basis of non-syndromic intellectual disability: a review. *Journal of neurodevelopmental disorders, 2*(4), 182-209.

Kevelam, S. H., van Harssel, J. J., van der Zwaag, B., Smeets, H. J., Paulussen, A. D., & Lichtenbelt, K. D. (2012). A patient with a mild holoprosencephaly spectrum phenotype and heterotaxy and a 1.3 Mb deletion encompassing GLI2. *American Journal of Medical Genetics Part A, 158*(1), 166-173.

Khan, M. A., Rafiq, M. A., Noor, A., Ali, N., Ali, G., Vincent, J. B., & Ansar, M. (2011). A novel deletion mutation in the TUSC3 gene in a consanguineous Pakistani family with autosomal recessive nonsyndromic intellectual disability. *BMC medical genetics, 12*(1), 56.

Khan, M. A., Rafiq, M. A., Noor, A., Hussain, S., Flores, J. V., Rupp, V., . . . Khan, F. S. (2012). Mutation in NSUN2, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability. *The American Journal of Human Genetics, 90*(5), 856-863.

Knoppers, B. M., Avard, D., Sénécal, K., & Zawati, M. n. H. (2014). Return of whole-genome sequencing results in paediatric research: a statement of the P3G international paediatrics platform. *Eur J Hum Genet, 22*(1), 3-5.

Kosugi, S., Hasebe, M., Tomita, M., & Yanagawa, H. J. P. o. t. N. A. o. S. (2009). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *106*(25), 10171-10176.

Krumm, N., Turner, T. N., Baker, C., Vives, L., Mohajeri, K., Witherspoon, K., . . . He, Z.-X. (2015). Excess of rare, inherited truncating mutations in autism. *Nature genetics, 47*(6), 582-588.

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic acids research, 42*(D1), D980-D985.

Law, R., Dixon-Salazar, T., Jerber, J., Cai, N., Abbasi, A. A., Zaki, M. S., . . . Khan, V. (2014). Biallelic truncating mutations in FMN2, encoding the actin-regulatory protein formin 2,

cause nonsyndromic autosomal-recessive intellectual disability. *The American Journal of Human Genetics, 95*(6), 721-728.

Lee, H., Deignan, J. L., Dorrani, N., Strom, S. P., Kantarci, S., Quintero-Rivera, F., . . . Yourshaw, M. (2014). Clinical exome sequencing for genetic identification of rare Mendelian disorders. *Jama, 312*(18), 1880-1887.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., . . . Cummings, B. B. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature, 536*(7616), 285-291.

Li, C., Lan, Y., Schwartz-Orbach, L., Korol, E., Tahiliani, M., Evans, T., & Goll, M. G. (2015). Overlapping Requirements for Tet2 and Tet3 in Normal Development and Hematopoietic Stem Cell Emergence. *Cell reports, 12*(7), 1133-1143.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.

Li, J., Cai, T., Jiang, Y., Chen, H., He, X., Chen, C., . . . Li, Z. (2016). Genes with *de novo* mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. *Molecular psychiatry*, 21(12, 290-297

MacDonald, J. R., Ziman, R., Yuen, R. K., Feuk, L., & Scherer, S. W. (2013). The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic acids research*, 42(D1), D986-D992 .

Madsen, G. F., Bilenberg, N., Cantio, C., & Oranje, B. (2014). Increased prepulse inhibition and sensitization of the startle reflex in autistic children. *Autism Research, 7*(1), 94-103.

Maki, T., Liang, A. C., Miyamoto, N., Lo, E. H., & Arai, K. (2013). Mechanisms of oligodendrocyte regeneration from ventricular-subventricular zone-derived progenitor cells in white matter diseases. *Frontiers in Cellular Neuroscience, 7*, 275.

Mancarci, B. O., Toker, L., Tripathy, S. J., Li, B., Rocco, B., Sibille, E., & Pavlidis, P. (2017). Cross-laboratory analysis of brain cell type transcriptomes with applications to interpretation of bulk tissue data. *Eneuro, 4*(6).

Marangi, G., Leuzzi, V., Manti, F., Lattante, S., Orteschi, D., Pecile, V., . . . Zollino, M. (2013). TRAPPC9-related autosomal recessive intellectual disability: report of a new mutation and clinical phenotype. *European Journal of Human Genetics, 21*(2), 229-232.

Marshall, C. R., Noor, A., Vincent, J. B., Lionel, A. C., Feuk, L., Skaug, J., . . . Ren, Y. (2008). Structural variation of chromosomes in autism spectrum disorder. *The American Journal of Human Genetics, 82*(2), 477-488.

Martin, H. C., Jones, W. D., McIntyre, R., Sanchez-Andrade, G., Sanderson, M., Stephenson, J. D., . . . Bruntraeger, M. (2018). Quantifying the contribution of recessive coding variation to developmental disorders. *Science, 362*(6419), 1161-1164.

Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T., & Saxena, S. (2011). Prevalence of intellectual disability: a meta-analysis of population-based studies. *Research in developmental disabilities, 32*(2), 419-436.

McCarthy, S. E., Gillis, J., Kramer, M., Lihm, J., Yoon, S., Berstein, Y., . . . Ghiban, E. (2014). *De novo* mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Molecular psychiatry, 19*(6), 652-658.

McLaren, J., & Bryson, S. E. (1987). Review of recent epidemiological studies of mental retardation: prevalence, associated disorders, and etiology. *American Journal on Mental Retardation*.

MD, I. B., M. T., & Wiestler, O. D. (2002). Ammon's horn sclerosis: a maldevelopmental disorder associated with temporal lobe epilepsy. *Brain Pathology, 12*(2), 199-211.

Meloni, I., Bruttini, M., Longo, I., Mari, F., Rizzolio, F., D'Adamo, P., . . . Renieri, A. (2000). A mutation in the Rett syndrome gene, MECP2, causes X-linked mental retardation and progressive spasticity in males. *The American Journal of Human Genetics, 67*(4), 982-985.

Metsu, S., Rooms, L., Rainger, J., Taylor, M. S., Bengani, H., Wilson, D. I., . . . Reyniers, E. (2014). FRA2A is a CGG repeat expansion associated with silencing of AFF3. *PLoS Genet, 10*(4), e1004242.

Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., . . . Epstein, C. J. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *The American Journal of Human Genetics, 86*(5), 749-764.

Miller, J. A., Ding, S.-L., Sunkin, S. M., Smith, K. A., Ng, L., Szafer, A., . . . Aiona, K. (2014). Transcriptional landscape of the prenatal human brain. *Nature, 508*(7495), 199-206.

Mir, A., Kaufman, L., Noor, A., Motazacker, M. M., Jamil, T., Azam, M., . . . Nasr, T. (2009). Identification of mutations in TRAPPC9, which encodes the NIK-and IKK-β-binding protein, in nonsyndromic autosomal-recessive mental retardation. *The American Journal of Human Genetics, 85*(6), 909-915.

Mir, A., Sritharan, K., Mittal, K., Vasli, N., Araujo, C., Jamil, T., . . . Rauf, S. (2014). Truncation of the E3 ubiquitin ligase component FBXO31 causes non-syndromic autosomal recessive intellectual disability in a Pakistani family. *Human genetics, 133*(8), 975-984.

Mirzaa, G. M., Chong, J. X., Piton, A., Popp, B., Foss, K., Guo, H., . . . Aldinger, K. A. (2020). *De novo* and inherited variants in ZNF292 underlie a neurodevelopmental disorder with features of autism spectrum disorder. *Genetics in Medicine, 22*(3), 538-546.

Mittal, K., Rafiq, M. A., Rafiullah, R., Harripaul, R., Ali, H., Ayaz, M., . . . So, J. (2016). Mutations in the genes for thyroglobulin and thyroid peroxidase cause thyroid dyshormonogenesis and autosomal-recessive intellectual disability. *Journal of Human Genetics*.

Mochida, G. H., Mahajnah, M., Hill, A. D., Basel-Vanagaite, L., Gleason, D., Hill, R. S., . . . Walsh, C. A. (2009). A truncating mutation of TRAPPC9 is associated with autosomal-recessive intellectual disability and postnatal microcephaly. *The American Journal of Human Genetics, 85*(6), 897-902.

Modell, B., & Darr, A. (2002). Genetic counselling and customary consanguineous marriage. *Nature Reviews Genetics, 3*(3), 225-229.

Musante, L., & Ropers, H. H. (2014). Genetics of recessive cognitive disorders. *Trends in Genetics, 30*(1), 32-39.

Myakhar, O., Unichenko, P., & Kirischuk, S. (2011). GABAergic projections from the subplate to Cajal–Retzius cells in the neocortex. *Neuroreport, 22*(11), 525-529.

Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S. S., Chen, W., . . . Jamali, P. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature, 478*(7367), 57-63.

Najmabadi, H., Motazacker, M. M., Garshasbi, M., Kahrizi, K., Tzschach, A., Chen, W., . . . Abedini, S. S. (2007). Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci. *Human genetics, 121*(1), 43-48.

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research, 31*(13), 3812-3814.

Noor, A., Whibley, A., Marshall, C. R., Gianakopoulos, P. J., Piton, A., Carson, A. R., . . . Pinto, D. (2010). Disruption at the PTCHD1 Locus on Xp22. 11 in Autism spectrum disorder and intellectual disability. *Science translational medicine, 2*(49), 49ra68-49ra68.

Noor, A., Windpassinger, C., Patel, M., Stachowiak, B., Mikhailov, A., Azam, M., . . . Paterson, A. D. (2008). CC2D2A, encoding a coiled-coil and C2 domain protein, causes autosomal-recessive mental retardation with retinitis pigmentosa. *The American Journal of Human Genetics, 82*(4), 1011-1018.

O'Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., . . . Smith, J. D. (2012). Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature, 485*(7397), 246-250.

Ofner, M., Coles, A., Decou, M., Do, M., Bienek, A., Snider, J., & Ugnat, A. (2018). *Autism spectrum disorder among children and youth in Canada 2018*: Public Health Agency of Canada Ottawa.

Omoyinmi, E., Melo Gomes, S., Standing, A., Rowczenio, D. M., Eleftheriou, D., Klein, N., . . . Brogan, P. A. (2014). Brief Report: Whole□Exome Sequencing Revealing Somatic NLRP3 Mosaicism in a Patient With Chronic Infantile Neurologic, Cutaneous, Articular Syndrome. *Arthritis & Rheumatology, 66*(1), 197-202.

Pagnamenta, A. T., Bacchelli, E., de Jonge, M. V., Mirza, G., Scerri, T. S., Minopoli, F., . . . Paracchini, S. (2010). Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia. *Biological psychiatry, 68*(4), 320-328.

Pang, A. W. C., MacDonald, J. R., Yuen, R. K., Hayes, V. M., & Scherer, S. W. (2014). Performance of high-throughput sequencing for the discovery of genetic variation across the complete size spectrum. *G3: Genes| Genomes| Genetics, 4*(1), 63-65.

Patel, S., Howard, D., Man, A., Schwartz, D., Jee, J., Felsky, D., . . . French, L. (2019). Transcriptomic analysis of brain hypometabolism links the cholinergic and gingipain hypotheses of Alzheimer's disease pathogenesis. *bioRxiv*.

Patwardhan, A., Harris, J., Leng, N., Bartha, G., Church, D. M., Luo, S., . . . Salit, M. (2015). Achieving high-sensitivity for clinical applications using augmented exome sequencing. *Genome medicine, 7*(1), 1.

Philippe, O., Rio, M., Carioux, A., Plaza, J.-M., Guigue, P., Molinari, F., . . . Smahi, A. (2009). Combination of linkage mapping and microarray-expression analysis identifies NF-κB signaling defect as a cause of autosomal-recessive mental retardation. *The American Journal of Human Genetics, 85*(6), 903-908.

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research, 35*(suppl 1), D61-D65.

Puffenberger, E. G., Jinks, R. N., Sougnez, C., Cibulskis, K., Willert, R. A., Achilly, N. P., . . . Lawrence, J. J. (2012). Genetic mapping and exome sequencing identify variants associated with five novel diseases. *PLoS One, 7*(1), e28936.

Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., . . . Kähler, A. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature, 506*(7487), 185-190.

Radnikow, G., Feldmeyer, D., & Lübke, J. (2002). Axonal projection, input and output synapses, and synaptic physiology of Cajal–Retzius cells in the developing rat neocortex. *Journal of Neuroscience, 22*(16), 6908-6919.

Rafiq, M. A., Kuss, A. W., Puettmann, L., Noor, A., Ramiah, A., Ali, G., . . . Garshasbi, M. (2011). Mutations in the alpha 1, 2-mannosidase gene, MAN1B1, cause autosomal-recessive intellectual disability. *The American Journal of Human Genetics, 89*(1), 176-182.

Rafiq, M. A., Leblond, C. S., Saqib, M. A., Vincent, A. K., Ambalavanan, A., Khan, F. S., . . . Ali, G. (2015). Novel VPS13B Mutations in Three Large Pakistani Cohen Syndrome Families Suggests a Baloch Variant with Autistic-Like Features. *BMC medical genetics, 16*(1), 41.

Rajab, A., Schuelke, M., Gill, E., Zwirner, A., Seifert, F., Gonzalez, S. M., & Knierim, E. (2015). Recessive DEAF1 mutation associates with autism, intellectual disability, basal ganglia dysfunction and epilepsy. *Journal of medical genetics, 52*(9), 607-611.

Rauch, A., Hoyer, J., Guth, S., Zweier, C., Kraus, C., Becker, C., . . . Rüschendorf, F. (2006). Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *American Journal of Medical Genetics Part A, 140*(19), 2063-2074.

Rehm, H. L., Bale, S. J., Bayrak-Toydemir, P., Berg, J. S., Brown, K. K., Deignan, J. L., . . . Lyon, E. (2013). ACMG clinical laboratory standards for next-generation sequencing. *Genetics in Medicine, 15*(9), 733-747.

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research, 47*(D1), D886-D894.

Reuter, M. S., Tawamie, H., Buchert, R., Gebril, O. H., Froukh, T., Thiel, C., . . . Zweier, C. (2017). Diagnostic yield and novel candidate genes by exome sequencing in 152 consanguineous families with neurodevelopmental disorders. *JAMA psychiatry, 74*(3), 293-299.

Reva, B., Antipin, Y., & Sander, C. (2010). Mutation Assessor. *Cancer*.

Riazuddin, S., Hussain, M., Razzaq, A., Iqbal, Z., Shahzad, M., Polla, D., . . . Tomas-Roca, L. (2017). Exome sequencing of Pakistani consanguineous families identifies 30 novel candidate genes for recessive intellectual disability. *Molecular Psychiatry, 22*(11), 1604-1614.

Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S. J., Zhong, Q., Sahni, N., . . . Mosca, R. (2014). A proteome-scale map of the human interactome network. *Cell, 159*(5), 1212-1226.

Ropers, H.-H., & Hamel, B. C. (2005). X-linked mental retardation. *Nature Reviews Genetics, 6*(1), 46-57.

Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., . . . Ayivi-Guedehoussou, N. (2005). Towards a proteome-scale map of the human protein–protein interaction network. *Nature, 437*(7062), 1173-1178.

Saadat, M. (2007). Consanguinity associated with child and adult mortality in 24 Asian and African countries, an ecological study. *Iranian Journal of Public Health, 36*, 35-39.

Safran, M., Dalah, I., Alexander, J., Rosen, N., Stein, T. I., Shmoish, M., . . . Krug, H. (2010). GeneCards Version 3: the human gene integrator. *Database, 2010*, baq020.

Sang, L., Miller, J. J., Corbit, K. C., Giles, R. H., Brauer, M. J., Otto, E. A., . . . Kwong, M. (2011). Mapping the NPHP-JBTS-MKS protein network reveals ciliopathy disease genes and pathways. *Cell, 145*(4), 513-528.

Schluth-Bolard, C., Labalme, A., Cordier, M. P., Till, M., Nadeau, G., Tevissen, H., . . . Sanlaville, D. (2013). Breakpoint mapping by next generation sequencing reveals causative gene disruption in patients carrying apparently balanced chromosome rearrangements with intellectual deficiency and/or congenital malformations. *J Med Genet, 50*(3), 144-150. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23315544. doi:10.1136/jmedgenet-2012-101351

Schmeisser, M. J., Ey, E., Wegener, S., Bockmann, J., Stempel, A. V., Kuebler, A., . . . Spilker, C. (2012). Autistic-like behaviours and hyperactivity in mice lacking ProSAP1/Shank2. *Nature, 486*(7402), 256-260.

Schwarz, J. M., Rödelsperger, C., Schuelke, M., & Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods, 7*(8), 575-576.

Scott, E. M., Halees, A., Itan, Y., Spencer, E. G., He, Y., Azab, M. A., . . . Abel, L. (2016). Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nature genetics*.

Seelow, D., Schuelke, M., Hildebrandt, F., & Nürnberg, P. (2009). HomozygosityMapper—an interactive approach to homozygosity mapping. *Nucleic acids research, 37*(suppl 2), W593-W599.

Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L., Edwards, K. J., . . . Gaunt, T. R. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human mutation, 34*(1), 57-65.

Singh, T., Kurki, M. I., Curtis, D., Purcell, S. M., Crooks, L., McRae, J., . . . Breen, G. (2016). Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nature neuroscience*.

Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shiel, J. A., Thomas, N. S., . . . Cooper, D. N. (2003). Human gene mutation database (HGMD®): 2003 update. *Human mutation, 21*(6), 577-581.

Stevens, S. J., van Ravenswaaij□Arts, C. M., Janssen, J. W., Klein Wassink□Ruiter, J. S., van Essen, A. J., Dijkhuizen, T., . . . Smeets, E. E. (2011). MYT1L is a candidate gene for intellectual disability in patients with 2p25. 3 (2pter) deletions. *American Journal of Medical Genetics Part A, 155*(11), 2739-2745.

Study, D. D. D., McRae, J. F., Clayton, S., Fitzgerald, T. W., Kaplanis, J., Prigmore, E., . . . Akawi, N. J. N. (2017). Prevalence and architecture of *de novo* mutations in developmental disorders. *542*(7642), 433-438.

Swaab, D. F. (2007). Sexual differentiation of the brain and behavior. *Best practice & research clinical endocrinology & metabolism, 21*(3), 431-444.

Tan, C., Topper, S., Del Gaudio, D., Nelakuditi, V., Shchelochkov, O., Nowaczyk, M., . . . Meeks, N. (2015). Characterization of patients referred for non□specific intellectual disability testing: the importance of autosomal genes for diagnosis. *Clinical genetics*.

Tarpey, P. S., Smith, R., Pleasance, E., Whibley, A., Edkins, S., Hardy, C., . . . Menzies, A. (2009). A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nature genetics, 41*(5), 535-543.

Teng, B. L., Nikolova, V. D., Riddick, N. V., Agster, K. L., Crowley, J. J., Baker, L. K., . . . Moy, S. S. (2016). Reversal of social deficits by subchronic oxytocin in two autism mouse models. *Neuropharmacology, 105*, 61-71.

Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J. J., Moser, M.-B., & Moser, E. I. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature, 561*(7721), 57-62.

Tzschach, A., Grasshoff, U., Beck-Woedl, S., Dufke, C., Bauer, C., Kehrer, M., . . . Di Donato, N. (2015). Next-generation sequencing in X-linked intellectual disability. *European Journal of Human Genetics*.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy□Moonshine, A., . . . Thibault, J. (2013). From FastQ data to high□confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics, 43*(1), 11.10. 11-11.10. 33.

van der Vaart, B., Franker, M. A., Kuijpers, M., Hua, S., Bouchet, B. P., Jiang, K., . . . Akhmanova, A. (2012). Microtubule plus-end tracking proteins SLAIN1/2 and ch-TOG promote axonal development. *The Journal of Neuroscience, 32*(42), 14722-14728a.

Van Scherpenzeel, M., Timal, S., Rymen, D., Hoischen, A., Wuhrer, M., Hipgrave-Ederveen, A., . . . Edvardson, S. (2014). Diagnostic serum glycosylation profile in patients with intellectual disability as a result of MAN1B1 deficiency. *Brain, 137*(4), 1030-1038.

Vissers, L. E., Gilissen, C., & Veltman, J. A. (2015). Genetic studies in intellectual disability and related disorders. *Nature Reviews Genetics*.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F., . . . Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome research, 17*(11), 1665-1674.

Wang, T., Guo, H., Xiong, B., Stessman, H. A., Wu, H., Coe, B. P., . . . Hoekzema, K. J. N. c. (2016). *De novo* genic mutations among a Chinese autism spectrum disorder cohort. *7*, 13316.

Ware, J. S., Samocha, K. E., Homsy, J., & Daly, M. J. J. C. p. i. h. g. (2015). Interpreting *de novo* variation in human disease using denovolyzeR. *87*(1), 7.25. 21-27.25. 15.

Weiss, L., Escayg, A., Kearney, J., Trudeau, M., MacDonald, B., Mori, M., . . . Meisler, M. (2003). Sodium channels SCN1A, SCN2A and SCN3A in familial autism. *Molecular Psychiatry, 8*(2), 186-194.

Wolf, E., Kim, P. S., & Berger, B. J. P. S. (1997). MultiCoil: a program for predicting two□and three□stranded coiled coils. *6*(6), 1179-1189.

Wright, A. (1997). Limbic system: amygdala. *Neuroscience Online. http://neuroscience. uth. tmc. edu/s4/chapter06. html*.

Wright, C. F., Fitzgerald, T. W., Jones, W. D., Clayton, S., McRae, J. F., Van Kogelenberg, M., . . . Bayzetinova, T. (2015). Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *The Lancet, 385*(9975), 1305-1314.

Wynshaw□Boris, A. (2007). Lissencephaly and LIS1: insights into the molecular mechanisms of neuronal migration and development. *Clinical genetics, 72*(4), 296-304.

Yang, H., & Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature protocols, 10*(10), 1556-1566.

Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A., . . . Niu, Z. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *New England Journal of Medicine, 369*(16), 1502-1511.

Yang, Y., Muzny, D. M., Xia, F., Niu, Z., Person, R., Ding, Y., . . . Buhay, C. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. *Jama, 312*(18), 1870-1879.

Yu, H., Tardivo, L., Tam, S., Weiner, E., Gebreab, F., Fan, C., . . . Yang, X. (2011). Next-generation sequencing to generate interactome datasets. *Nature methods, 8*(6), 478-480.

Zhang, R.-R., Cui, Q.-Y., Murai, K., Lim, Y. C., Smith, Z. D., Jin, S., . . . Wu, H.-P. (2013). Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell, 13*(2), 237-245.

Zhubi, A., Chen, Y., Dong, E., Cook, E., Guidotti, A., & Grayson, D. (2014). Increased binding of MeCP2 to the GAD1 and RELN promoters may be mediated by an enrichment of 5-hmC in autism spectrum disorder (ASD) cerebellum. *Translational psychiatry, 4*(1), e349.

Zyla, J., Marczyk, M., Domaszewska, T., Kaufmann, S. H., Polanska, J., & Weiner 3rd, J. (2019). Gene set enrichment for reproducible science: comparison of CERNO and eight other algorithms. *Bioinformatics, 35*(24), 5146-5154.

7 Appendices

7.1 Supplementary Material for Chapter 2

7.1.1 Methods

7.1.1.1 Family Recruitment

Families were recruited from the following sites, following local REB approval and written consent: COMSATS Institute of Information Technology, Islamabad, Pakistan (N=8 families), National University of Sciences and Technology (NUST), Islamabad, Pakistan (N=14), International Islamic University, Islamabad, Pakistan (N=43), Quaid-i-Azam University, Islamabad, Pakistan (N=30), Lahore Institute of Research and Development, Lahore, Pakistan (N=81), Qazvin University of Medical Sciences, Qazvin, Iran (N=10), and Pasteur Institute of Iran, Tehran, Iran (N=6).

7.1.1.2 HBD/Autozygosity mapping

Both of the below methods and the hg19 version of the genome were used to ensure a consistent and uniform genotyping.

Genotyping data was uploaded to the HomozygosityMapper server to determine putative homozygous-by-descent (HBD) regions based on the allele frequencies of the markers uploaded to the server from previous studies. HBD regions were identified by manual curation and only HBD regions larger than 1 Mb shared between all affected members (and not unaffected members) of the family were chosen. These regions were extracted based on SNP RS numbers and these dbSNP identifiers were converted to a genomic position for used to represent genomic regions with NGS data.

Previous methods used to identify HBD regions can vary among individual researchers due to the manual curation necessary to ensure the selected regions share both homozygosity and haploidentity between affecteds but not unaffecteds (i.e. to exclude false positives), and to try to limit false negative calls. These methods are also labor intensive and do not automatically take in family information. For these reasons, we used a software called FSuite for computational HBD mapping(S. Gazal et al., 2014). FSuite makes use of allele frequencies, genotyping and

pedigree structure integrated into a Hidden Markov model to determine the likelihood of a region being HBD(S. Gazal et al., 2014). In-house developed scripts were used to convert genotyping calls from the Affymetrix Genotyping Console, ChAS and Illumina GenomeStudio output to a family based PLINK input for FSuite. We also used 1000 Genomes Phase Three reference for the allele frequencies, extracting allele frequencies from the South Asian ethnic group specifically. This data was then used with standard run protocols of FSuite to generate HBD regions. Autozygosity mapping was done by extracting regions common between members of the family. These protocols were standardized and implemented on the Specialized Computing Cluster (SCC) at the Centre for Addiction and Mental Health (CAMH).

7.1.1.3 CNV Analysis

CNV analysis was performed to identify homozygous CNVs in HBD regions, or heterozygous/ homozygous CNVs that could indicate phenocopies that should then be excluded from the HBD/WES analyses. We have used five array types for homozygosity mapping including the Affymetrix SNP 5.0, SNP 6.0, CytoscanHD and Mapping 250K NspI, as well the Illumina HumanCoreExome platform. Due to the wide range of array types that were used in this study, we have used manufacturer-recommended and third-party software to call CNVs. The Affymetrix Genotyping Console was used to genotype and calls CNVs using the Affymetrix Birdsuite algorithm. CytoscanHD and SNP 6.0 arrays were analyzed and CNVs were called using the Chromosome Analysis Suite (ChAS). Illumina HumanCore Exome arrays were used for the largest proportion of families, and CNVs were called using the Illumina GenomeStudio cnvPartition plugin. PennCNV is an academically developed software that uses a Hidden Markov Method (HMM) to call CNVs(K. Wang et al., 2007). CNVs were called using PennCNV for all array types and represents a standard method for CNV calling across our study. CNVs derived from all algorithms were compared against the Database of Genomic Variants (DGV) to screen for variants that showed 30% or lower overlap with DGV control variants, or were not in the database(MacDonald et al., 2013).

7.1.1.4 Sequencing Alignment and Variant Calling

Illumina reads were aligned with BWA-mem (v 0.7.13) using the hg19 genome reference(H. Li, 2013). Once alignment was performed GATK (v 3.2.2) was used for base recalibration, indel

realignment and the GATK Unified Genotyper was used for variant calling. Variants called were generated in the standard VCF version 4.1 and annotated with Annovar(H. Yang & Wang, 2015), Alamut (v 1.4.4 Interactive Biosoftware, Rouen, France) and EDGC Annotator (v 1.0.0 Eone-Diagnomics Genome Center, Songdo Ichon, South Korea) software.

SOLiD sequencing was aligned to the hg19 reference genome using Shrimp2(David et al., 2011). Once alignment files were generated, variant calling and annotation were performed as previously described. Proton data was aligned using the TMAP algorithm and variants were called using the Ion Proton Variant Caller as part of the Torrent Server analysis pipeline (ThermoFisher Scientific, Waltham, MA, US). Ion Reporter was used in addition to the annotation software mentioned above for Ion Proton VCF files (ThermoFisher Scientific, Waltham, MA, US).

7.1.1.5 Variant Filtration

Once annotation was performed, filtration of variants was performed by computationally and manually extracting variants that were in HBD regions provided by the autozygosity mapping of FSuite and HomozygosityMapper. The next step was to filter variants based on sequencing quality, homozygosity of allele and mutation type. Prioritization was performed by scoring truncating mutations such as stop and frameshift loss of function mutations higher than missense mutations and inframe indels. Missense mutations were scored based on Sift and Polyphen scores for prioritization. Allele frequencies were then used to filter out variants that were too common in the population (>1 in 10 000). All variants surviving filtration were checked against the Exome Aggregation Consortium (ExAC), Cambridge, MA (URL: http://exac.broadinstitute.org) accessed 03,2016) OMIM(Hamosh et al., 2005) and Genecards(Safran et al., 2010) to determine expression level in tissues and other functional characteristics.

7.1.1.6 Anatomical expression analyses

To look for commonalities in the anatomical and temporal expression profiles for the genes identified here and previously reported genes for ID we used a number of human expression datasets, tools, and atlases.

The TopAnat Bgee tool was used to quantify tissue specific expression patterns before birth in human expression datasets (Bastian et al., 2008).TopAnat uses publically available expression profiles from reference atlases of normal tissue. Within a tissue, it tests for expression enrichment using binary calls of expression (expressed/present versus not). The ID Gene list was converted to Ensembl identifiers using the Ensembl Biomart tool (Ensembl release 84). The embryo developmental stage, RNA-Seq and Affymetrix data source settings were used with the "Data quality" option was set to "All". We removed redundant tissue terms and only present data from tissues with direct expression information (17 human fetal tissues).

7.1.1.7 Developmental brain expression analyses

Human spatiotemporal transcriptomic datasets were downloaded from the BrainSpan website (http://www.brainspan.org/static/download.html, December 2015). In this developmental atlas, age ranged from 8 post conception weeks to 40 years old (46% female). Brain specimens were from normal healthy donors that passed several strict exclusion criteria to ensure consistency. For example, specimens with evidence of malformations, lesions, neuronal loss, neuronal swelling, glioneuronal heterotopias, or dysmorphic neurons and neurites were excluded. Complete details of the methods used in the BrainSpan atlas are available as a technical paper (http://help.brain-map.org/download/attachments/3506181/Transcriptome_Profiling.pdf).

Both RNA sequencing and exon microarray gene expression profiles were used to test for specific temporal and regional expression of the ID genes. In both datasets we removed data from brain regions with expression profiles for less than 7 donors (10 regions). These regions are early structures that were dissected at early timepoints (earlier than 14 post conception weeks). After this filtering, 465 expression profiles from 35 brains (27 timepoints) in 16 unique regions were used from the exon array data. In the RNA sequencing data, 553 expression profiles from 41 brains (30 timepoints) in 16 unique regions remained. Expression data was previously summarized to genes by the BrainSpan consortium (exon array: 22,255, RNA sequencing: 17,282 genes). The provided RPKM (Reads Per Kilobase of transcript per Million mapped reads) expression values were log scaled.

Human prefrontal cortex expression profiles(Colantuoni et al., 2011) were downloaded from the supplemental files in "Mapping DNA methylation across development, genotype and

schizophrenia in the human frontal cortex"(Jaffe et al., 2015). This microarray dataset contains gene expression profiles of the dorsolateral prefrontal cortex across development. Provided mappings were used to link probes to genes (31,699 probes). Expression was mean averaged across multiple probes for a given gene. We limited our analyses to donor brains under 5 years old at death (supported by our results in the BrainSpan atlas of the developing brain). After this processing, 67 samples (34 timepoints) with information for 18419 genes were used in our analyses.

Expression values were z-score normalized across samples in each of the above datasets. Specifically, for each gene, its expression values were scaled (mean=0, standard deviation=1). This allows identification of region and timepoint specific expression patterns that are independent of absolute expression levels.

7.1.1.8 Pathway analyses

We used the Gene Ontology enRIchment anaLysis and visuaLizAtion tool (GOrilla) to test for enrichment of Gene Ontology (GO) gene groups in the ID gene list(Ashburner et al., 2000; Eden et al., 2009). All HGNC gene symbols were used as the background gene list and default options were used (biological process ontology). The GOrilla database was last updated on May 14, 2016 when we ran the analyses.

7.1.1.9 References

1. Gazal S, Sahbatou M, Babron MC, Genin E, Leutenegger AL. FSuite: exploiting inbreeding in dense SNP chip and exome data. *Bioinformatics.* 2014;30(13):1940-1941.

2. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome research.* 2007;17(11):1665-1674.

3. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic acids research.* 2013:gkt958.

4. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997.* 2013.

5. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature protocols.* 2015;10(10):1556-1566.

6. David M, Dzamba M, Lister D, Ilie L, Brudno M. SHRiMP2: sensitive yet practical short read mapping. *Bioinformatics.* 2011;27(7):1011-1012.

7. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic acids research.* 2005;33(suppl 1):D514-D517.

8. Safran M, Dalah I, Alexander J, et al. GeneCards Version 3: the human gene integrator. *Database.* 2010;2010:baq020.

9. Bastian F, Parmentier G, Roux J, Moretti S, Laudet V, Robinson-Rechavi M. Bgee: integrating and comparing heterogeneous transcriptome data among species. Paper presented at: Data Integration in the Life Sciences2008.

10. Colantuoni C, Lipska BK, Ye T, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature.* 2011;478(7370):519-523.

11. Jaffe AE, Gao Y, Deep-Soboslay A, et al. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nature neuroscience.* 2015.

12. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC bioinformatics.* 2009;10(1):48.

13. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nature genetics.* 2000;25(1):25-29.

7.1.2 Supplemental Figures and Tables

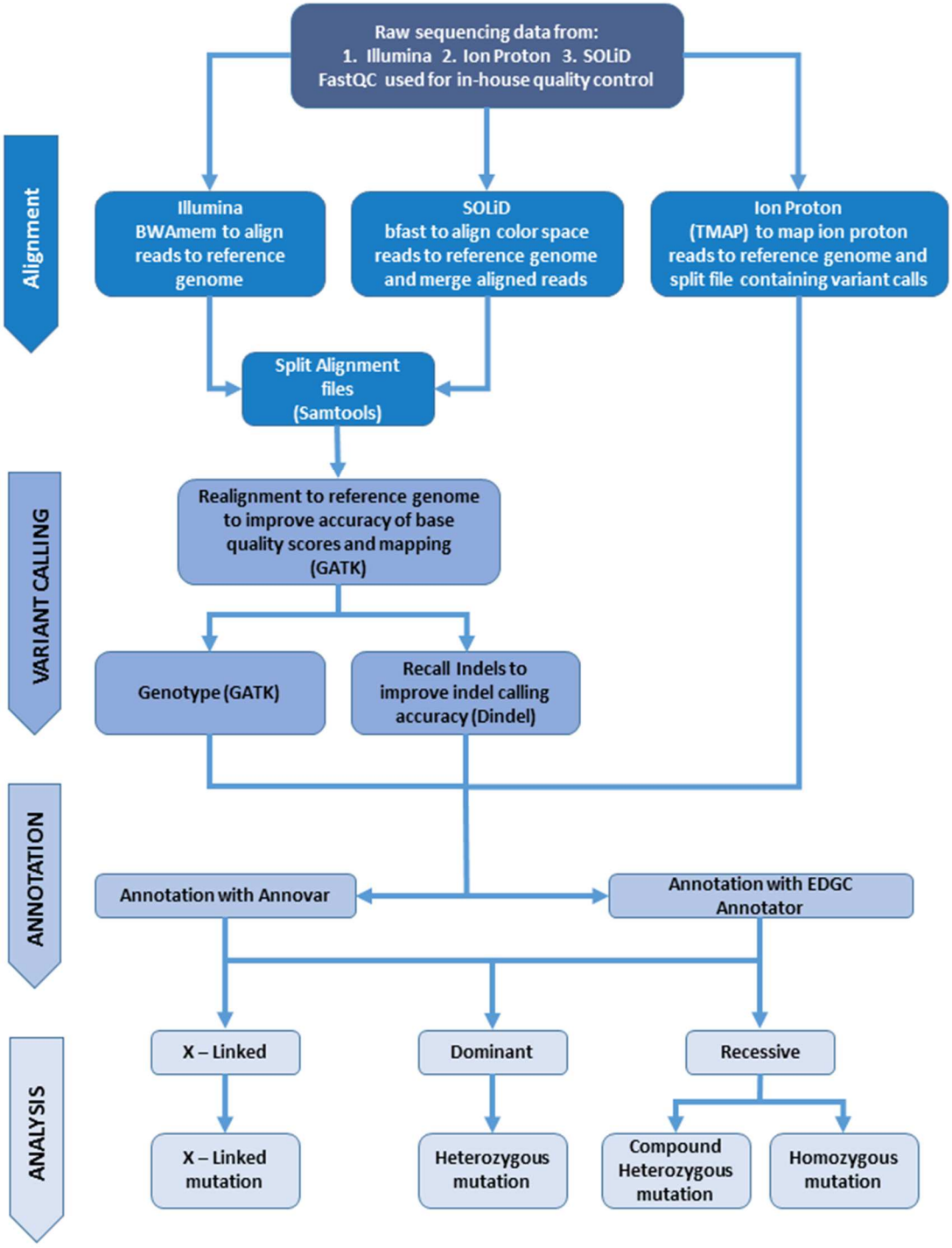
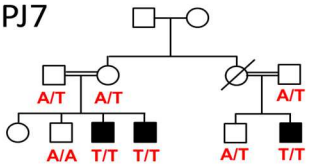
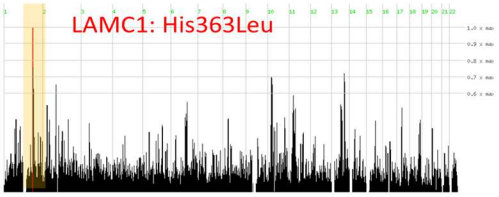


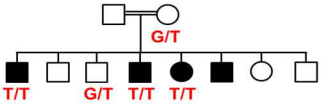
Figure A- 1: Bioinformatic pipeline for whole exome sequence (WES) analysis.

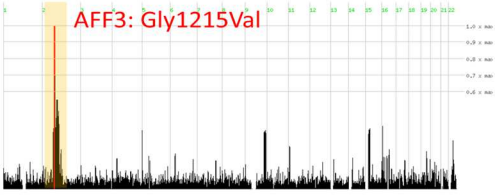
PJ7



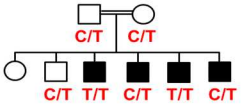


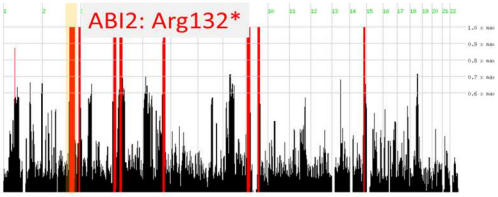
PK113



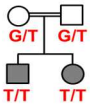


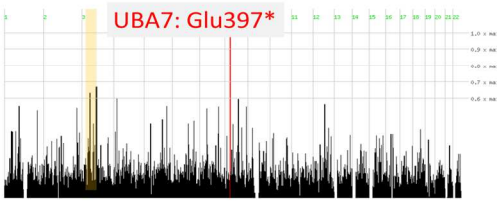
AS61[a]



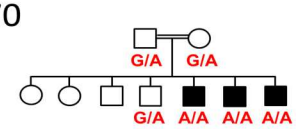


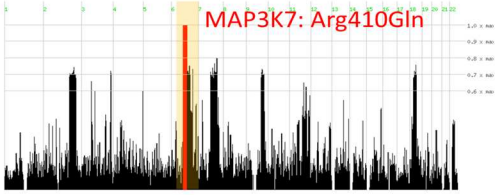
PK34[b]



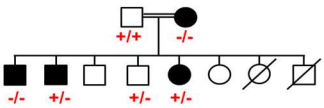


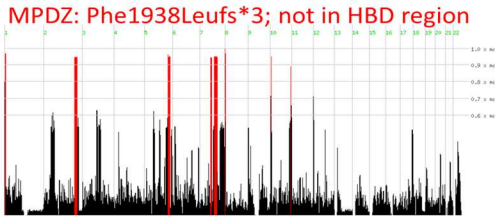
AS70



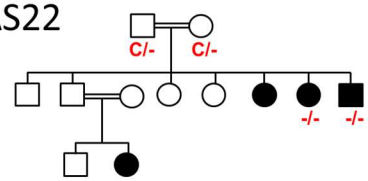


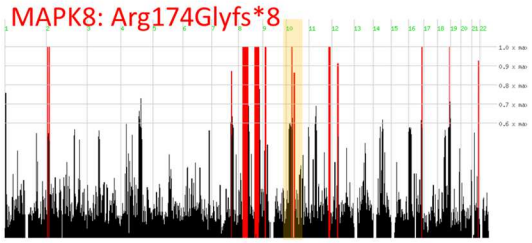
AS66[c]



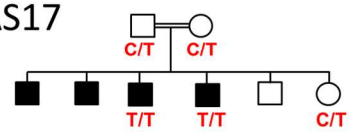


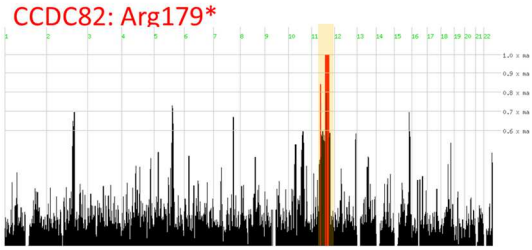
AS22



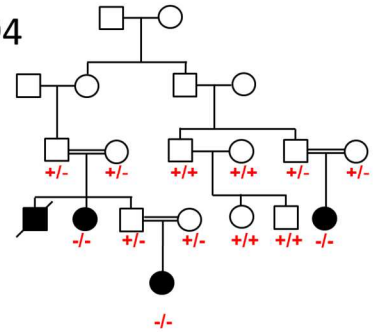


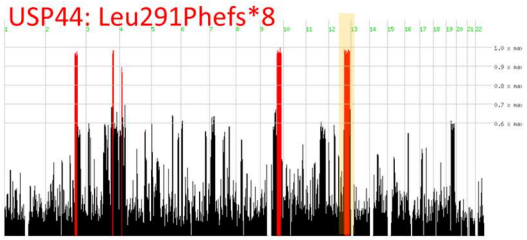
AS17



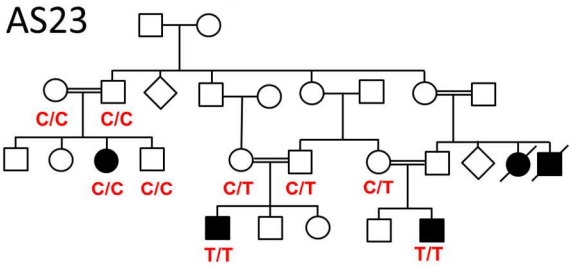


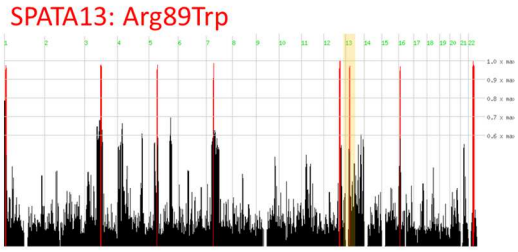
PK94



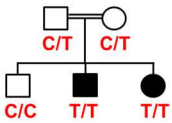


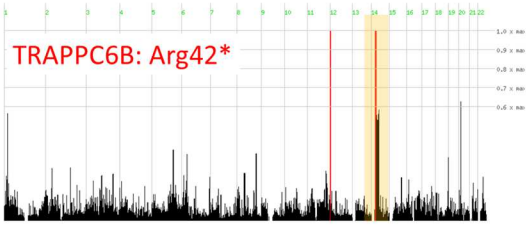
AS23



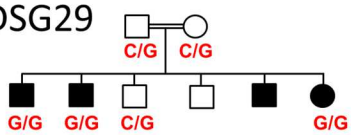


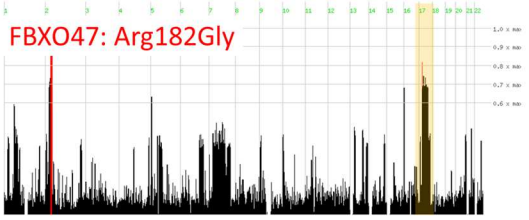
IDH3





IDSG29





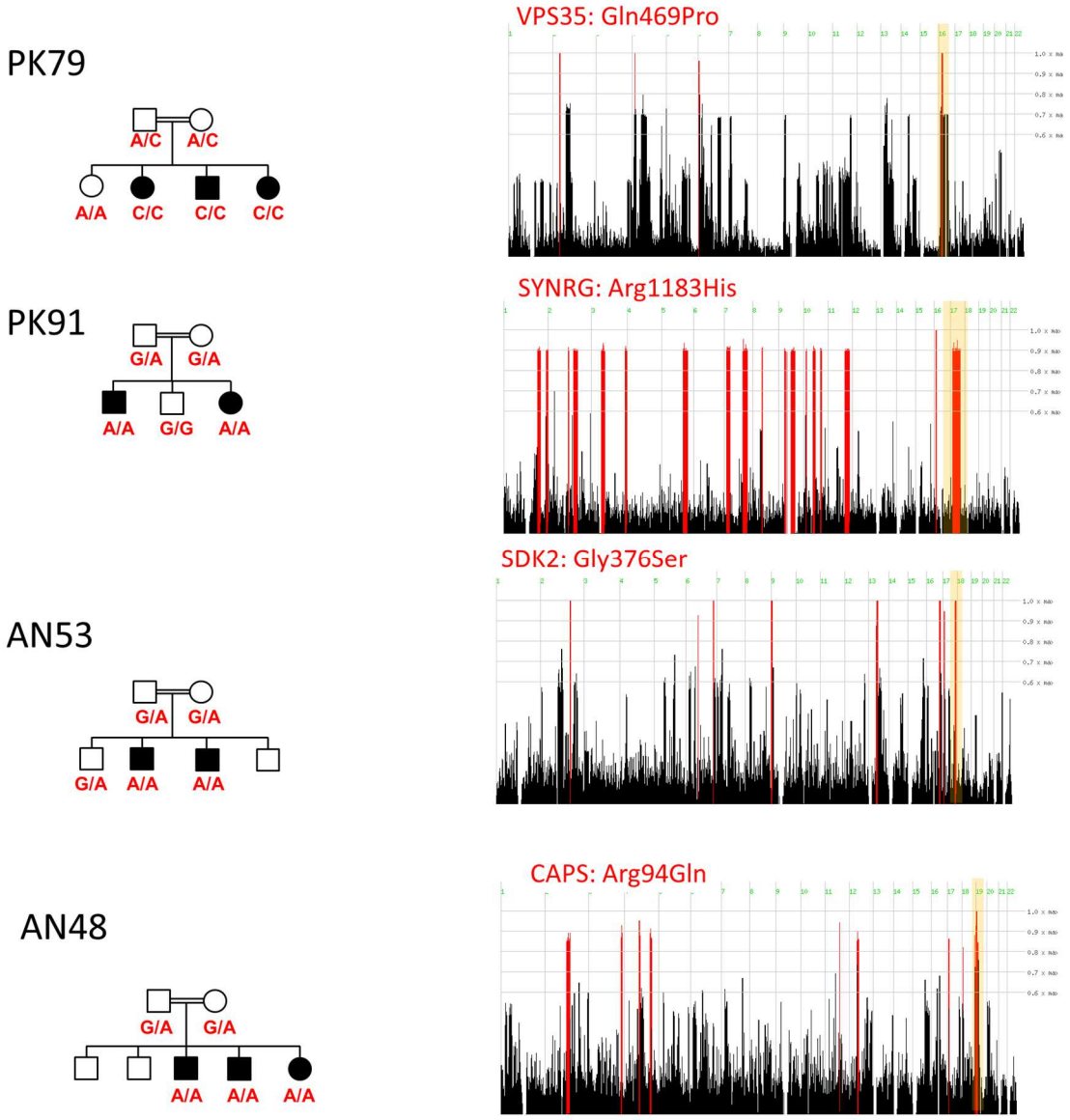


Figure A- 2: Pedigrees and HomozygosityMapper output for families with single variants identified, in addition to those shown in Figure 2.2 Pedigrees and HomozygosityMapper output for eight of the families

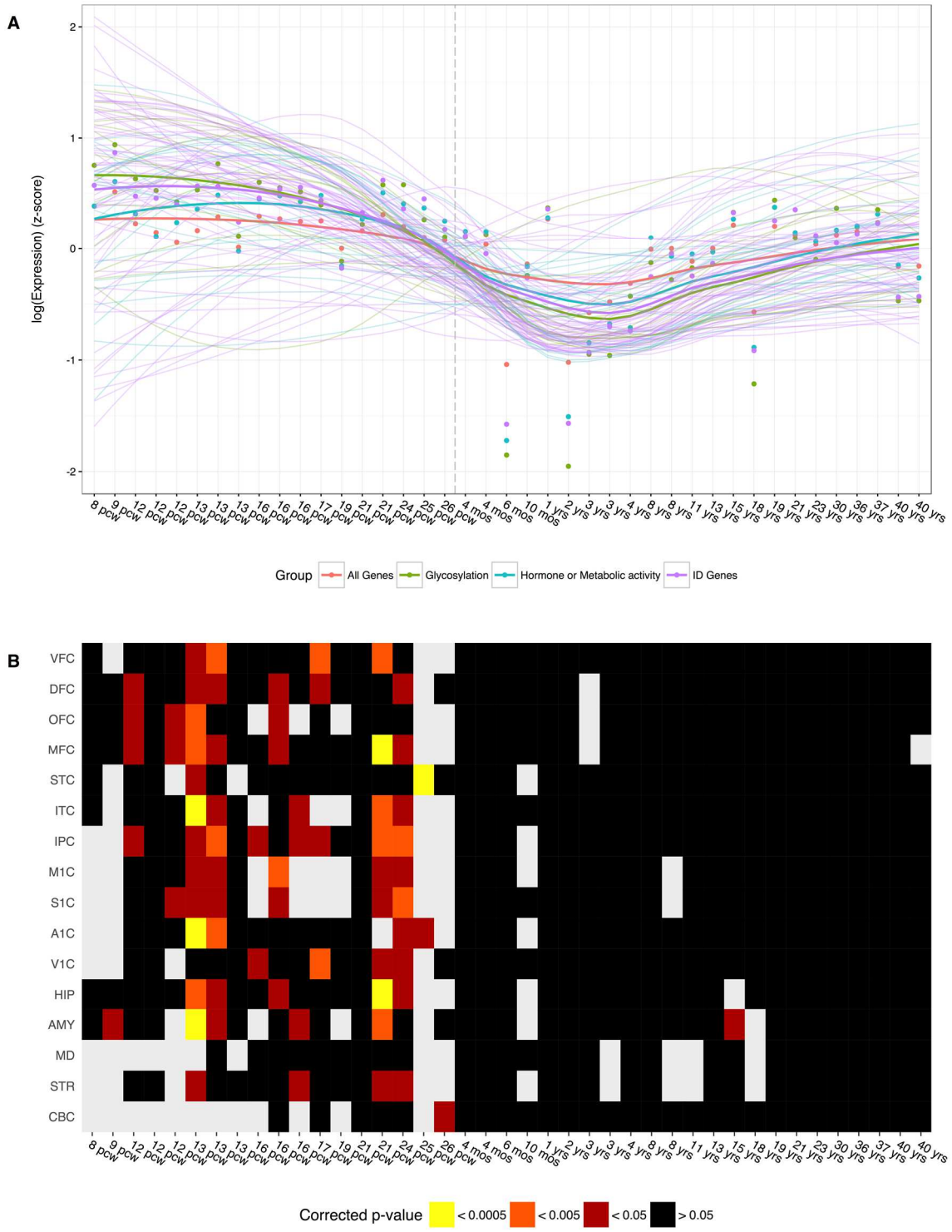


Figure A- 3: Spatiotemporal expression of ID genes in human development using RNA sequencing data

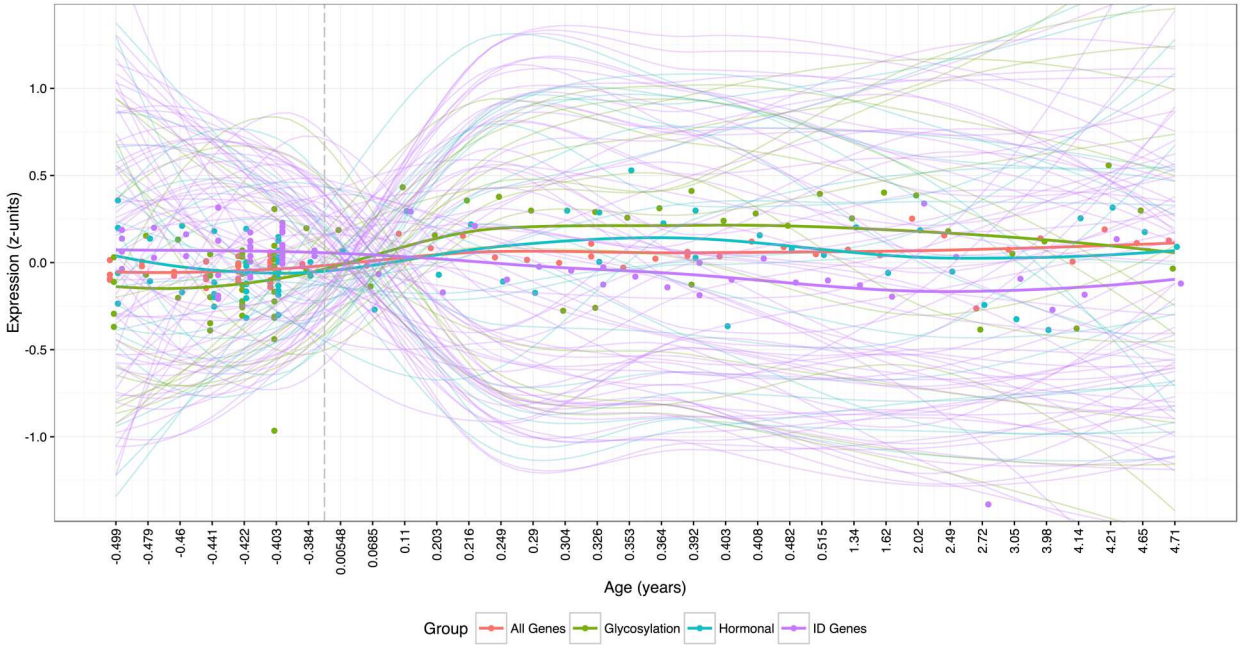


Figure A- 4: Developmental expression pattern of ID genes in the human prefrontal cortex.

Table A- 1 Family statistics denoting the number of affected and unaffected individuals as well as the f coefficient denoting the degree of consanguinity. The total number of individuals in the family are also listed.

Number	Family	Number of Members	Number of Affected	Number of Branches	*f* Coefficient
1	IDSG19	7	6	1	0.01
2	IDSG29	19	4	1	0.01
3	IDSG30	25	5	2	0.01
4	IDSG41	37	6	2	0.01
5	IDSG43	31	8	3	0.01
6	IDSG49	32	10	2	0.10464
7	IDH1	7	4	1	0.10072
8	IDH2	8	3	1	0.08025
9	IDH3	5	2	1	0.042333
10	IDH4	6	3	1	0.10505
11	IDH5	6	3	1	0.1417
12	IDH6	5	3	1	0.119667
13	IDH7	4	3	1	0.0511
14	IDH8	4	2	1	0.0983
15	IDH9	4	2	1	0.073533
16	IDH10	11	6	1	N/A
17	PK1	46	5	3	1
18	PK7	15	4	1	0.112333
19	PK8	42	10	3	1

20	PK9	7	4	1	0.03159
21	PK10	4	2	1	0.0933
22	PK11	15	3	1	0.073867
23	PK12	17	4	2	0.060067
24	PK13	10	2	2	0.054933
25	PK17	17	7	1	0.094831
26	PK18	13	4	2	0.01
27	PK19	7	4	1	0.01
28	PK22	4	2	1	0.01
29	PK25	4	2	1	0.01
30	PK26	5	2	1	0.085
31	PK27	7	4	1	0.03232
32	PK28	14	4	2	0.00112
33	PK29	5	2	1	0.0018
34	PK30	9	2	2	0.064333
35	PK31	5	2	1	0.077933
36	PK33	4	1	1	0.030067
37	PK34	4	2	1	0
38	PK35	11	4	3	0.0789
39	PK36	5	2	1	0.01
40	PK37	5	2	1	0.1648
41	PK38	5	2	1	0
42	PK39	20	8	5	0.01

43	PK40	5	2	2	0.01
44	PK42	5	2	1	0.01
45	PK46	8	5	2	0.01
46	PK48	9	5	1	0.004133
47	PK49	5	2	1	0.1666
48	PK50	9	3	3	0.01
49	PK51	5	2	1	0.1048
50	PK52	18	5	5	0.01
51	PK53	5	2	1	0.01
52	PK54	10	6	2	0.1799
53	PK55	11	3	2	0.08918
54	PK57	4	2	1	0.096933
55	PK58	6	2	2	0.073933
56	PK59	5	2	1	0.000933
57	PK60	11	6	3	0.042175
58	PK61	5	2	1	0.08
59	PK62	17	9	8	0.121052
60	PK63	8	4	2	0.036467
61	PK64	5	2	1	0.0946
62	PK65	5	2	1	0.129267
63	PK67	8	3	1	0.0584
64	PK68	5	3	1	0.111467
65	PK69	5	2	1	0.038667

66	PK70	3	2	1	0.1016
67	PK71	8	2	2	0.078533
68	PK72	6	3	1	0.0871
69	PK73	13	7	2	0.113505
70	PK74	9	5	3	0.120517
71	PK76	5	2	1	0.033667
72	PK77	8	5	1	0.1496
73	PK78	7	2	1	0.122467
74	PK79	6	3	1	0.15845
75	PK80	6	2	2	0.077333
76	PK81	4	2	1	0.0264
77	PK83	4	2	1	0.0387
78	PK85	5	3	1	0.083733
79	PK87	9	2	2	0.096867
80	PK91	5	2	1	0.0978
81	PK93	6	3	1	0.19255
82	PK94	13	3	3	0.087405
83	PK95	9	5	2	0.117167
84	PK98	6	5	1	0.07588
85	PK100	6	2	2	0.014822
86	PK101	4	2	1	0.084133
87	PK102	5	3	1	0.2174
88	PK108	7	6	2	0.091475

89	PK111	4	2	1	0.0432
90	PK113	5	4	1	0.05655
91	PK117	1	2	1	0.004417
92	PK121	1	2	1	0.064333
93	PK123	5	2	1	0.0511
94	PK127	1	2	1	0.00805
95	PK135	1	2	1	0.2068
96	PK138	4	2	1	0.025267
97	PK144	8	3	2	0.021931
98	IAID1	5	2	1	0.152733
99	IAID2	10	4	1	0.072533
100	IAID3	5	2	1	0.076667
101	PJ1	11	4	3	0.13685
102	PJ2	6	3	2	0.1869
103	PJ3	10	4	3	0.098983
104	PJ5	16	4	2	1
105	PJ6	11	6	3	1
106	PJ7	8	3	3	1
107	PJ8	11	5	3	1
108	PJ9	12	5	2	0.111133
109	PJ10	10	4	3	1
110	PJ11	4	2	2	1
111	PJ12	7	2	2	1

112	RQ1	8	3	1	0.09352
113	RQ2	18	4	1	0.0943
114	RQ6	7	4	2	0.08608
115	RQ7	10	3	3	1
116	RQ8	4	2	2	1
117	RQ9	5	2	2	1
118	RQ10	7	4	2	0.083475
119	RQ11	7	3	2	0.12222
120	AN13	14	5	4	0.037
121	AN14	8	3	2	0.03285
122	AN16	26	8	5	0.07236
123	AN17	11	5	1	0.110667
124	AN18	8	8	2	0.11828
125	AN19	7	3	2	0.1601
126	AN20	8	4	2	0.07392
127	AN21	8	4	2	0.09636
128	AN22	9	3	3	0.0724
129	AN24	5	3	1	0.0503
130	AN27	7	3	2	0.146
131	AN28	9	3	3	0.069067
132	AN30	13	3	5	0.07348
133	AN37	10	4	3	0.06908
134	AN43	8	3	2	0.01

135	AN45	11	4	4	0.0606
136	AN48	7	3	2	0.0686
137	AN49	5	3	2	0.07245
138	AN50	14	5	4	0.0704
139	AN51	20	9	4	0.09186
140	AN52	7	3	1	0.09825
141	AN53	6	2	1	0.077933
142	AS3	11	5	4	0.56195
143	AS5	5	2	2	0.073933
144	AS6	2	3	2	0.0105
145	AS7	5	2	1	0.086667
146	AS8	4	3	2	0.03155
147	AS11	11	4	3	1
148	AS15	4	3	1	0.1164
149	AS16	7	4	3	1
150	AS17	7	4	2	0.01
151	AS18	6	3	1	1
152	AS19	20	7	3	0.044225
153	AS20	5	3	1	0.15012
154	AS21	5	2	3	0.01
155	AS22	5	2	3	0.01
156	AS23	8	3	3	1
157	AS25	5	3	3	0.1049

158	AS26	6	4	5	0.01
159	AS27	5	4	1	0.146
160	AS28	11	6	1	0.069067
161	AS30	14	6	5	0.07348
162	AS55	8	2	3	0
163	AS56	5	4	1	0.1182
164	AS58	10	3	3	0.10125
165	AS61	8	4	2	0.10056
166	AS62	10	5	4	0.1248
167	AS64	5	3	2	0.0457
168	AS66	6	4	2	0.06772
169	AS68	6	3	2	0.01
170	AS70	6	3	2	0.059133
171	AS72	11	5	3	0.167833
172	AS74	11	4	4	0.1511
173	AS101	6	3	1	0.0429
174	AS102	5	3	1	0.086933
175	AS103	12	6	4	0.1451
176	AS104	5	3	1	0
177	AS105	6	2	1	0.039733
178	AS107	4	2	1	0.08335
179	AS109	6	4	1	0.07496
180	AS110	13	8	2	0.0438

181	AS111	6	4	2	0.03368
182	AS112	10	6	2	0.020333
183	AS113	6	2	2	0
184	AS114	9	5	2	0.0349
185	ZA5	28	6	3	N/A
186	ZA7	23	4	1	0.06005
187	ZA8	4	3	1	0.03815
188	ZA9	6	2	1	0.009667
189	ZA10	6	2	1	0.018533
190	ZA12	44	8	1	0.031933
191	ZA16	28	5	3	0.15895
192	ZA17	30	10	3	N/A

Table A- 2: Regions selected by microarray platform. Regions were determined as HBD based on shared homozygous haplotype using HomozygosityMapper analysis cross-referenced with FSuite analysis. Although all regions were explored, only HBD regions ≥1.0Mb in size are reported here.

Family #	Family ID	X-linkage compatible ?	Chr	start SNP	end SNP	start bp	end bp	size	HM Score	Genes	Number of Family Members Tested	Number of Affected Tested (Homozygous for Variant)	CNVs
Illumina Core Exome													
1	PK9	N	No HBD										
2	PK12	Y	8	rs17074954	rs435904	5679269	31789101	26.11	101				
			16	rs9923480	rs1420585	31,544,369	48,995,896	17.452	169	**GPT2**	17	7 (7)	
			12	rs1523118	rs7964150	39,229,699	40,388,337	1.159	38				
			12	rs1635544	rs7138803	48,380,031	50,247,468	1.867	115				
3	PK17	N	No HBD										PK17-45 has partial 7p monosomy
4	PK31	N	5	rs10514237	rs348900	81,615,350	83,406,683	1.791	370				
			6	rs9389212	rs9399178	135,051,407	136,350,699	1.299	428				
			11	rs11246002	rs3861787	211,447	4,647,770	4.436	500				

			11	rs679333	rs6483079	88,642,586	90,652,932	2.01	394				
			17	rs513643	rs4313838	72,599,418	77,698,582	5.099	480	**METTL23**	5	2(2)	
			20	rs3920525	rs6104958	11,134,907	12,707,998	1.573	500				
5	PK35	N	7	rs1601987	rs2525734	118,919,137	120,624,165	1.705	324				
			19	rs2361019	rs2217662	22,729,613	23,743,106	1.013	344				
6	PK37	N	No HBD										
7	PK40	N	3	rs1512909	rs9847884	89,336,417	94,031,084	4.694	182				
			6	rs9470094	rs9366917	35,752,792	36,790,666	1.038	241				
			20	rs2387578	rs845787	25,096,494	26,197,931	1.101	137				
8	PK42	Y	2	rs4280394	rs13016931	192,704,699	207,914,021	15.209	500				
			22	rs5755996	rs4822379	22,257,652	23,645,718	1.388	482				
9	PK48	N	No HBD										
10	PK55	N	No HBD										
			X							**MECP2**	11	5 (5 hemizygous/heterozygous)	
11	PK58	N	1	rs16833075	rs6587761	236,157,277	249,198,060	13.041	250				PK58-5 removed (poor

													microarray)
			2	rs13423995	rs9309559	33,012	27,141,515	27.109	250				
			2	rs7423968	rs4852279	59,467,016	72,264,864	12.798	250				
			3	rs6782243	rs6807270	143,409,272	148,587,489	5.178	250				
			5	rs17631488	rs10051091	18,777,746	29,028,901	10.251	250				
			5	rs1979980	rs10062617	98,107,067	109,517,541	11.41	250				
			6	rs4441992	rs13218440	6,559,568	12,059,954	5.5	250				
			6	rs389674	rs17643290	147,813,016	148,869,370	1.056	250				
			7	rs6953751	rs1985369	155,377,430	159,119,220	3.742	250				
			10	rs10905155	rs1762757	7,411,558	9,409,770	1.998	250				
			11	rs175115	rs2510282	60,825,214	82,522,085	21.697	250				
			11	rs619521	rs11222109	128,578,157	130,336,236	1.758	191				
			12	rs10783433	rs11177160	51,755,293	68,706,883	16.952	250				
			13	rs9578416	rs12018729	22,543,935	55,846,574	33.303	250				
			13	rs1929742	rs9561704	76,820,308	95,534,065	18.714	250				
			13	rs716307	rs16972217	106,510,947	108,935,447	2.425	250				
			16	rs2133783	rs2306237	56,973,335	58,881,897	1.909	250	**GPR56**	5	2(2)	
			18	rs644435	rs1954825	355,944	2,068,872	1.713	229				
			18	rs2654173	rs623546	75,472,992	77,595,228	2.122	250				
12	PK64	N	4	rs3864210	rs6449144	7,376,694	9,944,650	2.57	255				

			6	rs11961985	rs2499762	21,824,770	33,948,624	12.12	257	**TDP2**	5	2(2)	
			9	rs6560041	rs10120372	90,495,370	108,440,542	17.94	10				
			9	rs4837957	rs4838177	119,629,567	127,218,258	7.589	12				
			11	rs11382548	rs661864	61,165,732	78,509,582	17.35	259				
			18	rs12185438	rs7227333	31,671,602	57,645,433	25.97	366				
			21	rs2824688	rs1826318	19,552,880	20,754,462	1.2	340				
13	PK65	N	1	rs1149331	rs12047015	7,528,669	9,061,382	1.533	288				
			2	rs17015647	rs9308990	35,013,851	36,571,060	1.557	276				
			17	rs6565705	rs8079404	13,905	6,742,672	6.729	500	**SLC13A5**	5	2(2)	
14	PK67	Y	No HBD										
15	PK68	N	9	rs10125775	rs12376992	36,993,071	81,059,643	44.066	750				
			12	rs2204849	rs1492330	38,094,338	41,025,447	2.931	643				
			13	rs7999277	rs596249	55,034,281	79,699,392	24.665	750	**SLAIN1**	5	3(3)	
16	PK71	N	No HBD										
17	PK72	Y	8	rs4840674	rs2188014	7,007,415	16,862,202	9.855	750				
			X							**ATRX**	6	3(3)	
18	PK73	N	1	rs4658913	rs10737832	231,160,646	239,341,587	8.181	1045				PK73-5 & 8 removed

													(CNVs)
19	PK74	Y	5	rs10045267	rs6555852	166,370,461	168,697,807	2.327	1244				
20	PK76	N	11	rs10789833	rs10892178	111,300,044	117,685,343	6.385	160				
			21	rs11700947	rs1018754	20126875	27025078	6.898	156				
21	PK77	N	21	rs11088960	rs11701571	46,135,051	47,963,149	1.828	330				
		Exclude PK77-2 & 9 (MAN1B1)	10	rs11259181	rs10763707	14,596,739	29,594,791	14.998	240				
			12	rs12307191	rs11061274	128,257,358	131,478,239	3.221	240				
22	PK78	N	1	rs12732088	rs1001737	175283119	234417623	59.135	80				PK78-7 removed (CNV)
			2	rs1030321	rs11681842	60135937	66072634	5.937	80				
			3	rs12639182	rs1168704	38736230	54897396	16.161	80				
			3	rs10510881	rs13096446	62620324	66488134	3.868	71				
			3	rs11128078	rs10934885	90030538	129298994	39.599	80				
			3	rs10212419	rs11714312	139004920	143335479	4.331	57				
			4	rs1115917	rs1023730	16104248	37240580	21.136	80				
			5	rs1124418	rs1279912	160384176	180690937	20.306	80				
			7	rs10242703	rs10270076	6870635	25363037	18.492	80				
			9	rs10992622	rs10120219	95830197	98364547	2.534	80				

			9	rs10820273	rs10978734	105518491	109926894	4.408	80				
			10	rs10458802	rs10795860	7618993	11362317	3.743	80				
			10	rs10826795	rs10751327	30661341	44137293	13.476	80				
			11	rs11600417	rs11038476	4639292	5659680	1.02	80				
			11	rs11529659	rs12291056	20224852	33415151	13.19	80				
			11	rs11020080	rs11217150	92578985	98707145	6.128	80				
			14	rs1319954	rs10137361	20347564	34396100	14.049	80				
			20	rs11698167	rs12329492	1848203	16582448	14.734	80				
23	PK79	N	6	rs12202342	rs1001290	2,759,201	5,910,329	3.151	230				
			16	rs11645094	rs11860145	17,635,895	53,120,022	35.484	240	**VPS35**	6	3(3)	
24	PK80	N	No HBD										
25	PK81	N	1	rs10902661	rs12040932	25,908,241	28,311,224	2.403	86				
			13	rs1000588	rs12871748	64,141,737	69,782,361	5.641	76				
			18	rs12965475	rs12457018	31,976,364	40,337,801	8.361	160				
26	PK83	N	5	rs11749259	rs12652071	16,993,278	18,913,137	1.92	84				
			8	rs10107946	rs10094970	111,087,094	113,432,766	2.346	62				
			10	rs10796127	rs10764480	13,826,242	24,870,596	11.044	160				
			13	rs1258	rs1286	91,717	96,138	4.421	83				

				4656	6697	,352	,087						
			16	rs1186 5835	rs1244 7345	31509 816	48,867 ,548	17.358	54				
27	PK85	N	1	rs6660 757	rs3054 57	83971 065	88,195 ,672	4.225	503				
			12	rs7958 517	rs1399 440	99,837 ,070	101,22 1,375	1.384	133				
			16	rs7195 219	rs1164 7703	27,277 ,056	50,913 ,577	23.637	268	**GPT2**	5	3(3)	
			16	rs1214 9832	rs9932 117	53,842 ,908	54,906 ,895	1.064	258				
			18	rs2658 761	rs1296 4725	54,059 ,039	55,655 ,487	1.596	172				
28	PK87	Y	2	rs1093 3405	rs7260 16	233,54 3,537	234,72 5,502	1.191	253				
			5	rs5834 42	rs4590 183	41,586 ,405	42,626 ,060	1.039	83				
			10	rs2459 420	rs1090 9011	37,302 ,621	42,436 ,301	5.134	88				
29	PK91	N	2	rs4663 509	rs6752 442	235,71 7,794	237,07 3,776	1.356	163				
			3	rs9833 256	rs4687 530	187,79 3,410	193,92 1,560	6.128	260				
			7	rs1023 9445	rs4720 103	13,674 ,145	33,166 ,070	19.492	260				
			7	rs1024 4533	rs2057 16	105,56 2,881	130,61 6,236	25.053	270				
			8	rs2843 894	rs9642 769	36,025 ,847	37,048 ,234	1.022	61				
			8	rs7837 156	rs7011 286	42,057 ,370	57,222 ,136	15.165	257				
			10	rs1373 795	rs2138 555	49,783 ,975	64,220 ,494	14.437	260				
			10	rs3890 159	rs2817 656	93,060 ,218	98,905 ,223	5.845	257				

			10	rs181632	rs12242686	119,364,995	120,739,449	1.374	146				
			13	rs4586294	rs619542	93,546,788	94,784,082	1.237	155				
			15	rs7171039	rs3762168	82,244,828	85,142,647	2.898	152				
			17	rs12325725	rs1561901	13,095,576	54,612,056	41.516	268	**SYNRG**	5	2(2)	
30	PK93	Remove PK93-4, Turner's	13	rs11148274	rs10507467	22,506,277	38,737,092	16.231	160				
			14	rs11627948	rs10873077	30,588,110	54,451,302	23.845	160				
31	PK94	N	12	rs10745708	rs11110663	95,284,726	101,507,917	6.223	167	**USP44**	13	3(3)	
			12	rs10774913	rs11059303	117,846,406	128,145,564	10.299	167				
32	PK95		20	rs6133123	rs6045177	4,449,931	18,048,084	13.598	775				PK95-7 removed (CNV)
													NB Not HBD in PK95-8
33	PK98	N	18	rs7237664	rs9967445	25,928,664	27,301,274	1.373	856				
34	PK100	N	No HBD										
35	PK101	Y	1	rs425181	rs2772304	61,235,684	68,636,580	7.4	276				
			1	rs10737633	rs6684597	193,808,664	195,085,692	1.277	165				
			6	rs9388801	rs17063374	130,850,699	134,434,375	3.584	274				
			7	rs12674093	rs2392412	33,554,162	36,073,889	2.519	276				
			8	rs1010	rs1842	70,863	74,818	3.954	298				

				5638	665	,894	,529						
			9	rs1079783	rs10745401	136,462,863	138,182,186	1.719	267				
			10	rs7100815	rs6584977	110,348,376	111,954,970	1.607	128				
			16	rs3095600	rs437947	52,431,954	57,942,866	5.511	272	**GPR56**	4	2(2)	
			16	rs235130	rs7190307	66,230,608	68,218,336	1.988	121				
36	PK102	Y	1	rs742671	rs6426752	15,519,417	22,718,991	7.2	750				
			2	rs6546619	rs41386450	70,938,472	106,637,411	35.699	750				
			2	rs13410347	rs13035389	196,472,286	199,666,111	3.194	632				
			3	rs2102313	rs1471738	72,026,826	100,915,962	28.889	750	**HTRA2**	5	3(3)	
			3	rs12637098	rs12489012	194,551,660	197,833,758	3.282	732	**DNAH7**	5	3(3)	
			4	rs744731	rs4406006	17,492,843	25,991,628	8.499	750				
			4	rs733778	rs1396066	41,819,608	55,289,826	13.47	750				
			4	rs2850373	rs10488899	102,168,234	107,864,811	5.697	750				
			9	rs10858121	rs11137376	138,445,655	141,025,328	2.58	616				
			10	rs3740057	rs2768312	101,658,585	119,200,948	17.542	750				
			13	rs264729	rs9551232	22,197,910	26,600,933	4.403	750				
			13	rs912518	rs11616977	29,197,494	52,092,247	22.895	750				
			19	rs11666603	rs8105737	12,496,934	15,198,631	2.702	736				
			19	rs1552223	rs3829647	41,525,952	52,216,563	10.691	750				

			22	rs2001134	rs4823603	27,337,048	47,557,638	20.221	750				
37	PK108	N	HBD										
38	PK111	N	5	rs4702235	rs1971391	5,187,371	10,512,330	5.324	500				
			5	rs158851	rs7708643	163,085,666	167,529,727	4.444	500				
			6	rs9504505	rs13192093	5,746,943	12,263,693	6.515	500				
			14	rs898927	rs807733	98,263,147	99,688,409	1.425	500				
			13	rs17075632	rs9546489	83,077,704	84,284,111	1.206	286				
39	PK113	N	2	rs6729991	rs2222186	79,641,996	105,296,110	25.654	411	**AFF3**	5	3(3)	
40	PK117		3	rs9289390	rs1501202	131,219,070	134,961,213	3.742	250				PK117-5 removed (CNV)
			3	rs2118097	rs572190	161,523,244	163,907,728	2.384	190				
			15	rs2705480	rs932066	87,861,276	89,595,518	1.743	185				
41	PK121	Y	1	rs867921	rs2819316	156,352,974	162,278,552	5.926	276				
			4	rs11933950	rs10518436	118,043,617	125,653,594	7.61	288				
			6	rs4945630	rs9402140	119,430,311	129,876,535	10.446	276				
			9	rs2453556	rs3010744	27,586,162	80,981,110	50.395	304	**GNE**	4	2(2)	
			12	rs11065733	rs4238061	111,294,520	119,763,890	8.469	500				

42	PK123	N	no HBD										
43	PK127	Y	1	rs931413	rs2154298	79,348,506	80,791,787	1.443	170				
44	PK135	Y	2	rs6749002	rs1257193	128,196,424	134,993,718	6.797	500				
			3	rs4133027	rs6550191	28,624,327	33,062,415	4.438	500				
			3	rs4370006	rs4621329	178,522,077	180,488,885	1.967	370				
			3	rs9825532	rs6803023	89,087,160	95,649,705	6.562	160				
			6	rs4711895	rs7745524	47,627,027	80,432,965	32.806	500				
			8	rs4562370	rs6471277	90,635,791	92,360,193	1.274	363				
			10	rs2802457	rs12775323	7,886,084	19,428,643	11.543	500				
			10	rs6482253	rs2480277	23,396,191	30,703,513	7.307	500				
			11	rs12786220	rs653343	21,378,056	64,606,177	43.228	500	**BDNF**	3	2(2)	
			11	rs4433578	rs12294124	130,580,691	134,934,063	4.353	500				
			12	rs7959625	rs16944338	111,623,138	114,894,728	3.272	500				
			15	rs11858967	rs12593926	87,125,552	88,831,599	1.706	406				
			16	rs8056400	rs12447373	95,254	5,194,458	5.099	500				
			16	rs7201505	rs11866970	28,071,953	79,164,452	51.093	500				
			16	rs12929880	rs1532471	84,600,595	86,185,911	1.585	500				
			18	rs8094315	rs4396631	60,936,047	65,852,638	4.917	500				
			21	rs10439884	rs451885	10,971,951	17,744,507	6.773	500				

45	PK138	Y	12	rs7972149	rs10744798	98,094,620	114,059,260	15.965	286				
46	PK144	Y	7	rs10241665	rs4730967	118,429,175	120,273,226	1.844	216				
47	PJ3	N	5	rs7715172	rs3733768	38452894	43297032	4.844	277				
			17	rs12450758	rs2472717	17,260,170	18,879,649	1.619	208				
			19	rs8101646	rs4802619	46,505,526	50,274,961	3.769	256				
			19	rs8111074	rs2603701	51,776,117	55,600,736	3.825	269	**MBOAT7**	10	4(4)	
48	PJ9	N	3	rs1880363	rs16851901	778288555	105697533	27.409	272	**TBC1D23**	12	5(5)	
49	RQ2	N	11	rs198444	rs10899186	61,505,168	75,962,755	14.457	755				
50	AS5	Y	1	rs6683683	rs17021310	103,671,460	104,858,324	1.187	306				
			8	rs4840674	rs7004322	7,007,415	8,761,295	1.754	292				
			8	rs4272426	rs7828391	11,763,932	12,728,705	0.965	287				
			8	rs13274488	rs6530659	12,969,055	13,496,786	0.528	345				
			19	rs3787062	rs8101981	6,229,287	16,144,586	9.915	500				
			16	rs16946340	rs2356736	48,667,717	49,897,295	1.23	381				
51	AS7	N	7	rs2390256	rs12534599	20,177,882	22,258,795	2.081	500				

			17	rs11657409	rs2015561	37,368,156	38,593,294	1.225	253				
			17	rs9900213	rs17599948	40,375,881	41,353,410	0.978	254				
			17	rs9909324	rs2109019	58,046,756	59,475,888	1.429	255				
			18	rs9957673	rs12454816	57,566,388	58,542,737	0.976	348				
			21	rs2831534	rs2832409	29,514,609	30,981,245	1.467	248				
			22	rs5747010	rs2000463	16,855,618	23,803,722	6.948	500				
52	AS15	Y	2	rs17648731	rs11164111	77,721,499	95,535,544	17.814	258				
			2	rs624817	rs6430600	135,278,207	136,836,365	1.558	259				
			6	rs4713605	rs568644	32985992	52817144	19.831	263				
			6	rs9388416	rs2254235	125,648,345	134,365,043	8.717	262				
			6	rs9371201	rs4495282	150,145,001	156,207,989	6.063	262				
			11	rs10767886	rs2957705	4,218,391	10,375,899	6.158	750				
			17	rs916625	rs8082029	5,975,043	12,761,374	6.786	261				
53	AS18	N	1	rs10752832	rs3118644	145,657,708	156,863,043	11.205	265				
			6	rs2746319	rs218291	80,415,428	108,467,024	28.052	507				
			11	rs308828	rs6483270	88,152,876	93,624,380	5.472	262				
			16	rs7202007	rs8060595	26,954,662	51,083,636	24.129	507				
			17	rs11078241	rs8082107	14,250,136	15,074,606	0.824	507				

54	AS20	N	3	rs1512903	rs6783837	145,912,548	155,048,352	9.136	500				
			7	rs10486106	rs10950497	10,957,622	13,947,289	2.99	500				
			8	rs757	rs12114308	28,244,724	35,920,869	7.676	500				
55	AS27	Y	No HBD										
56	AS28	N	No HBD										
	branch1	Y	1	rs10912551	rs1351963	173,119,967	175,343,297	2.223	635				
			2	rs13423995	rs2314703	33,012	4,479,479	4.446	1000				
			2	rs2685200	rs12469609	82,205,627	83,383,702	1.1708	284				
			2	rs6745441	rs7559678	96,206,153	98,755,252	2.549	332				
			3	rs6439578	rs7622879	135,280,770	136,771,382	1.489	248				
			3	rs11707053	rs9852425	139,035,269	144,215,405	5.18	1000				
			4	rs7690743	rs13150700	180,988,794	186,610,123	5.621	514				
			5	rs4573006	rs27863	44,611,650	50,165,489	5.554	646				
			6	rs9377009	rs7749357	146,919,485	151,655,001	4.736	756				
			8	rs12676279	rs2306975	43,784,378	48,353,258	4.569	326				
			12	rs10850933	rs6489171	118,340,875	127,882,841	9.542	532				
			16	rs2075637	rs735919	4,938,280	6,022,280	1.084	929				
			18	rs4399595	rs4797402	8,227,344	10,121,727	1.894	1000				
57	AS55	N	13	rs3013	rs4098	56,032	57,503	1.47	197				

				342	269	,422	,045						
58	AS56	N	2	rs1038570	rs7568533	62,890,171	64,547,448	1.657	180				
			5	rs10040259	rs10055072	18,797,589	21,391,244	2.594	270				
			5	rs11952815	rs7732200	112,502,941	113,279,043	0.776	333				
			7	rs4724335	rs4730274	44,978,350	107,479,719	62.501	504				
			12	rs4473005	rs11115162	75,521,362	82,426,466	6.905	505				
			14	rs7156008	rs3915546	63,509,861	64,617,958	1.108	179				
			16	rs957676	rs12447345	19,680,546	48,867,548	29.187	750				
			16	rs8059348	rs360774	51,000,977	55,401,333	4.4	262				
			17	rs1292032	rs7215706	57,992,343	59,366,049	1.374	128				
			18	rs1031676	rs644016	54,042,206	55,164,547	1.122	361				
59	AS66	N	5	rs868641	rs4868949	150,485,241	165,138,708	14.653	513				
			7	rs39203	rs12667819	89,734,282	106,546,657	16.812	508				
			7	rs6959729	rs10249619	107,446,180	109,827,761	2.382	504				
													NB AS66-3 has homozygous truncating mutation in MPDZ, also in affected mother, but father is WT-non-paternity indicated

			2	rs4148195	rs10198826	44,039,633	60,910,033	16.87	264				
			5	rs2189087	rs868641	136,934,484	150,485,241	13.551	261				
			7	rs10239100	rs39203	87,564,277	89,734,282	2.17	254				
60	AS101	Y	No HBD										
61	AS102	N	6	rs6456719	rs6939259	10,863,325	14,014,915	3.152	509				
			8	rs7463805	rs6980880	110,753,672	123,787,061	13.033	267				
			11	rs1045454	rs16937956	204,228	8,404,501	8.2	533	**PGAP2**	5	3(3)	
			12	rs4762437	rs2279766	98,205,710	117,595,978	19.39	270				
			12	rs2649978	rs3782257	119,288,160	124,854,903	5.567	263				
62	AS103	N	1	rs11577402	rs6668037	237,578,970	239,332,114	1.753	1256				
			1	rs1509701	rs12069813	239,214,483	240,712,898	1.498	1256				
			1	rs1509701	rs10802523	239,214,483	247,745,313	8.531	1256				
			14	rs797326	rs1742868	34,606,636	36,536,949	1.93	790				
63	AS104	Y	1	rs2275697	rs1572993	205,027,737	205,045,087	0.017	6	**CNTN2**	5	3(3)	
			3	rs4128782	rs6123	89,237,732	93,593,119	4.355	155				
			4	rs11722181	rs1380341	33,354,946	34,534,283	1.179	96				
64	AS107	N	1	rs515194	rs6689840	108,657,934	157,077,577	48.42	258				

			4	rs1157893	rs921853	31,744,052	34,984,841	3.421	206				
			5	rs12515078	rs2545394	53,659,481	66,406,143	12.747	257				
			6	rs1535011	rs812134	1,558,449	4,950,275	3.392	500				
			7	rs6465306	rs11981219	90,878,249	102,339,168	11.461	259				
			9	rs6476874	rs6477256	4,527,201	8,065,584	3.538	256				
			9	rs7027771	rs7021395	101,552,955	110,645,412	9.092	258				
			12	rs10859872	rs17216905	95,715,994	106,630,904	10.915	11				
			14	rs6573644	rs10146483	66,428,638	67,897,671	1.469	128				
65	AS109	N	3	rs1014229	rs11917617	47,652,795	55,952,782	8.3	758				
			15	rs2594934	rs8040885	28,182,715	31,740,294	3.558	995				
			20	rs6070667	rs6090228	57,542,883	61,706,750	4.164	1000				
66	AN17	N	No HBD										
67	AN24	Y	3	rs11716050	rs11720627	5,069,843	13,855,350	8.786	255				
			7	rs10488600	rs6464211	136,605,458	151,873,853	15.268	259				
			12	rs10129050	rs12368662	26,222,768	42,122,993	15.9	257				
			13	rs7326634	rs9517465	86,740,599	88,170,033	1.429	256				
			18	rs2156058	rs17813753	63,312,311	64,433,748	1.121	253				
68	IAID1	N	1	rs1086	rs6074	221,39	234,76	13.376	258				

				3595	07	3,875	9,499						
			2	rs6434455	rs6708542	192,085,141	205,256,040	13.171	256				
			4	rs13120580	rs10015045	172,352,348	178,498,527	6.146	261				
			5	rs436321	rs4957684	95,754,483	106,390,482	10.636	260				
			5	rs10038452	rs6601089	174,384,256	179,603,337	5.219	263				
			10	rs7904044	rs7897567	7,632,504	11,975,662	4.434	260				
			10	rs1650169	rs1863824	81,388,19	88,112,939	6.725	255				
			13	rs9511890	rs9513088	26,295,986	28,934,707	2.639	500				
			19	rs10412317	rs2290745	48,507,720	52,909,303	4.402	257				
			21	rs2223080	rs459617	29,801,854	31,140,753	1.339	120				
69	IAID2	Y	4	rs6854954	rs7657882	137,397,046	146,893,493	9.496	257				
			5	rs3112400	rs423333	71,394,869	73,218,954	1.824	146				
			7	rs37949	rs7777070	7,996,282	9,283,620	1.287	258				
			7	rs1526491	rs993900	83,433,484-	85,005,940	1.572	254				
			13	rs1011038	rs2335562	20,695,554	23,468,089	2.773	500				
			15	rs3848182	rs2043380	79,918,711	87,612,368	7.694	259				
			17	rs11868785	rs6607292	34,139,218	36,124,082	1.984	144				
			17	rs3785931	rs4341799	47,578,118	55,361,584	7.783	258				
70	IAID3	Y	8	rs2336695	rs2527074	1,033,625	5,509,449	4.476	256				

			13	rs7339152	rs9594670	37,308,647	42,631,489	5.323	256				
			15	rs2946365	rs1023781	86,332,964	98,833,418	12.5	256				
			17	rs1995523	rs6504803	31,788,443	51,020,706	19.232	257				
			17	rs720959	rs2411042	70,924,681	74,913,937	3.989	257				
71	IDH1	N	9	rs4841991	rs9410073	138386597	141022295	2.635	635				
72	IDH2	N	4	rs10032936	rs4521322	45,843,672	47,018,520	1.175	509				
73	IDH3	N	7	rs6972992	rs7802908	35,556,427	47,951,430	12.395	64				
			12	rs11063263	rs4766096	191,619	3,403,999	3.212	500				
			14	rs8006914	rs17091451	23,310,485	56,895,162	33.585	500	**TRAPPC6B**	5	2(2)	
74	IDH4	N	3	rs9840933	rs12639517	173,010,075	185,574,131	12.564	287	**ALG3**	6	3(3)	
			4	rs11733560	rs7656678	67,572,774	78,970,864	11.398	407				
			14	rs915363	rs1890982	101,600,931	104,637,327	3.036	81				
75	IDH5	Y	1	rs11102184	rs12566888	111,485,207	156,869,047	45.384	513				
			2	rs17386061	rs17019636	17,307,871	36,959,486	19.652	750				
			8	rs6991346	rs13262483	13,118,537	18,976,000	8.858	366				
			10	rs549167	rs4948627	43,186,821	44,923,802	1.737	256				

			10	rs1936602	rs4919105	92,445,253	99,280,233	6.835	292				
76	IDH6	N	6	rs200997	rs2158279	27,811,815	29,366,265	1.554	364				
77	IDH7	N	1	rs10907360	rs1036851	144,474,542	146,633,750	2.159	66				
			1	rs12040255	rs997985	173,538,503	175,010,010	1.472	88				
			3	rs1496522	rs34309901	89,012,893	93,721,783	4.709	145				
			11	rs366066	rs1791468	72,284,569	106,408,071	34.124	500				
			15	rs7163927	rs516628	23,915,420	29,710,092	5.795	500				
			16	rs8059783	rs2035490	83,012,842	85,459,188	2.446	500				
			16	rs237997	rs16945874	34,686,820	48,175,235	13.488	99				
			16	rs4889544	rs8047974	31,423,820	34,221,699	2.797	68				
78	IDH8	N	3	rs13324142	rs2952830	48,689,285	53,002,134	4.313	205				
			9	rs2807580	rs4579578	126,926,107	133,241,327	6.315	500				
			11	rs7124648	rs657554	47,044,249	56,629,376	9.585	500				
			11	rs7925428	rs12294124	128724209	134934063	6.21	500				
			16	rs8056400	rs2283492	95,254	4,037,525	3.942	500				
			16	rs2072076	rs8045920	19,236,181	61,107,891	41.872	500	ABCC12	4	2(2)	
			17	rs2047803	rs8067227	7,237,346	35,006,825	27.769	500				
			18	rs11080775	rs6506998	14,582,736	19,710,668	5.128	115				

79	IDH9	Y	2	rs1401383	rs12614652	86,122,874	87,430,727	1.307	147				
			2	rs12992413	rs1857745	97,277,489	99,016,975	1.739	165				
			2	rs11675333	rs4851455	100,637,331	102,083,183	1.445	159				
			2	rs6746088	rs826556	107,987,035	109,401,259	1.414	171				
			2	rs10171282	rs3789090	110,256,391	111,789,092	1.533	107				
			2	rs2952738	rs6728975	117,407,288	118,520,639	1.113	123				
			3	rs10865956	rs9825535	49,581,559	50,772,664	1.191	156				
			3	rs7628120	rs9823980	65,652,373	68,337,634	2.685	500				
			3	rs11533825	rs4632567	81,177,939	82,316,849	1.139	140				
			3	rs9875783	rs4270515	82,477,805	83,920,089	1.442	176				
			3	rs9857831	rs4682429	105,370,343	112,521,973	7.152	500				
			6	rs7739952	rs6935570	56,202,243	62,705,029	6.502	266				
			6	rs13212642	rs2496540	71,595,571	72,927,964	1.332	241				
			9	rs1536608	rs636548	223,613	3,248,543	3.025	500				
			17	rs4796768	rs1405952	40,224,118	41,753,789	1.53	211				
80	IDH10	N	2	rs1869829	rs7573156	135,877,562	149,850,409	13.973	1000	**HNMT**	11	4(4)	
			13	rs1336666	rs1475276	103,051,897	106,052,498	3.001	1000				
81	IDSG19	N	No HBD										

	IV:5 and IV:6 only:		2	rs7588344	rs17042407	75,846,558	113,558,914	37.712	304				
			10	rs2420994	rs11245008	123,862,952	128,193,368	4.33	268	**DMBT1**	6	3(3)	
			12	rs11063263	rs16933918	191,619	4,977,980	4.786	504				
82	IDSG29	N	2	rs4434038	rs2215972	30,551,366	40,137,942	9.587	750				
			13	rs9512673	rs1360485	27,981,532	31,031,884	3.05	336				
			17	rs2453580	rs3859192	19,438,321	38,128,648	18.69	612	**FBXO47**	5	3(3)	
83	IDSG30	N	8	rs12545900	rs12548021	24,666,099	29,344,462	4.678	1000				
84	IDSG41	N	15	rs2892067	rs12900387	91,582,079	97,520,230	5.938	553				
			15	rs2573643	rs11247329	100,470,089	102,398,631	1.929	527				
85	IDSG43	N	7	rs766544	rs42045	90,802,001	92,250,354	1.448	657				
			8	rs2444860	rs777808	99,052,895	116,203,234	17.15	594	**VPS13B**	8	2(2)	
86	IDSG49	N	6	rs3094112	rs9357234	30761667	37348920	6.587	734				
87	ZA7	Y	no HBD										
88	ZA8	N	no HBD										
89	ZA9	Y	22	rs5752638	rs2213645	28,188,203	29,361,159	1.173	114				

90	ZA10	Y	1	rs10493140	rs11205716	49,407,295	50,835,606	1.428	136				
			6	rs1407317	rs7754440	62,531,424	63,899,051	1.368	167				
			8	rs6468303	rs6415407	34,953,826	35,999,591	1.046	70				
			14	rs10137318	rs4274352	44,903,433	46,028,940	1.126	146				
			17	rs12600680	rs8064681	58,037,374	59,155,210	1.118	88				
			22	rs5752638	rs2213645	28,188,203	29,361,159	1.173	65				
91	ZA12	Y	no HBD										
92	ZA16	Y	13	rs4770052	rs9510445	21,095,720	23,535,168	2.439	750				
			13	rs9526073	rs9596598	46,006,512	52,433,786	6.427	725				
			17	rs12602627	rs11871181	53,595,972	54,957,775	1.362	354				
			21	rs2823990	rs992087	18,081,151	29,540,412	11.459	750				
Affymetrix 250K NspI													
93	PK1	N	4	rs2191685	rs7664104	14,392,437	25,594,335	11.202	1000	**CC2D2A**	46	5(5)	
94	PK10	N	7	rs2396676	rs1123065	113,327,166	114,664,779	1.338	401				
			3	rs11916855	rs1038544	96,145,675	97,452,064	1.306	230				
			3	rs16843144	rs1654925	135,283,979	136,690,391	1.406	378				
			16	rs1244	rs1761	31,860	48,063	16.203	622				

				7780	4940	,399	,755						
95	PK13	Y	2	rs17044776	rs9308741	116,370,016	117,721,562	1.352	375				
			17	rs10432016	rs2703813	11,990,679	20,115,315	8.125	500	**ALDH3A2**	10	2(2)	
			10	rs2749592	rs7084706	38,258,842	43,894,339	5.635	382				
96	PK18	Y	7	rs4721923	rs2237342	20,637,721	27,796,242	7.159	750				
			8	rs1012180	rs705998	57,429,005	70,132,590	12.704	750				
			8	rs979670	rs1470684	104,117,387	106,583,872	2.466	652				
97	PK19	N	3	rs2236975	rs11712164	50,483,437	52,244,051	1.761	597				
			4	rs9884979	rs12505447	120,869,741	129,760,450	8.891	413	**BBS7**	7	4(4)	
98	PK22	Y	2	rs4522651;	rs10495544	34,049	6,956,133	6.922	500	**TPO**	4	2(2)	
			4	rs7672789	rs7699169	57,482,846	92,093,824	34.611	500				
			13	rs1557071;	rs4773392	106,089,052	112,227,358	6.138	500				
			14	rs1243631;	rs1384963	21,043,283	22,970,327	1.927	464				
			16	rs10521033;	rs12929759	51,408,850	54,410,447	3.002	500				
			17	rs6502187;	rs17707404	12,190,744	14,627,309	2.437	359				
			19	rs4801483;	rs6510156	57,953,121	59,039,699	1.087	118				
99	PK25	N	4	rs4324578	rs16859193	44,363,889	46,196,104	1.832	380				

			5	rs261125;	rs10513025	4,369,887	9,623,622	5.254	500				
			8	rs1982691;	rs17590021	13,456,548	17,507,299	4.051	500				
			12	rs4765870	rs12370980	2,090,876	3,849,484	1.759	312				
			12	rs1594737	rs11105209	87,518,836	89,613,012	2.094	369				
			12	rs1672301	rs11059876	127,430,069	129,238,618	1.809	443				
			14	rs8006326	rs1031790	82,638,829	84,694,347	2.055	439				
			15	rs12900029;	rs4965897	95,953,267	102,085,600	6.312	465				
			19	rs771361;	rs10402271	23,260,883	45,329,214	22.068	500				
			22	rs139170;	rs17001225	44,599,452	50,021,820	5.422	500				
100	PK26	N	8	rs11783149	rs2978997	99,445,122	101,396,145	1.951	116				
			8	rs4500059	rs12541010	111,432,405	112,988,570	1.556	92				
			10	rs1035211	rs1408344	101,345,516	102,648,784	1.303	94				
			16	rs12447780	rs17811434	31,860,399	47,462,851	15.602	165				
101	PK27	N	2	rs13424724	rs6712463	184,914,442	186,282,807	1.368	708				
			2	rs207942	rs4450588	217,043,965	221,289,656	4.245	1000	**GMPPA**	6	4(4)	
102	PK30	N	4	rs2270269	rs16894270	8,580,952	10,165,779	1.585	319				
			5	rs12110233	rs4916663	87,092,766	88,302,330	1.21	337				
			5	rs171528	rs6888673	97,928,295	99,169,359	1.241	315				

			6	rs1334706	rs11155050	138,479,214	139,651,545	1.172	362				
			6	rs9389717	rs605671	140,087,307	141,179,655	1.092	308				
103	PK36	Y	1	rs765751	rs10737812	219,669,226	237,058,828	17.39	500				
			2	rs4076190	rs4079660	69,328,389	89,285,770	19.957	500				
			2	rs6543294	rs11884156	106,036,457	123,262,388	17.226	500				
			3	rs9863076	rs10511147	88,033,939	94,030,669	5.997	340				
			3	rs2178532	rs6782743	171,508,977	186,159,130	14.65	500				
			4	rs17595784	rs7666487	45,840,803	47,243,737	1.403	340				
			5	rs1875018	rs2455458	10,451,417	14,665,615	4.214	500				
			5	rs4245982	rs7720165	41,756,930	43,301,544	1.545	245				
			7	rs4727881	rs1072198	118,305,379	120,327,349	2.022	394				
			7	rs10487463	rs7806653	126,619,872	138,438,487	11.819	500				
			8	rs2304297	rs10087659	42,608,199	49,820,430	7.212	322				
			9	rs7873275	rs10869719	38,768,168	78,785,890	40.017	500				
			10	rs1668169	rs7475340	73,825,652	76,839,355	3.014	248				
			14	rs12432204	rs2268337	90,609,207	95,083,379	4.474	500				
			17	rs6502259	rs473502	13,320,413	14,349,379	1.029	344				
			17	rs1015673	rs741071	34,477,386	47,591,690	13.114	500				
			18	rs12456734	rs1785427	4,430,692	5,579,427	1.149	380				
			19	rs8105	rs4655	261,03	6,941,	6.681	278				

				536	43	3	867						
			19	rs12460985	rs1697531	45,690,889	51,745,142	6.054	432				
			21	rs390344	rs243617	15,332,033	19,204,969	3.873	500				
104	PK46	Y	16	rs12596707	rs6498105	8,242,146	10,816,321	2.574	1250	**PMM2**	8	5(5)	
			3	rs3792558	rs2188151	46,903,094	50,201,924	3.299	814				
			3	rs172087	rs9815442	81,466,954	82,045,528	0.579	836				
105	PK50	N	2	rs13023579	rs3911238	186,190,812	187,470,344	1.279	371				
			2	rs10197088	rs919538	193,724,143	195,863,427	2.139	370				
			2	rs6723635	rs41499250	197,124,182	199,101,213	1.977	330				
			2	rs7577929	rs870068	205,231,284	206,625,266	1.394	343				
			3	rs732096	rs2569993	11,893,950	12,926,096	1.032	341				
			4	rs11133030	rs12500049	175,017,835	176,661,808	1.644	295				
			7	rs407488	rs10263367	68,872,414	70,082,089	1.21	329				
			11	rs10835308	rs326783	28,338,360	29,633,243	1.295	329				
			11	rs17147220	rs11228758	55,468,553	56,508,476	1.04	387				
			11	rs17157266	rs1791687	62,199,817	66,227,184	4.027	302				
			16	rs8048000	rs1566466	31,845,697	48,512,961	16.667	222				
			19	rs2052854	rs683856	48,566,037	50,822,408	2.256	305				

106	PK52	Y	No HBD										
107	PK54	N	8	rs10112802	rs4737311	70,968,179	72,120,122	1.152	1356				
			8	rs1994035	rs2514706	79,003,651	101,110,719	22.107	1369	**IMPA1**	11	6(6)	
			18	rs2850935	rs2578203	75,072,931	76,251,238	1.178	1380				
108	PJ1	Y	3	rs938382	rs1201763	134,589,067	138076206	3.487	432				
			3	rs9825064	rs12635525	139,049,804	157,432,037	18.382	750				
			9	rs12341394	rs7849050	22,148,055	24,750,179	2.602	750				
			10	rs1831183	rs734187	26,864,609	33,484,696	6.62	645				
			16	rs7190651	rs211504	23,757,600	26,799,280	3.041	545				
			21	rs118512	rs12481771	15,580,334	16,958,897	1.379	662				
109	PJ2	N	7	rs9642374	rs6958850	53,293,700	70,166,762	16.873	750				
			10	rs1925574	rs4746094	68,691,199	73,540,915	4.85	750				
			14	rs11158445	rs17756627	63155404	69090437	5.935	750				
			16	rs1088638	rs8050261	91,010	6303301	6.212	618				
110	PJ5	N	11	rs6589964	rs10894834	122,870,683	134,376,325	11.506	500	**DCPS**	16	4	
			14	rs4060079	rs4982731	20,419,930	23,585,333	3.165	500				
			17	rs10512514	rs11652864	65,211,876	70,296,265	5.084	500				

111	PJ6	Y	No HBD										
	Left branch only	Y	1	rs34075261	rs1408586	112,233,741	119,079,652	6.846	500				
			4	rs6532686	rs1546176	77,804,009	79,483,435	1.679	413				
			5	rs274714	rs852585	6,723,546	10,955,893	4.232	500				
			16	rs1019257	rs1868983	6,749,612	8,482,800	1.733	500				
			17	rs8064924	rs733099	18,901	4,658,971	4.64	391				
	Right branch only		4	rs11098903	rs1709425	127,831,579	130,138,940	2.307	905				
			10	rs7911009	rs2170147	50,138,459	53,027,556	2.889	901				
			12	rs2012848	rs4321008	94,428,565	101,945,736	7.517	1000				
			X							**GPR56**	11	6(6)	
112	PJ7	Y	1	rs1281338	rs10798030	182,112,106	185,997,086	3.885	750	**LAMC1**	8	3(3)	
			1	rs4276872	rs269720	188,251,894	189,491,402	1.24	469				
			2	rs9309386	rs10209422	25,727,910	28,430,531	2.703	394				
			13	rs9598249	rs269574	61,818,094	64,632,805	2.815	388				
113	PJ8	Y	No HBD										
		PJ8-11 only	11	rs498337	rs1498553	1,854,571	5,709,028	3.854	0	**TH**	11	5(5)	
		All other affected	1	rs17541444	rs12032036	115,917,891	119,930,368	4.012	1000				
			4	rs4426781	rs6820406	98,062,047	99,552,178	1.49	823				
			4	rs1353309	rs7654790	135,012,093	138,990,667	3.979	1000				

			6	rs17511672	rs9491317	107,776,430	125,470,012	17.694	1000				
114	PJ10	N	No HBD										
115	PJ11	N	8	rs2466089	rs7844223	32,526,678	49,966,660	17.44	500	**HGSNAT**	4	2(2)	
			9	rs10965941	rs17771857	23,583,321	28,209,091	4.626	500				
			16	rs4553646	rs13335504	31,536,419	48,252,973	16.716	211				
			16	rs8048836	rs383384	58,320,844	65,965,884	7.645	500				
			22	rs132404	rs5768344	45,130,373	48,491,160	3.361	500				
116	PJ12	Y	2	rs17265949	rs10803888	169,664,458	176,504,793	6.84	500	**LRP2**	7	2(2)	
			3	rs3856923	rs7619326	188,844,697	192,814,240	3.97	500				
			4	rs7664024	rs800453	18,944,656	24,818,509	5.874	500				
			4	rs4470656	rs11099551	80,576,083	83,700,755	3.124	500				
			4	rs1789911	rs12648912	100,263,778	108,517,790	8.254	500				
			4	rs7663356	rs1503539	118,489,947	120,811,913	2.322	472				
			6	rs1744415	rs1886979	39,215,601	43,904,901	4.689	500				
			8	rs10090760	rs1842627	82,158,027	84,302,146	2.144	388				
			10	rs10902911	rs1710313	124,994,243	127,734,513	2.74	500				
			11	rs10896444	rs611221	68,967,247	75,309,571	6.342	500				
			13	rs8002097	rs972477	62,008,597	65,300,618	3.292	480				

			16	rs399573	rs1428764	62,651,566	64,916,648	2.265	193				
117	RQ1	N	4	rs565867	rs3796857	13121722	15983799	2.862	500				
			16	rs17259551	rs30842	54504019	58743454	4.239	500	**GPR56**	8	3(3)	
			16	rs12447780	rs11861127	31860399	48186373	16.326	212				
118	RQ7												
119	RQ8	N	9	rs11103117	rs12238423	138,527,913	140,691,283	2.163	802	**MAN1B1**	4	2(2)	
120	RQ9												
121	AS6	N	5	rs3776569	rs266591	36,639,458	64,347,543	27.708	750	**ERCC8**	4	3(3)	
122	AS8	N	4	rs1948828	rs1026305	70540370	72423817	1.883	295				
			7	rs11762869	rs7803406	19,200,619	20,469,685	1.269	540				
			7	rs41326151	rs245905	27,648,621	29,199,462	1.551	625				
			10	rs1953092	rs10763607	59,799,112	61,204,465	1.405	256				
123	AS16	N	No HBD										
124	AS17	Y	5	rs6873130	rs4129562	108,214,413	109,562,274	1.348	363				
			7	rs4727881	rs10953925	118,305,379	120,715,154	2.41	334				
			8	rs4740033	rs980086	85,762,603	87,010,204	1.248	114				
			8	rs3779804	rs16875013	87,623,585	88,925,380	1.302	164				

			8	rs17720712	rs989108	89,105,615	90,281,984	1.176	230				
			10	rs10508861	rs11239629	38,064,069	43,001,367	4.937	262				
			11	rs11232256	rs642603	80,344,863	104,707,732	24.363	500	**CCDC82**	7	4(4)	
			11	rs4080494	rs11228758	50,218,185	56,667,628	6.449	420				
			13	rs9564421	rs9572031	68,233,366	69,433,995	1.2	228				
			14	rs17224456	rs1256056	63,700,859	64,712,626	1.012	158				
			18	rs1456609	rs17823562	41,844,641	42,988,088	1.143	211				
125	AS21	Y	14	rs7143288	rs4905446	92,191,783	96,597,957	4.26	500				
			5	rs692759	rs41454653	98,192,545	102,044,661	3.852	118				
			6	rs7746199	rs9368571	27261324	28677758	1.416	314				
			8	rs4507802	rs1499452	136,634,547	137,976,964	1.342	196				
126	AS22	N	2	rs1870130;	rs375793	6,584,163	10,029,440	3.445	500				
			2	rs17368670;	rs6723291	15,091,566	19,033,965	3.942	500				
			7	rs38841	rs10259910	116,319,926	117,503,982	1.184	330				
			8	rs11995542;	rs10808690	27,347,371	59,104,280	31.757	500				
			8	rs13251429;	rs11779779	96,423,252	123,767,773	27.344	500				
			9	rs7027322;	rs4977237	14,785,224	19,849,715	5.064	500				
			10	rs2771247;	rs10508782	30,635,748	33,733,976	3.098	500				
			10	rs1578875	rs2663058	45,550,145	50,126,899	4.577	431	**MAPK8**	5	2(2)	

			11	rs574951;	rs11220985	116,253,844	127,464,040	11.21	500				
			16	rs10871364;	rs11150354	79,460,567	81,200,651	1.74	500				
			21	rs4817936;	rs1006890	39,657,076	41,265,404	1.608	462				
127	AS26	N	1	rs623817	rs7532617	20,986,078	37,089,923	16.104	750				
			10	rs3740452	rs737314	60,565,959	64,274,177	3.708	750				
			12	rs10082845	rs2160667	8,507,181	13,052,677	4.545	750				
			13	rs16966394	rs9515106	105,986,281	110,356,151	4.37	750				
			14	rs11850361	rs10140682	24,220,155	26,538,360	2.318	750				
			14	rs2180375	rs10150519	27,454,218	29,914,167	2.46	750				
128	AS58	Y	4	rs342459	rs981544	88,059,415	97,931,374	9.872	750				
			13	rs4943392	rs1160321	37,146,599	40,155,143	3.009	750				
129	AS61	Y	4	rs2366715	rs17829765	73,854,103	75,196,731	1.342	439				
			5	rs2304057	rs155941	137,897,174	139,433,638	1.536	656				
			10	rs2505171	rs2565200	37,323,786	43,622,933	6.299	478				
			16	rs12447780	rs933544	31,860,399	48,859,532	16.999	457				
	Splitting by ABI2 genotype, ASMR61-1& 3, and 2 &4												
		2&4	7	rs2040901	rs17241389	80,959,480	83,764,309	2.805	500				
			9	rs12335892	rs10960462	10,821,734	12,082,380	1.26	382				

			11	rs11246002	rs17330155	211,447	4,911,774	4.7	500				
			13	rs9599137	rs678487	67,290,628	69,410,200	2.12	469				
			17	rs9900839	rs885688	6,991,442	12,183,999	5.193	500				
			18	rs16974887	rs487455	10,555,012	24,367,973	13.813	500				
			18	rs2298614	rs8094653	29,057,915	37,686,961	8.629	500				
		1&3	2	rs259814	rs1079943	180,718,145	206,464,465	25.746	500	**ABI2**	7	4(4)	
			4	rs4292329	rs11721812	10,144,459	25,730,546	15.586	500				
			4	rs17654423	rs997407	47,722,775	62,878,316	15.156	500				
			5	rs4836446	rs891987	128,464,956	141,467,196	13.002	500				
			9	rs7859893	rs4879990	6,568,116	27,977,820	21.41	500				
			9	rs6560207	rs2490211	74,106,055	76,216,694	2.111	500				
			13	rs993726	rs791166	55,243,102	57,003,808	1.761	339				
130	AS62	N	7	rs10486722	rs12673613	41,811,908	67,434,884	25.623	750	**PSPH**	10	5(5)	
131	AS64	N	8	rs4546626	rs2087457	110,125,762	127,715,881	17.59	750				
			14	rs1959801	rs10459466	101,769,738	107,145,272	5.376	503				
132	AS68	Y	5	rs25920	rs11956061	110,793,142	126,245,898	15.453	500				
			10	rs596406	rs2505453	11,224,665	26,023,828	14.799	500				
			14	rs8008151	rs10131137	59,813,702	61,592,366	1.779	372				

			14	rs7159878	rs2296316	64,466,684	65,520,246	1.054	351				
133	AS70	Y	3	rs6548853	rs1248844	82,823,120	84,880,238	2.057	449				
			6	rs6921049	rs532230	73,785,937	100,045,602	26.26	750	**MAP3K7**	6	3(3)	
			7	rs6466665	rs12706310	118,213,868	120,699,177	2.485	596				
			11	rs3816724	rs10501331	47,312,045	51,429,291	4.117	294				
			11	rs34288506	rs17615252	59,248,250	60,251,402	1.003	315				
			20	rs6138473	rs6120970	25,012,155	30,232,452	5.22	270				
134	AS72	N	16	rs11076059	rs818406	54,850,716	65,477,189	10.626	1250				
			16	rs1562110	rs4785775	86,759,160	90,163,275	3.404	1022	**FBXO31**	11	5(5)	
135	AS74	N	1	rs17545392	rs4300268	73,055,268	77,047,383	3.992	750				
			8	rs2729459	rs769031	39,820,230	70,351,166	30.531	750	**HGSNAT**	11	3(3)	
136	AS105	N	6	rs339361	rs17505997	117162515	119239308	2.077	337				
			8	rs9297957	rs1603941	96393714	118359997	21.966	500	**PIDD**	6	2(2)	
			8	rs10109650	rs6470831	123880227	131641096	7.761	500				
			11	rs11246002	rs16929027	211447	3033192	2.822	238				
			15	rs8032120	rs8038073	82117891	91604722	9.487	500				
137	AS110	N	11	rs1124	rs4929	21144	25282	2.137	1422	**PIDD**	13	8(8)	

				6002	992	7	33						
138	AS111	Y	2	rs1825497	rs6430096	139,688,866	146,454,914	6.766	833				NB not HBD in 111-4
139	AS112	N	No HBD										
140	AS113	Y	No HBD										
141	AS114	N	NB. ASMR114-4 has homozygous truncating mutation in FMN2; excluded as phenocopy										
			1	rs2422137	rs1909067	72,619,961	74,170,011	1.55					
			17	rs7216721	rs11655754	36,782,185	42,697,484	5.915	789				
			17	rs10445383	rs1503059	48,985,443	50,204,456	1.22	355				
142	AN13	Y	No HBD										
143	AN14	N	5	rs2441949	rs2914296	2,627,075	7,657,537	5.03	750	**NSUN2**	8	3(3)	
144	AN16	N	No HBD in all 4 branches										
		3 out of 4 branches	7	rs1557782	rs17345800	95,099,157	105,277,793	10.179	768	**AP4M1**	29	8(8)	
145	AN18	N	2	rs601255	rs12712499	31,452,231	36,590,531	5.138	750				
			7	rs11765805	rs1079515	153,260,518	154,866,259	1.606	669				
			7	rs6947568	rs2178820	154,900,894	159,105,036	4.204	687				
			9	rs3217986	rs10511813	22,005,330	27,664,162	5.658	750				

146	AN19	N	1	rs12120384	rs10442682	83,292,220	115,864,549	32.572	750				
			1	rs6587562	rs863028	151,246,241	159,104,085	7.858	750				
			1	rs269728	rs16832366	189,520,248	190,449,380	0.929	490				
			1	rs2400604	rs10922193	195,531,171	197,240,482	1.709	493				
			2	rs4953938	rs9287475	135,694,848	138,306,981	2.612	744				
			2	rs6755879	rs1430271	82,432,143	83,378,207	0.946	420				
			3	rs6790153	rs7614661	155,088,878	168,136,690	13.048	750				
			5	rs17461314	rs6891533	110,570,176	111,935,877	1.366	666				
			7	rs10267037	rs7789449	154,389,895	157,227,627	2.838	675				
			7	rs11561862	rs17688583	39,885,541	40,865,696	0.98	566				
			7	rs2072071	rs7810838	94,054,223	95,278,056	1.224	599				
			8	rs10093553	rs2061565	99,402,038	101,002,820	1.601	379				
			10	rs11259283	rs7909447	14,740,813	27,221,632	12.481	750				
			11	rs1791687	rs12271290	66,227,184	68,326,236	2.099	342				
			11	rs11236915	rs10899421	76,440,869	77,708,667	1.268	412				
			13	rs684587	rs9315762	30,619,171	40,741,907	10.123	750				
			16	rs1857943	rs1566466	31,938,455	48,512,961	16.575	445				
			18	rs16963382	rs1941689	30,032,385	31,272,403	1.24	506				
			20	rs6140226	rs6074396	7,226,483	12,224,896	4.998	750				
			22	rs1772	rs1326	33,655	36,579	2.924	750				

				1787	93	,597	,561						
147	AN20	N	12	rs3858620	rs2712542	114,929,252	127,730,787	12.802	1000	**TCTN2**	7	4(4)	
148	AN21	N	6	rs6923488	rs6904814	7,918,405	11,324,196	3.406	750				
			8	rs17079488	rs17094082	2,986,574	21,341,176	18.355	750	**TUSC3**	8	4(4)	
149	AN22	N	3	rs2173068	rs41486248	162,363,222	163,242,806	0.88	586				
150	AN27	N	6	rs2326365	rs4714567	3,846,444	12,430,430	8.584	500				
			7	rs12667669	rs17219864	17,425,895	20,773,567	3.348	500				
			7	rs17157815	rs7801799	29,400,704	32,032,526	2.632	282				
			8	rs4045956	rs11778615	219,677	3,360,214	3.141	461				
			9	rs2812276	rs4740401	130,058,579	133,938,676	3.88	301				
			10	rs821678	rs10794108	108,123,430	128,423,873	20.3	500				
			11	rs7927138	rs10840269	7,727,886	9,628,802	1.901	401				
			13	rs727984	rs1411754	45,462,542	92,891,557	47.429	500				
			15	rs11854354	rs1484198	23,799,284	46,677,273	22.878	500	**DUOX2**	7	3(3)	
			15	rs4965178	rs7169385	97,910,418	102,374,592	4.464	500				
			19	rs1600361	rs1993883	23,572,497	28,467,076	4.895	368				
			19	rs10424306	rs2932766	35,685,739	48,390,424	12.705	500				

151	AN28	N	18	rs747175	rs11150946	72,154,720	74,074,707	1.92	750				
			20	rs6061015	rs2424955	30,473,188	31,692,076	1.219	159				
			20	rs6087456	rs6087577	31,740,804	32,955,423	1.214	439				
152	AN30	N	2	rs11694172	rs231747	203,532,304	204,688,432	1.156	190				
153	AN37	N	17	rs1880756	rs2083797	43,832,367	44,921,897	1.09	626				
		NB. In 3 out of 4 affected:	19	rs2191667	rs17654300	9,348,586	17,788,648	8.44	788	**TRMT1**	10	4(4)	
154	AN43	N	9	rs943423	rs11137379	137,437,183	141,027,939	3.591	616	**MAN1B1**	8	3(3)	
			16	rs8048000	rs17752673	31,845,697	48,477,093	16.631	578				
155	AN45	N	no HBD										
156	AN48	N	1	rs17192059	rs3737741	45,699,20	47,000,566	1.301	293				
			2	rs10931359	rs6761221	189,308,284	190,798,939	1.491	390				
			3	rs9829940	rs2317212	96,167,649	97,328,284	1.161	225				
			7	rs12534529	rs7795646	118,660,614	120,323,727	1.663	404				
			12	rs4435086	rs10858818	88,344,477	89,426,732	1.082	282				
			16	rs6565284	rs1566466	31,804,434	48,512,961	16.709	411				
			18	rs11659761	rs4891027	73,186,974	74,273,213	1.086	437				
			19	rs4807949	rs807969	1,676,419	7,838,484	6.162	605	**CAPS**	7	3(3)	

157	AN49	N	1	rs638672	rs12137158	225,458,963	240,660,818	15.202	750	**FMN2**	5	3(3)	
			12	rs729784	rs11066476	109,981,264	113,465,367	3.484	594				
158	AN50	N	3	rs12492625	rs797078	71,687,846	109,144,543	37.457	1250	**ZBTB11**	14	5(5)	
159	AN51	N	8	rs4292660	rs2679749	99,796,296	103,858,488	4.062	1793	**VPS13B**	21	9(9)	NB. HBD in 7 out of 8 affected
160	AN52	N	1	rs3930209	rs7531369	15,126,214	22,639,161	7.513	453				
			1	rs11102058	rs11578062	110,666,410	114,870,307	4.204	689				
			5	rs11746959	rs1376180	34,266,680	40,939,640	6.673	689				
			5	rs7723361	rs749073	49,974,032	51,615,913	1.642	535				
			5	rs117931	rs4865610	52,453,866	54,810,847	2.357	664				
			6	rs9403683	rs1338167	100,802,068	101,901,973	1.1	358				
			8	rs10488368	rs1035633	190,568	3,923,744	3.733	750				
			8	rs7834078	rs4735765	68,579,813	78,097,805	9.518	750				
			8	rs1519342	rs2583764	78,691,400	79,663,800	0.972	261				
			9	rs600951	rs7023760	224,742	13,881,467	13.657	750				
			12	rs1860360	rs2401389	493,900	9,866,682	9.373	711				
			12	rs2034337	rs2044316	10,031,880	20,471,228	10.439	750				
			13	rs1977704	rs1407992	19,906,422	31,708,167	11.802	750				

			18	rs12458980	rs11082650	44,199,217	45,649,628	1.45	406				
			19	rs12983082	rs7252635	11,216,561	13,552,942	2.336	418				
			19	rs7259659	rs1673917	48,606,054	52,993,620	4.388	639				
161	AN53	N	6	rs7767794	rs4707962	67,222,654	68,959,224	1.737	436				
			6	rs12210297	rs9371807	151,639,691	154,760,029	3.12	500				
			7	rs7784067	rs10225488	31,084,099	32,128,882	1.045	379				
			8	rs604449	rs6469644	116,005,882	117,481,751	1.476	286				
			9	rs3008110	rs4742409	186,603	7,774,180	7.587	500				
			9	rs10959367	rs10809674	10,917,884	12,015,741	1.098	334				
			16	rs811049	rs2562125	72,260,046	77,349,653	5.09	500				
			17	rs9916009	rs7210592	71,001,479	75,966,594	4.965	500	**SDK2**	6	2(2)	
			22	rs4818972	rs2839373	46,176,838	48,069,930	1.893	272				
162	ZA5	N	2	rs7608111	rs1509364	21,607,079	22,621,296	1.014	342				
			8	rs7826857	rs2061565	99,436,341	101,002,820	1.566	296				
			10	rs4024198	rs7084413	22,092,310	23,274,148	1.182	427				
			11	rs1074014	rs1487805	108,447,598	109,548,952	1.101	722				
			12	rs625684	rs7965166	4,037,169	7,730,186	3.693	938				
			12	rs12227686	rs1489347	86,113,178	87,254,962	1.142	438				

163	ZA17	N	2	rs7569119	rs10496298	82,472,911	83,694,607	1.222	287				
			4	rs6531606	rs12506905	33,376,313	34,761,595	1.385	243				
			6	rs9373448	rs362848	145,607,995	146,682,485	1.074	267				
			7	rs10249530	rs3802059	90,992,321	92,915,394	1.923	364				
			7	rs2706848	rs6977743	95,492,611	104,107,166	8.615	500				
			7	rs425013	rs10953607	107,840,971	109,052,359	1.211	323				
			7	rs4727881	rs1072198	118,305,379	120,327,349	2.022	394				
			16	rs8048000	rs1566466	31,845,697	48,512,961	16.667	217				
			21	rs3788169	rs11702704	46,580,779	47,868,321	1.288	309				
Affymetrix 5.0													
164	PK7	N	7	rs17561679	rs6465946	84,739,093	103,588,152	18.849	1706	**PEX1**	**9**	**5(5)**	
465	PK8	N	No HBD										
166	PK11	Y	3	rs11709020	rs12487108	46,680,063	52,063,886	5.384	801				
			5	rs2429191	rs156322	129,376,192	131,653,925	2.278	614				
			7	rs6466665	rs2244863	118,213,868	120,622,836	2.409	672				
			8	rs10097518	rs10503764	23,615,779	24,849,522	1.238	441				
			9	rs11142955	rs669296	74,283,609	79,297,906	5.014	1500				
			11	rs782409	rs650393	44,468,534	73,351,285	28.883	1500				
			12	rs1525	rs1947	33,966	39,450	5.484	629				

				895	234	,496	,491						
			16	rs12447780	rs17752673	31,860,399	48,477,093	16.617	593				
			16	rs825837	rs11861857	73,539,548	82,679,532	9.14	1500				
			17	rs17687849	rs12452064	43,759,719	44,868,187	1.108	365				
			17	rs1024637	rs17597484	58,038,893	59,296,426	1.258	256				
		PK11-4 only	8	rs2898421	rs13279752	15,476,431	15,509,793	0.03	5	**TUSC3**	15	3(3)	
167	AS3	N	3	rs2220504	rs17606712	1,235,201	3,101,524	1.866	1243				
168	AS11	N	No HBD							**ATP8A2**	11	4(4)	
169	AS19	N	8	rs1519368	rs11990111	139,389,995	142,709,706	3.32	1229	**TRAPPC9**	20	7(7)	
170	AS23	N	No HBD										
	Excluding branch 1 (non-consanguineous marriage):												
		Y	1	rs3736330	rs12094945	2,327,075	4,933,154	2.606	871				
			1	rs17027186	rs10754855	9,068,213	14,742,723	5.675	1000				
			4	rs12645808	rs16882428	28,851,586	29,903,074	1.051	779				
			5	rs6883959	rs10940474	51,493,935	54,880,701	3.387	1000				
			13	rs12429402	rs7334690	22,388,941	28,081,535	5.693	1000	**SPATA13**	9	3(3)	
			13	rs9530615	rs1324790	77,540,355	78,741,934	1.202	669				
			16	rs1731083	rs2719726	8,824,150	10,779,741	1.956	1000				

			22	rs5760752	rs5755492	25,345,537	35,395,126	10.05	1000				
171	AS25	Y	3	rs13089564	rs4685485	310,946	2,186,551	1.876	1000				NB. Cousin from non-consanguineous marriage excluded
			12	rs6582179	rs10746227	74,029,874	82,637,350	8.607	1000				
			19	rs2617815	rs3786654	6,795,147	14,557,821	7.763	1000				
172	AS30	Y	No HBD										
Affymetrix CytoScanHD													
173	PK28	N	14	rs1838332	rs11158653	66,517,249	67,658,895	1.142	901				
			16	rs11150644	rs17821664	31,860,504	48,202,503	16.342	617				
			17	rs2531899	rs8064681	58,007,193	59,155,210	1.148	640				
174	PK29	N	1	rs17107837	rs10489663	53,537,699	55,186,686	1.649	500				
			1	rs1322374	rs10920277	186,114,325	189,970,368	3.856	500				
			1	rs2119335	rs4846819	214,428,811	217,385,077	2.956	500				
			4	rs11099695	rs6814401	149,354,139	153,575,339	4.221	500				
			6	rs10946317	rs9366426	19,452,510	22,064,639	2.612	500				
			6	rs4469587	rs7766285	58,727,508	63,830,759	5.103	377				
			11	rs1396621	rs17151327	55,575,934	56,653,038	1.077	392				
			12	rs1077	rs1117	33,841	39,543	5.701	366				

				2110	0872	,755	,250						
175	PK33	Y	1	rs2274327	rs7517230	9,009,406	11,060,799	2.051	608				
			2	rs11695058	rs4516485	46,978,349	48,940,300	1.962	913				
			11	rs3802985	rs10838474	198,510	5,659,341	5.467	1000				
			12	rs2694414	rs2076022	105,041,664	120,875,790	15.834	1000				
			14	rs9805961	rs807736	97,891,268	99,690,907	1.8	1000				
			17	rs854767	rs1989229	18,018,263	43,312,893	25.295	1000	**NAGS**	4	1(1)	
176	PK34	N	3	rs13095891	rs13062228	48,381,826	52,932,744	4.551	586	**UBA7**	4	2(2)	
			3	rs13077937	rs9289536	88,537,491	93,548,479	5.011	620				
			4	rs16870359	rs9291437	21,097,417	22,165,255	1.068	556				
			5	rs67611103	rs17102667	143,368,861	144,454,943	1.083	512				
			7	rs10251196	rs10953925	119,588,295	120,715,154	1.127	508				
			8	rs4739562	rs13263200	33,844,064	35,916,718	2.073	930				
			8	rs10101998	rs16875714	85,137,646	87,053,655	1.916	552				
			8	rs6992633	rs4909666	137,351,166	138,404,925	1.054	510				
			12	rs10879502	rs10732757	73,220,644	74,857,957	1.637	522				
177	PK38	Y	7	rs36064812	rs17139528	62,829,382	64,855,177	2.025	376				
			18	rs4083288	rs8083133	15,101,910	19,543,272	4.441	178				

178	PK39	N	No HBD										
179	PK49	Y	2	rs7602185	rs2303845	212,463,306	216,861,273	4.398	1000				
			3	rs17073477	rs2399106	65,685,805	106,337,570	40.651	1000				
			4	rs874950	rs1706214	33,299,963	34,873,377	1.573	789				
			4	rs1328925	rs12508239	159,190,290	163,267,782	4.077	1000				
			6	rs9460387	rs580921	19,629,921	32,287,496	12.658	1000	**TDP2**	5	2(2)	
			6	rs4715249	rs16885212	51,702,799	54,179,404	2.477	1000				
			6	rs236232	rs16876230	79,126,929	86,132,363	7.005	1000				
			13	rs9562607	rs57387518	46,169,039	52,077,517	5.908	1000				
			16	rs12919913	rs4149782	948,777	3,027,296	2.079	1000				
			16	rs34286603	rs13330387	7,047,669	19,193,876	12.146	1000				
			21	rs11088650	rs430098	14,386,013	20,258,518	5.872	1000				
			22	rs140378	rs5746636	16,877,135	18,896,301	2.019	980				
180	PK51	N	2	rs71426727	rs1227131	165,389,424	168,412,798	3.023	1000				
			2	rs2099159	rs58279250	209,511,062	210,880,219	1.369	830				
			2	rs207942	rs34577239	217,043,965	242,775,910	25.731	1000				
			3	rs1542531	rs7622802	162,256,364	163,350,517	1.094	704				
			5	rs17285919	rs585950	86,908,920	88,099,671	1.191	699				

			5	rs249496	rs10428561	102,201,590	112,007,180	9.806	1000				
			6	rs7744623	rs12198996	15,062,971	25,166,010	10.103	1000				
			6	rs6902872	rs166716	135,087,023	137,403,527	2.317	1000				
			6	rs6935507	rs35498687	149,814,584	154,683,383	4.869	1000				
			8	rs2956917	rs1020347	1,514,749	3,101,584	1.587	1000				
			9	rs455178	rs72739012	106,498,091	107,788,943	1.291	768				
			12	rs10744648	rs3782744	586,407	3,690,249	3.104	1000				
			22	rs17486173	rs2138150	43,153,703	45,658,035	2.504	1000				
181	PK53	N	2	rs17807135	rs219221	203,633,245	205,008,562	1.375	673				
			4	rs6821790	rs11734660	6,029,163	7,093,834	1.065	749				
			5	rs1030660	rs11958187	145,312,933	146,435,163	1.122	610				
			6	rs62426816	rs1343236	79,376,572	82,018,209	2.642	560				
			6	rs9444453	rs12190865	87,599,030	89,737,500	2.138	518				
			6	rs13195365	rs9499111	100,627,492	102,958,132	2.331	800				
			8	rs1449546	rs35135295	76,584,886	77,761,425	1.177	474				
			11	rs1545896	rs61882415	15,801,938	17,861,400	2.059	750				
			22	rs130503	rs3859840	34,338,666	35,359,702	1.021	472				
182	PK57	Y	2	rs6744641	rs1809636	88,760,177	95,421,697	6.662	554				
			2	rs9678800	rs6875	96,009,318	98,372,980	2.364	626				

			4	rs7661775	rs3017728	102,525,045	105,508,676	2.984	880				
			7	rs4096669	rs10239141	118,352,199	132,098,335	13.746	1000				
			11	rs1504677	rs17151327	55,475,356	56,653,038	1.178	872				
			12	rs10770633	rs4564422	20,372,499	25,417,418	5.045	1000				
			16	rs13330911	rs12921822	89,561	7,052,774	6.963	1000				
			16	rs4553646	rs10468296	31,536,419	48,484,530	16.948	727				
			17	rs9895150	rs8065270	45,307,928	75,246,941	29.939	1000				
			18	rs34431440	rs2615867	3,354,822	4,440,419	1.086	744				
			18	rs1728030	rs17799201	10,843,729	28,610,506	17.767	1000				
			18	rs2051308	rs17827157	48,194,641	66,902,015	18.707	1000				
			19	rs11669411	rs10423261	7,797,612	15,006,487	7.209	1000				
			19	rs2915921	rs10405242	22,714,327	24,374,419	1.66	730				
183	PK59	Y	1	rs859380	rs10489474	175,476,995	176,543,149	1.066	126				
			1	rs9425456	rs10913870	177,321,939	179,687,646	2.366	300				
			1	rs10911492	rs10911679	183,985,203	185,086,991	1.102	144				
			1	rs16825335	rs2495420	186,387,744	187,768,072	1.38	284				
			1	rs10922049	rs435062	196,172,936	197,949,082	1.776	400				
			3	rs17352589	rs11919883	101,881,100	102,980,751	1.1	380				
			19	rs4933000	rs12327766	22,730,026	23,919,307	1.189	349				
			1	rs6780	rs2279	149,99	151,82	1.829	348				

				7996	503	5,265	4,089						
184	PK60	N	No HBD across all affecteds. Splitting: Br2 and 3 have ID + epilepsy, Br1 doesn't										
			2	rs12465363	rs34577239	239,751,204	242,775,910	3.025	1000				
			6	rs7356805	rs9490809	112,266,347	123,869,607	11.603	1000				
			9	rs10977977	rs1182665	9,789,362	19,736,871	9.948	1000				
			9	rs991557	rs7043029	81,290,543	104,123,762	22.833	1000				
			13	rs9318316	rs484312	24,684,163	30,601,799	5.918	1000				
			16	rs13330911	rs9938377	89,561	5,861,006	5.771	1000				
			18	rs2552488	rs7407770	38,759,157	57,598,173	18.839	1000				
			7	rs6974327	rs4993699	64,917,900	66,887,996	1.97	728				
185	PK61	N	3	rs6785832	rs13071976	49,224,316	52,139,910	2.916	808				
			6	rs7770648	rs17080517	110,669,620	119,032,673	8.363	1000				
			10	rs10885372	rs35229952	114,436,559	122,407,822	7.971	1000				
			18	rs12458775	rs12959896	13,432,341	14,937,620	1.505	544				
186	PK62	Y	11	rs3793948	rs7930900	83,556,174	89,918,373	6.362	2500	**TMEM135**	21	9(9)	
			11	rs9645608	rs11603266	90,380,573	91,034,922	0.654	2175				
187	PK63	N	3	rs9810201	rs1807193	192,811,289	194,238,390	1.427	925				
			8	rs6989976	rs1481513	76,230,087	79,008,608	2.779	958				

			8	rs4476968	rs530696	92,241,460	97,140,193	4.899	1000				
			8	rs1458564	rs10216995	50,415,120	51,549,845	1.135	593				
			12	rs11065716	rs1426434	111,207,445	115,666,362	4.459	1000				
188	PK69	Y	6	rs9322787	rs6934122	104,520,286	105,997,617	1.477	734				
			6	rs352849	rs7748253	109,504,086	110,701,451	1.197	682				
189	PK70	N	2	rs12471820	rs17189126	43,333,993	53,283,528	9.95	1000				
			2	rs13411115	rs17040266	56,594,576	70,789,532	14.195	1000				
			4	rs2920180	rs2276877	4,215,172	5,851,343	1.636	1000				
			4	rs9312239	rs7686607	179,962,359	190,921,709	10.959	1000				
			8	rs16905258	rs1566084	135,878,725	142,624,215	6.745	1000				
			10	rs16919480	rs265518	65,871,419	83,099,241	17.228	1000	**TET1**	3	2(2)	
			11	rs8186576	rs11227598	50,573,869	56,000,451	5.427	607				
			11	rs1192923	rs10898904	69,504,277	72,910,772	3.406	1000				
			11	rs10789856	rs10790240	111,875,773	118,009,279	6.134	1000				
			11	rs10893056	rs2510723	123,531,495	127,043,856	3.152	1000				
			12	rs10844387	rs10876149	33,073,853	39,228,083	6.154	842				
			14	rs4982270	rs4900881	35,880,818	49,610,464	13.73	1000				
			15	rs2433283	rs17238566	50,490,078	56,344,287	5.854	1000				
			17	rs67049855	rs11078773	5,841,351	8,904,563	3.063	1000				

			17	rs8078399	rs8079136	38,224,498	53,907,805	15.683	1000				
			21	rs28564821	rs3747011	42,950,560	48,084,628	5.134	1000				
Affymetrix 6.0													
190	RQ6	Y	1	rs10752632	rs16840028	159,875,455	161,990,106	2.115					
			8	rs9297745	rs16909781	127,669,856	139,912,945	12.243		**TG**	7	4(4)	
191	RQ10	N	8	rs1994035	rs7002473	79,003,651	87,128,258	8.125	2869	**VPS13B**	7	4(4)	
192	RQ11	N	7	rs6955220	rs16904	3,093,500	7,060,950	3.967	3000				
			9	rs1811647	rs7851161	13,354,300	35,244,061	21.89	3000				
			9	rs11793049	rs1555972	116,773,869	126,796,404	10.023	3000				

Table A- 3 Full list of homozygous sequence variants (or hemizygous for ChrX variants) identified for the Pakistani and Iranian families in the study, coordinates, frequency in control population and in silico predictions of the effects of the variant. Variants highlighted in bold are previously unreported in association with ID, and asterisks indicate genes/variants already reported by us in this cohort. Minor allele frequency of variants in a control population used the Exome Aggregation Consortium database (ExAC; Cambridge, MA; http://exac.broadinstitute.org; May 2016). Control Nonsense-mediated mRNA decay (NMD) was predicted using MutationTaster (www.mutationtaster.org). Missense changes were evaluated using Polyphen2 (PPH2), both HumDiv and HumVar, and using ConDel-an integrated analysis that uses prediction using five different algorithms, including PolyPhen2, SIFT, and Mutation Assessor[10] was also performed (D=deleterious; N=neutral). Potential effects on splicing called by MutationTaster were checked using Splice Predictor (www.fruitfly.org), and are shown as yes (Y), no (N), possibly (P), or not checked (-) for LoF mutations.

A. Families with single homozygous (or hemizygous for ChrX variants) mutation identified; no homozygous loss of function mutations reported in ExAC controls												
	Fam. ID	**Gene**	**Disease; MIM #**	**Cytoband**	**Coordinates: Genome cDNA Protein**	**ExAC MAF (S. Asian ExAC MAF)**	**NMD for LoF?**	**PPH2 HumDiv**	**PPH2 HumVar**	**ConDel**	**Splice Effect? (Y/N/P)**	**Ref**
1	**PJ7**	***LAMC1***		**1q25.3**	**Chr1:183083732A>T NM_002293.3: c.1088A>T p. His363Leu**	**0 (0)**	**-**	**0.997**	**0.995**	**N**	**N**	**N**

2[a]	AS104	*CNTN2*	FAME5; 615400	1q32.1	Chr1:205034976C>A (0/0) NM_005076.3: c.1755C>A p. Tyr585*	0 (0)	Y	-	-	-	-	1
3	AN49	*FMN2**	MRT47; 616193	1q43	chr1:240370627_240370628delinsG NM_020066.4: c.2515_2517delinsCG p. Thr839Argfs*48	0 (0)	Y	-	-	-	-	2
4[b]	AS114	*FMN2*	MRT47; 616193	1q43	Chr1:240370641_240370642delCT NM_020066.4: c.2529_2530delCT p. Ser844Cysfs*6	0 (0)	Y	-	-	-	-	2
5	PK22	*TPO**	TDH2A; 274500	2p25.3	Chr2:1497591G>T NM_000547.5: c.1786G>T p. Glu596*	6.64E-05 (4.89E-04)	Y	-	-	-	-	3
6	**PK113**	***AFF3***		**2q11.2**	**Chr2:100167973C>A** **NM_002285.2: c.3644G>T p. Gly1215Val**	**0 (0)**	**-**	**1**	**0.983**	**D**	**P**	**N**

7	PJ12	*LRP2**	222448	2q31.1	Chr2:170027106C>T NM_004525.2: c.11335G>A p. Asp3779Asn	2.47E-04 (9.08E-04)	-	1	1	D	N	4
8	IDH10	*HNMT**	MRT51; 616739	2q22.1	Chr2:138727776G>A NM_006895.2: c.179G>A p. Glu60Asp	1.66E-05 (0)	-	1	0.996	D	N	5
9[c]	**AS61**	***ABI2***		**2q33.2**	**Chr2:204245039C>T** **NM_005759.4: c.394C>T** **p. Arg132***	**0 (0)**	**Y**	**-**	**-**	**-**	**-**	**N**
10	PK27	*GMPPA*	AAMR; 615510	2q35	Chr2:220368860G>A NM_013335.3: c.545G>A p. Gly182Asp	0 (0)	-	0.996	0.988	D	N	6
11[d]	**PK34**	***UBA7***		**3p21.31**	**Chr3:49848458C>A** **NM_003335.2: c.1189G>T** **p. Glu397***	**8.01E-04 (5.45E-03)**	**Y**	**-**	**-**	**-**	**-**	**N**

12	**PJ9**	***TBC1D23***		**3q12.1-q12.2**	**Chr3:100037953delA** **NM_018309.4: c.1683delA p. Glu562Argfs*3**	**0 (0)**	**Y**	**-**	**-**	**-**	**-**	**N**
13	**AN50**	***ZBTB11***		**3q12.3**	**Chr3:101371344A>C** **NM_014415.3:** **c.2640T>G p.** **His880Gln**	**0 (0)**	**-**	**1**	**0.999**	**D**	**N**	**N**
14	IDH4	*ALG3*	CDG1D; 601110	3q27.1	Chr3:183960676A>G NM_005787.5: c.1079T>C p. Phe360Ser	0 (0)	-	1	1	D	N	7
15	PK1	*CC2D2A**	JBTS9; 612285	4p15.32	Chr4:15552604G>C NM_020785.1: c.2191+1G>C p. Val728Glufs*741	0 (0)	Y	-	-	-	Y	8,9
16	PK19	*BBS7*	BBS7; 615984	4q27	Chr4:122754463delT NM_176824.2: c.1599delA p. Lys534Asnfs*8	0 (0)	Y	-	-	-	-	10
17	AN14	*NSUN2**	MRT5; 611091	5p15.31	Chr5:6600308C>T NM_017755.5: c.2035G>A p. Gly679Arg	8.26E-06 (0)	-	1	0.998	D	N	11,12

18	AS6	*ERCC8*	CSA; 216400	5q12.1	Chr5:60214118T>C NM_000082.3: c.373A>G p. Lys125Glu	0 (0)	-	1	0.993	D	N	13
19	PK49	*TDP2*		6p22.3	Chr6:24651244_24651245insA NM_016614.2: c.860_861insT p. Leu287Phefs*5	0 (0)	Y	-	-	-	-	14
20	PK64	*TDP2*		6p22.3	Chr6:24658106G>A NM_016614.2: c.451C>T p. Gln151*	0 (0)	Y	-	-	-	-	14
21	**AS70**	***MAP3K7***		**6q15**	**Chr6:91254333C>T** **NM_145331.2: c.1229G>A p. Arg410Gln**	**1.65E-05 (1.21E-04)**	**-**	**0.994**	**0.885**	**N**	**N**	**N**
22	AS62	*PSPH**		7p11.2	Chr7:56088803C>T NM_004577.3 c.103G>A p. Ala35Thr	0 (0)	-	1	0.997	D	N	15

23	PK7	*PEX1*	PBD1A; 214100	7q21.2	Chr7:92136405T>G NM_000466.2: c.1706A>C p. His569Pro	8.25E-06 (0)	-	0.995	0.873	D	N	16
24	AN16	*AP4M1*	SPG50; 612936	7q22.1	Chr7:99700343_99700344delTA NM_004722.3: c.193_194delTA p. Tyr65Phefs*50	0 (0)	Y	-	-	-	-	17
25[e]	PK11	*TUSC3*	MRT7; 611093	8p22	Chr8:15480694C>T NM_006765.3: c.244C>T p. Arg82*	0 (0)	Y	-	-	-	-	18,19
26	AN21	*TUSC3**	MRT7: 611093	8p22	chr8:15,521,688_15,692,362del NM_178234 del exons 5-10	n.d.	Y	-	-	-	-	18,19
27	PJ11	*HGSNAT*	Sanfilippo syndrome type C; 252930	8p11.21	Chr8:43047544delG NM_152419.2: c.1348delG (p. Asp450Ilefs*32)	8.75E-6 (0)	Y	-	-	-	-	20

28 f	AS74	*HGSNAT*	Sanfilippo syndrome type C; 252930	8p11.21	del Exon 8 NM_152419.2: Exon 8 del p. Ile249Alafs*4	N/A	Y	-	-	-	-	20
29	PK54	*IMPA1*		8q21.13	Chr8:82583195_82583196delinsC NM_005536.3: c.544_545delinsG p. Leu182Valfs*54	0 (0)	Y	-	-	-	-	21
30	AN51	*VPS13B**	COH1; 216550	8q22.2	chr8:100728884_100779911 del NM_017890.4: exons 37-40del p. Gly2177Alafs*16	NA	Y	-	-	-	-	22
31	RQ10	*VPS13B**	COH1; 216550	8q22.2	Chr8:100732719delT NM_017890.4: c.6879delT p. Phe2293Leufs*24	0 (0)	Y	-	-	-	-	22
32	IDSG43	*VPS13B*	COH1; 216550	8q22.2	chr8:100883029_100883030insG NM_017890.4: c.11484_11485insG (p. Leu3830Thrfs*13	0 (0)	Y	-	-	-	-	22
33	RQ6	*TG**	TDH3; 274700	8q24.22	Chr8:134034365C>T NM_003235.4: c.7006C>T p. Arg2336*	1.65E-05 (0)	Y	-	-	-	-	3

34	AS19	*TRAPPC9* *	MRT13; 613192	8q24.3	Chr8:141407724G>A NM_031466.7: c.1423C>T p. Arg475*	8.24E-06 (6.06E-05)	Y	-	-	-	-	23,24,25
35 [g]	**AS66**	***MPDZ***	**HYC2; 615219**	**9p23**	**Chr9:13109992_13109995delGAAA** **NM_003829.4: c.5811_5814delTTTC p. Phe1938Leufs*3**	**0 (0)**	**Y**	**-**	**-**	**-**	**-**	**26**
36	PK121	*GNE*	Nonaka myopathy 605820; Sialuria, 269921	9p13.3	Chr9:36246342C>T NM_005476.4: c.302G>A p. Arg101His	8.25E-06 (0)	-	0.999	0.961	D	N	27
37 [h]	RQ7	*MAN1B1**	MRT15; 614202	9q34.3	Chr9:139996059G>A NM_016219.4: c.1189G>A p. Glu397Lys	0 (0)	-	1	0.996	D	N	28
38 [h]	RQ8	*MAN1B1**	MRT15; 614202	9q34.3	Chr9:139996059G>A NM_016219.4: c.1189G>A p. Glu397Lys	0 (0)	-	1	0.996	D	N	28
39 [h]	RQ9	*MAN1B1**	MRT15; 614202	9q34.3	Chr9:139996059G>A NM_016219.4: c.1189G>A p. Glu397Lys	0 (0)	-	1	0.996	D	N	28

40	AN43	*MAN1B1**	MRT15; 614202	9q34.3	Chr9:140000740G>A NM_016219.4: c.1418G>A p. Trp473*	0 (0)	Y	-	-	-	-	28
41 [i]	PK77	*MAN1B1*	MRT15; 614202	9q34.3	Chr9:140001244C>T NM_016219.4: c.1549C>T p. Arg517Cys	9.13E-05 (5.45E-04)	-	0.937	0.293	D	N	28
42	**AS22**	***MAPK8***		**10q11.22**	**Chr10:49628265_49628265delC** **NM_002750.3: c.518delC p. Arg174Glyfs*8**	**0 (0)**	**Y**	**-**	**-**	**-**	**-**	**N**
43	**PK70**	***TET1***		**10q21.3**	**Chr10:70451328G>T** **NM_030625.2: c.6168G>T p. Lys2056Asn**	**0 (0)**	**-**	**0.979**	**0.756**	**N**	**N**	**N**
44	**IDSG19**	**DMBT1**		**10q26.13**	**Chr10:124356559C>G** **NM_007329.2:** **c.2906C>G p.** **Thr969Arg**	**0 (0)**	**-**	**0.982**	**0.764**	**N**	**N**	**N**

45 j	**AS105**	***PIDD1***		**11p15.5**	**chr11:799453G>A** **NM_145886.3: c.2587C>T** **p. Gln863***	**0 (0)**	**N**	**-**	**-**	**-**	**-**	**N**
46 j	**AS110**	***PIDD1***		**11p15.5**	**chr11:799453G>A** **NM_145886.3: c.2587C>T** **p. Gln863***	**0 (0)**	**N**	**-**	**-**	**-**	**-**	**N**
47	**PK135**	***BDNF***		**11p14.1**	**Chr11:27679747A>G** **NM_001709.4: c.365T>C** **p. Met122Thr**	**1.65E-04 (1.21E-03)**	**-**	**0.956**	**0.549**	**N**	**N**	**N**
48	AS102	*PGAP2*	HPMRS3, MRT17 & MRT21; 614207	11p15.4	Chr11:3846254G>C NM_014489.3: c.713G>C p. Arg238Pro	4.95E-05 (3.64E-04)	-	0.894	0.613	N	N	29,30
49	PK62	*TMEM135*		11q14.2	Chr11:87032300delTT NM_022918.3: c.1304_1305delTT p. Phe435Serfs*31	0 (0)	N	-	-	-	-	17

50	**AS17**	***CCDC82***		**11q21**	**Chr11:96117377G>A** **NM_024725.3: c.535C>T** **p. Arg179***	**1.65E-05 (6.05E-05)**	**Y**	**-**	**-**	**-**	**-**	**N**
51	PJ5	*DCPS**	ARS; 616459	11q24.2	Chr11:126208295G>A NM_014026.3: c.636+1G>A 15 amino acid insertion	8.25E-06 (0)	N	-	-		Y	31,32
52	**PK94**	***USP44***		**12q22**	**Chr12:95927147_95927160delinsA** **NM_032147.3: c.873_886delinsT p. Leu291Phefs*8**	**0 (0)**	**Y**	**-**	**-**	**-**	**-**	**N**
53	AN20	*TCTN2**	JBTS24; 616654	12q24.31	Chr12:124191376C>T NM_024809.4: c.1873C>T p. Gln625*	0 (0)	Y	-	-	-	-	33
54	**AS23**	***SPATA13***		**13q12.12**	**Chr13:24797332C>T** **NM_001166271.1:** **c.265C>T p. Arg89Trp**	**4.71E-05 (0)**	**-**	**1**	**0.987**	**D**	**N**	**N**

55[k]	AS11	*ATP8A2*	CAMRQ4; 615268	13q12.13	Chr13:26343221_26343221delG NM_016529.4: c.2422delG p. Val808*	0 (0)	Y	-	-	-	-	34
56	**PK68**	***SLAIN1***		**13q22.3**	**chr13:78320722_78320722delA** **NM_144595.3: c.135delA p. Thr49Hisfs*96**	**0 (0)**	**Y**	**-**	**-**	**-**	**-**	**N**
57	**IDH3**	***TRAPPC 6B***		**14q21.1**	**Chr14:39628712G>A** **NM_177452.3: c.124C>T** **p. Arg42***	**1.65E-05 (6.06E-05)**	**Y**	**-**	**-**	**-**	**-**	**N**
58	**AN27**	***DUOX2***	**TDH6; 607200**	**15q21.1**	**Chr15:45396542C>T** **NM_014080.4: c.2356G>A p. Asp786Asn**	**5.40E-05 (3.24E-04)**	**-**	**0.983**	**0.552**	**D**	**N**	**N**

59	PK46	*PMM2*	CDG1A; 212065	16p13.2	Chr16:8898636A>G NM_000303.2: c.191A>G p. Tyr64Cys	0 (0)	-	0.998	0.953	D	N	35
60	**PK79**	***VPS35***	**PARK17; 614203**	**16q11.2**	**Chr16:46705735T>G** **NM_018206.4: c.1406A>C p. Gln469Pro**	**0 (0)**	**-**	**0.983**	**0.938**	**N**	**N**	**N**
61[1]	PK12	*GPT2*	MRT49; 616281	16q11.2	Chr16:46918865C>G NM_133443.2: c.238C>G p. Gln80Glu	0 (0)	-	0	0	N	P	36
62	PK85	*GPT2*	MRT49; 616281	16q11.2	Chr16:46956326C>T NM_133443.3: c.1210C>T p. Arg404*	3.33E-05 (1.89E-04)	Y	-	-	-	-	36
63	RQ1	*GPR56*	BFPP; 606854	16q21	chr16:57689345_57689346ins CCATGGAGGTGCT NM_005682.6: c.803_804insCCATGGAGGTGCT p. Leu269Hisfs*21	0 (0)	Y	-	-	-	-	37

64	PK58	*GPR56*	BFPP; 606854	16q21	Chr16:57693446C>T NM_005682.5: c.1426C>T p. Arg476*	0 (0)	Y	-	-	-	-	37
65	PK101	*GPR56*	BFPP; 606854	16q21	Chr16:57693480T>C NM_005682.5: c.1460T>C p. Leu487Pro	0 (0)	-	0.927	0.615	D	N	37
66	**AS72**	***FBXO31***	**MRT45; 615979**	**16q24.2**	**Chr16:87369054_87369059delinsT** **NM_024735.3: c.847_852delinsA p. Cys283Asnfs*81**	**0 (0)**	**Y**	**-**	**-**	**-**	**-**	**38**
67	PK65	*SLC13A5*	EEIE25; 615905	17p13.1	Chr17:6610075T>C NM_177550.3: c.254A>G p. Asp85Gly	0 (0)	-	0.999	0.989	N	N	39
68	PK13	*ALDH3A2*	SLS; 270200	17p11.2	Chr17:19561072G>A NM_000382.2: c.695G>A p. Gly232Glu	0 (0)	-	1	1	D	N	40

69	**PK91**	***SYNRG***	**607291**	**17q12**	**chr17:35896199C>T** **NM_007247.4: c.3548G>A p. Arg1183His**	**8.66E-06 (0)**	**-**	**1**	**0.998**	**D**	**N**	**N**
70	**IDSG29**	***FBXO47***		**17q12**	**Chr17:37107906G>C** **NM_001008777.2:** **c.544C>G p.** **Arg182Gly**	**0 (0)**	**-**	**1**	**0.995**	**D**	**N**	**N**
71	PK33	*NAGS*	NAGSD; 237310	17q21.31	chr17:42085083C>T NM_153006.2: c.1393C>T p. Arg465Trp	7.74E-05 (4.54E-04)	-	0.992	0.786	D	N	41,42
72	**AN53**	***SDK2***		**17q25.1**	**Chr17:71431658C>T** **NM_001144952.1:** **c.1126G>A p.** **Gly376Ser**	**5.82E-05 (0)**	**-**	**0.999**	**0.982**	**D**	**N**	**N**
73	PK31	*METTL23* *	MRT44; 615942	17q25	Chr17:74729449C>T NM_001080510: c.397C>T p. Gln133*	0 (0)	Y	-	-	-	-	43,44

74	**AN48**	***CAPS***		**19p13.3**	**Chr19:5914970G>A** **NM_004058.3: c.281G>A** **p. Arg94Gln**	**2.17E-04 (1.27E-04)**	**-**	**1**	**1**	**D**	**N**	**N**
75	AN37	*TRMT1*		19p13.2	Chr19:13223779_13223810del NM_017722.3: c.657_688del p. Gln219Hisfs*22	0 (0)	Y	-	-	-	-	17
76	**PJ3**	***MBOAT7***		**19q13.42**	**Chr19:54684518_54684524del** **NM_024298.3: c.820_826del** **p. Gly274Profs*47**	**0 (0)**	**Y**	**-**	**-**	**-**	**-**	**N**
77	**PJ6 branch 2**	***GPR64***		**Xp22.13**	**ChrX:19055718T>C** **NM_005756.3: c.191A>G** **p. Asn64Ser**	**8.45E-05 (0)**	**-**	**0.971**	**0.624**	**N**	**N**	**N**
78	PK59	*ATRX*	ATRX; 301040	Xq21.1	ChrX:76778780T>C NM_000489.3: c.6799A>G p. Thr2267Ala	0 (0)	-	0.905	0.521	D	N	45

79	PK72	*ATRX*	ATRX; 301040	Xq21.1	ChrX:76940077G>A NM_000489.3: c.671C>T p. Ala224Val	0 (0)	-	0.999	0.985	D	N	45
80	AS101	*ARHGEF 6*	MRX46; 300436	Xq26.3	ChrX:135829705A>G NM_004840.2: c.296T>C p. Val99Ala	1.14E-05 (9.91E-05)	-	1	1	D	N	46
81	**PK87**	***MAGEA1 1***		**Xq28**	**ChrX:148796213C>T** **NM_005366.4:** **c.169T>C** **p. Ser57Pro**	**0 (0)**	**-**	**0.984**	**0.86**	**n.d.**	**N**	**N**
82	PK55	*MECP2*	RTT; 312750; MRXS13; 300055	Xq28	ChrX:153296710C>T NM_004992.3: c.569G>A p. Arg190His	0 (0)	-	0.994	0.975	D	N	47,48

B. Families with two to four segregating variants identified, all passing criteria. Families with five or more variants are not shown.

1	**AS64**	***EIF3H***		**8q24.11**	**Chr8:117738410A>C** **NM_003756.2: c.134T>G** **p. Val45Gly**	**0 (0)**	**-**	**0.999**	**0.986**	**D**	**N**	**N**
1	**AS64**	***DEPTOR***		**8q24.12**	**Chr8:120940642T>G** **NM_022783.2: c.125T>G** **p. Leu42Arg**	**0.000207** **(0.001534)**	**-**	**0.925**	**0.649**	**N**	**N**	**N**
2	**AS8**	***CSMD2***		**1p35.1**	**Chr1:34002660G>A** **NM_052896.3: c.9409C>T p. Pro3137Ser**	**0 (0)**	**-**	**1**	**1**	**D**	**N**	**N**
2	AS8	*ALG13*	EIEE36; 300884	Xq23	ChrX:111003031C>G NM_001099922.2: c.3218C>G p. Pro1073Arg	0 (0)	-	0.986	0.741	N	N	49,50
3	**PK102**	***DNAH7***		**2q32.3**	**chr2:196799481C>T** **NM_018897.2: c.3305G>A p. Gly1102Glu**	**0 (0)**	**-**	**1**	**1**	**D**	**N**	**N**
3	**PK102**	***HTRA2***		**2p13.1**	**chr2:74757907G>T** **NM_013247.4: c.670G>T** **p. Asp224Tyr**	**0 (0)**	**-**	**1**	**1**	**D**	**N**	**N**

3	PK102	*DCX*	LISX1; 300067	Xq23	Chr12:13715989G>A NM_000555.3: c.158A>G p. Glu53Gly	0 (0)	-	0.001*	0.001*	N	N	51,52
4	**IDH5**	***ALDH18 A1***		**10q24.1**	**chr10:97366587A>C** **NM_002860.3: c.2320T>G p. Phe774Val**	**0 (0)**	**-**	**0.999**	**0.978**	**D**	**N**	**N**
4	**IDH5**	***C2orf70***		**2p23.3**	**chr2:26802188G>C** **NM_001105519.1: c.488G>C p. Trp163Ser**	**0 (0)**	**-**	**1**	**0.996**	**N**	**N**	**N**
5	**PK51**	***ALPI***		**2q37.1**	**Chr2:233322708G>T** **NM_001631.3: c.857G>T** **p. Gly286Val**	**0 (0)**	**-**	**1**	**0.999**	**D**	**N**	**N**
5	**PK51**	***ITFG2***		**12p13.33**	**Chr12:2929317G>T** **NM_018463.3: c.472G>T** **p. Glu158***	**0 (0)**	**Y**	**-**	**-**	**-**	**-**	**N**

5	**PK51**	***NEU4***		**2q37.3**	**Chr2:242757403G>A** **NM_080741.2: c.520G>A** **p. Gly174Ser**	**3.74E-05** **(9.05E-05)**	**-**	**1**	**0.993**	**D**	**N**	**N**
5[m]	PK51	*SCN1A*	EIEE8; 607208	2q24.3	chr2:166848561C>T NM_006920.4: c.5191G>A p. Asp1731Asn	8.23E-06 (0)	-	0.907	0.784	D	N	53
6	**AS109**	***CISH***		**3p21.2**	**Chr12:13715989G>A** **NM_013324.5: c.314C>T** **p. Thr105Met**	**3.48E-05** **(7.36E-05)**	**-**	**0.998**	**0.997**	**D**	**N**	**N**
6	**AS109**	***GATA5***		**20q13.33**	**Chr20:61039908G>A** **NM_080473.4:** **c.1178C>T** **p. Ala393Val**	**8.71E-06** **(0)**	**-**	**0.968**	**0.994**	**D**	**N**	**N**
7	**AS15**	***MYH8***		**17p13.1**	**Chr17:10305039G>A** **NM_002472.2:** **c.2752C>T p.** **Leu918Phe**	**1.65E-05** **(0)**	**-**	**0.594**	**0.177**	**D**	**N**	**N**

7	**AS15**	***MYH4***		**17p13.1**	**Chr17:10358578C>A** **NM_017533.2: c.2209G>T p. Gly737Cys**	**0 (0)**	**-**	**1**	**1**	**D**	**N**	**N**
8[n]	**IDH8**	***ABCC12***		**16q12.1**	**Chr16:48117841_48117842insA** **NM_033226.2: c.3971_3972insT p. Leu1325Profs*10**	**1.65E-05 (0)**	**Y**	**-**	**-**	**-**	**-**	**N**
8	**IDH8**	***TPSD1***		**16p13.3**	**Chr16:1306527G>C** **NM_012217.2: c.93G>C** **p. Gln31His**	**8.48E-06 (0)**	**-**	**0.993**	**0.797**	**N**	**N**	**N**
8	**IDH8**	***EPN2***		**17p11.2**	**Chr17:19235271G>A** **NM_014964.4: c.1517G>A p. Arg506Gln**	**4.94E-05 (2.42E-04)**	**-**	**1**	**0.961**	**N**	**N**	**N**

9	**PK60**	***NARS***		**18q21.3**	**Chr18:55268898G>A** **NM_004539.3: c.1633C>T p. Arg545Cys**	**2.47E-05 (6.06E-05)**	**-**	**0.984**	**0.67**	**D**	**N**	**N**
9	**PK60**	***SLC14A2***		**18q21.3**	**Chr18:43262375C>T** **NM_007163.3: c.2654C>T p. Pro885Leu**	**4.94E-05 (1.82E-04)**	**-**	**0.979**	**0.573**	**N**	**N**	**N**
C. Dominant mutations segregating in ID families												
1	PK76	*GRIN2B*	MRD6; 613970		Chr12:13715989G>A NM_000834.3: c.4183C>T p. Gln1395*	0 (0)	Y	-	-	-	-	54

a. Chr1:205034976C>A at *CNTN2* segregates fully, homozygous, in recessive pattern of inheritance, but is not within a significant (>1Mb) HBD region.
b. Affected individual in AS114 with *FMN2* mutation is a phenocopy, and genes/mutations for the other four affected individuals in this pedigree have yet to be established.
c. The homozygous nonsense mutation reported for family AS61 is located within a large HBD region shared between just two of the four affected individuals. Support for this gene comes from functional studies, including with *Abi2* null mice (Grove et al, 2004).
d. Two homozygotes for this variant are indicated in the ExAC South Asian population (N=8,251 individuals). N.B. affected individuals in PK34 have learning disability rather than ID. The South Asian cohort in ExAC were part of the Pakistan Risk of Myocardial Infarction Study (PROMIS), and would not have excluded subjects in this category.
e. Homozygous nonsense mutation in *TUSC3* is not in shared HBD region, and does not segregate with other affected members of family PK11.
f. The effect of deletion of *HGSNAT* exon 8 on the HGSNAT protein is predicted as p. Ile249Alafs*4.
g. The *MPDZ* mutation reported here is not within a significant region of HBD, and is present as homozygous in the affected mother and the first-born affected son. The affected brother and sister are both heterozygous, but the father is wild type homozygous, suggesting non paternity for the first child. Also, of note, a previous report has linked a biallelic Love with non-syndromic autosomal recessive hydrocephaly[26]. However, hydrocephaly is not present in family AS66.
h. Families RQ7, RQ8 and RQ9, sharing the same mutation in *MAN1B1*, all originate from the same region, however are not thought to be closely related.
i. PK77-9 is the sole affected member of the family carrying this homozygous *MAN1B1*, and genes/mutations for the other four affected individuals in this pedigree have yet to be established.
j. Families AS105 and AS110 both carry the same mutation in *PIDD1*, and share a common haplotype, however, these families are apparently not closely related.
k. Homozygous nonsense mutation identified in cerebellar ataxia, mental retardation and disequilibrium syndrome (CAMRQ4) was outside of shared HBD regions, and in only one affected family member, thus is a probable phenocopy.
l. Missense change in *GPT2* is also predicted to affect a splice donor site (BDGP (www.fruitfly.org) splice predictor score: WT: 0.78; mut: 0.58).
m. This variant is homozygous. Mutations have previously been reported in this gene in association with autosomal dominant EIEE8 (MIM 607208), GEFSP2 (MIM 604403), and have been reported *de novo* in autism spectrum disorders[55] and ID[56].

n. ExAC database indicated there are homozygous Love variants elsewhere in ABCC12.
n.d.= data not available for this variant.

References

1. Stegmann, E., Rein haler, E., Entail, S., El Tribe, M. A., Hemmed, M., El Nahass, N., Gaber, A. M., Fouad, A., Edris, S., Benet-Pages, A., Eck, S. H., Patrai, E., Mei, D., Brice, A., Lesage, S., Guerrini, R., Simpich, F., Strom, T. M., Simpich, A. Autosomal recessive cortical myoclonic tremor and epilepsy: association with a mutation in the potassium channel associated gene CNTN2. Brain 136: 1155-1160, 2013.
2. Law, R., Dixon-Salazar, T., Jerker, J., Cai, N., Abbasi, A. A., Zakie, M. S., Mittal, K., Gabriel, S. B., Rafiq, M. A., Khan, V., Nguyen, M., Ali, G., and 11 others. Biallelic truncating mutations in FMN2, encoding the actin-regulatory protein forming 2, cause no syndromic autosomal-recessive intellectual disability. Am. J. Hum. Genet. 95: 721-728, 2014.
3. Mittal K, Rafiq MA, Rafiullah R, Harripaul R, Ali A, Ayaz M, Aslam M, Naeem F, Amin-us-din M, Waqas A, So J, Rapp old GA, Vincent JB, Ayyub M. Mutations in the Genes for Thyroglobulin and Thyroid Peroxidase Cause Thyroid Dyshormonogenesis and Autosomal Recessive Intellectual Disability. J Hum Genet [in press]
4. Vasli N, Ahmed I, Mittal K, Ohadi M, Mikhailov A, Rafiq MA, Bhatti A, Carter MT, Andrade DM, Ayub M, Vincent JB, John P Identification of a homozygous missense mutation in LRP2 and a hemizygous missense mutation in TSPYL2 in a family with mild intellectual disability. Psychiatr Genet. 2016 Apr;26(2):66-73.
5. Heidari, A., Tongsook, C., Najafipour, R., Musante, L., Vasli, N., Garshasbi, M., Hu, H., Mittal, K., McNaughton, A. J. M., Sritharan, K., Hudson, M., Stehr, H., and 22 others. Mutations in the histamine N-methyltransferase gene, HNMT, are associated with nonsyndromic autosomal recessive intellectual disability. Hum. Molec. Genet. 24: 5697-5710, 2015.
6. Koehler, K., Malik, M., Mahmood, S., Giebelmann, S., Beetz, C., Hennings, J. C., Huebner, A. K., Grahn, A., Reunert, J., Nurnberg, G., Thiele, H., Altmuller, J., and 20 others. Mutations in GMPPA cause a glycosylation disorder characterized by intellectual disability and autonomic dysfunction. Am. J. Hum. Genet. 93: 727-734, 2013.
7. Korner, C., Knauer, R., Stephani, U., Marquardt, T., Lehle, L., von Figura, K. Carbohydrate deficient glycoprotein syndrome type IV: deficiency of dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase. EMBO J. 18: 6816-6822, 1999.
8. Noor, A., Windpassinger, C., Patel, M., Stachowiak, B., Mikhailov, A., Azam, M., Irfan, M., Siddiqui, Z. K., Naeem, F., Paterson, A. D., Lutfullah, M., Vincent, J. B., Ayub, M. CC2D2A, encoding a coiled-coil and C2 domain protein, causes autosomal-recessive mental retardation with retinitis pigmentosa. Am. J. Hum. Genet. 82: 1011-1018, 2008.
9. Gorden, N. T., Arts, H. H., Parisi, M. A., Coene, K. L. M., Letteboer, S. J. F., van Beersum, S. E. C., Mans, D. A., Hikida, A., Eckert, M., Knutzen, D., Alswaid, A. F., Ozyurek, H., and 23 others. CC2D2A is mutated in Joubert syndrome and interacts with the ciliopathy-associated basal body protein CEP290. Am. J. Hum. Genet. 83: 559-571, 2008.
10. Badano, J. L., Ansley, S. J., Leitch, C. C., Lewis, R. A., Lupski, J. R., Katsanis, N. Identification of a novel Bardet-Biedl syndrome protein, BBS7, that shares structural features with BBS1 and BBS2. Am. J. Hum. Genet. 72: 650-658, 2003.
11. Abbasi-Moheb, L., Mertel, S., Gonsior, M., Nouri-Vahid, L., Kahrizi, K., Cirak, S., Wieczorek, D., Motazacker, M. M., Esmaeeli-Nieh, S., Cremer, K., Weissmann, R., Tzschach, A., Garshasbi, M., Abedini, S. S., Najmabadi, H., Ropers, H. H., Sigrist, S. J., Kuss, A. W. Mutations in NSUN2 cause autosomal-recessive intellectual disability. Am. J. Hum. Genet. 90: 847-855, 2012.
12. Khan, M. A., Rafiq, M. A., Noor, A., Hussain, S., Flores, J. V., Rupp, V., Vincent, A. K., Malli, R., Ali, G., Khan, F. S., Ishak, G. E., Doherty, D., Weksberg, R., Ayub, M., Windpassinger, C., Ibrahim, S., Frye, M., Ansar, M., Vincent, J. B. Mutation in NSUN2, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability. Am. J. Hum. Genet. 90: 856-863, 2012.
13. Henning, K. A., Li, L., Iyer, N., McDaniel, L. D., Reagan, M. S., Legerski, R., Schultz, R. A., Stefanini, M., Lehmann, A. R., Mayne, L. V., Friedberg, E. C. The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIH. Cell 82: 555-564, 1995.

14. Gómez-Herreros F, Schuurs-Hoeijmakers JH, McCormack M, Greally MT, Rulten S, Romero-Granados R, Counihan TJ, Chaila E, Conroy J, Ennis S, Delanty N, Cortés-Ledesma F, de Brouwer AP, Cavalleri GL, El-Khamisy SF, de Vries BB, Caldecott KW. TDP2 protects transcription from abortive topoisomerase activity and is required for normal neural function. Nat Genet. 2014 May;46(5):516-21.
15. Vincent, J. B., Jamil, T., Rafiq, M. A., Anwar, Z., Ayaz, M., Hameed, A., Nasr, T., Naeem, F., Khattak, N. A., Carter, M., Ahmed, I., John, P., Wiame, E., Andrade, D. M., Schaftingen, E. V., Mir, A., Ayub, M. Phosphoserine phosphatase (PSPH) gene mutation in an intellectual disability family from Pakistan. Clin. Genet. 87: 296-298, 2015.
16. Reuber, B. E., Germain-Lee, E., Collins, C. S., Morrell, J. C., Ameritunga, R., Moser, H. W., Valle, D., Gould, S. J. Mutations in PEX1 are the most common cause of peroxisome biogenesis disorders. Nature Genet. 17: 445-448, 1997.
17. Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S. S., Chen, W., Hosseini, M., Behjati, F., Haas, S., Jamali, P., Zecha, A., Mohseni, M., and 33 others. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature 478: 57-63, 2011.
18. Garshasbi, M., Hadavi, V., Habibi, H., Kahrizi, K., Kariminejad, R., Behjati, F., Tzschach, A., Najmabadi, H., Ropers, H. H., Kuss, A. W. A defect in the TUSC3 gene is associated with autosomal recessive mental retardation. Am. J. Hum. Genet. 82: 1158-1164, 2008.
19. Molinari, F., Foulquier, F., Tarpey, P. S., Morelle, W., Boissel, S., Teague, J., Edkins, S., Futreal, P. A., Stratton, M. R., Turner, G., Matthijs, G., Gecz, J., Munnich, A., Colleaux, L. Oligosaccharyltransferase-subunit mutations in nonsyndromic mental retardation. Am. J. Hum. Genet. 82: 1150-1157, 2008.
20. Fan, X., Zhang, H., Zhang, S., Bagshaw, R. D., Tropak, M. B., Callahan, J. W., Mahuran, D. J. Identification of the gene encoding the enzyme deficient in mucopolysaccharidosis IIIC (Sanfilippo disease type C). Am. J. Hum. Genet. 79: 738-744, 2006.
21. Figueiredo T, Melo US, Pessoa AL, Nobrega PR, Kitajima JP, Rusch H, Vaz F, Lucato LT, Zatz M, Kok F, Santos S. A homozygous loss-of-function mutation in inositol monophosphatase 1 (IMPA1) causes severe intellectual disability. Mol Psychiatry. 2015 Sep 29. [Epub ahead of print]
22. Kolehmainen, J., Black, G. C. M., Saarinen, A., Chandler, K., Clayton-Smith, J., Traskelin, A.-L., Perveen, R., Kivitie-Kallio, S., Norio, R., Warburg, M., Fryns, J.-P., de la Chapelle, A., Lehesjoki, A.-E. Cohen syndrome is caused by mutations in a novel gene, COH1, encoding a transmembrane protein with a presumed role in vesicle-mediated sorting and intracellular protein transport. Am. J. Hum. Genet. 72: 1359-1369, 2003.
23. Mir, A., Kaufman, L., Noor, A., Motazacker, M. M., Jamil, T., Azam, M., Kahrizi, K., Rafiq, M. A., Weksberg, R., Nasr, T., Naeem, F., Tzschach, A., Kuss, A. W., Ishak, G. E., Doherty, D., Ropers, H. H., Barkovich, A. J., Najmabadi, H., Ayub, M., Vincent, J. B. Identification of mutations in TRAPPC9, which encodes the NIK- and IKK-beta-binding protein, in nonsyndromic autosomal-recessive mental retardation. Am. J. Hum. Genet. 85: 909-915, 2009.
24. Mochida, G. H., Mahajnah, M., Hill, A. D., Basel-Vanagaite, L., Gleason, D., Hill, R. S., Bodell, A., Crosier, M., Straussberg, R., Walsh, C. A. A truncating mutation of TRAPPC9 is associated with autosomal-recessive intellectual disability and postnatal microcephaly. Am. J. Hum. Genet. 85: 897-902, 2009.
25. Philippe, O., Rio, M., Carioux, A., Plaza, J.-M., Guigue, P., Molinari, F., Boddaert, N., Bole-Feysot, C., Nitschke, P., Smahi, A., Munnich, A., Colleaux, L. Combination of linkage mapping and microarray-expression analysis identifies NF-kappa-B signaling defect as a cause of autosomal-recessive mental retardation. Am. J. Hum. Genet. 85: 903-908, 2009.
26. Al-Dosari, M. S., Al-Owain, M., Tulbah, M., Kurdi, W., Adly, N., Al-Hemidan, A., Masoodi, T. A., Albash, B., Alkuraya, F. S. Mutation in MPDZ causes severe congenital hydrocephalus. J. Med. Genet. 50: 54-58, 2013.
27. Seppala, R., Lehto, V.-P., Gahl, W. A. Mutations in the human UDP-N-acetylglucosamine 2-epimerase gene define the disease sialuria and the allosteric site of the enzyme. Am. J. Hum. Genet. 64: 1563-1569, 1999.

28. Rafiq, M. A., Kuss, A. W., Puettmann, L., Noor, A., Ramiah, A., Ali, G., Hu, H., Kerio, N. A., Xiang, Y., Garshasbi, M., Khan, M. A., Ishak, G. E., and 12 others. Mutations in the alpha 1,2-mannosidase gene, MAN1B1, cause autosomal-recessive intellectual disability. Am. J. Hum. Genet. 89: 176-182, 2011. Note: Erratum: Am. J. Hum. Genet. 89: 348 only, 2011.
29. Krawitz, P. M., Murakami, Y., Riess, A., Hietala, M., Kruger, U., Zhu, N., Kinoshita, T., Mundlos, S., Hecht, J., Robinson, P. N., Horn, D. PGAP2 mutations, affecting the GPI-anchor-synthesis pathway, cause hyperphosphatasia with mental retardation syndrome. Am. J. Hum. Genet. 92: 584-589, 2013.
30. Hansen, L., Tawamie, H., Murakami, Y., Mang, Y., Rehman, S., Buchert, R., Schaffer, S., Muhammad, S., Bak, M., Nothen, M. M., Bennett, E. P., Maeda, Y., Aigner, M., Reis, A., Kinoshita, T., Tommerup, N., Baig, S. M., Abou Jamra, R. Hypomorphic mutations in PGAP2, encoding a GPI-anchor-remodeling protein, cause autosomal-recessive intellectual disability. Am. J. Hum. Genet. 92: 575-583, 2013.
31. Ahmed, I., Buchert, R., Zhou, M., Jiao, X., Mittal, K., Sheikh, T. I., Scheller, U., Vasli, N., Rafiq, M. A., Brohi, M. Q., Mikhailov, A., Ayaz, M., and 11 others. Mutations in DCPS and EDC3 in autosomal recessive intellectual disability indicate a crucial role for mRNA decapping in neurodevelopment. Hum. Mol. Genet. 24: 3172-3180, 2015.
32. Ng, C. K. L., Shboul, M., Taverniti, V., Bonnard, C., Lee, H., Eskin, A., Nelson, S. F., Al-Raqad, M., Altawalbeh, S., Seraphin, B., Reversade, B. Loss of the scavenger mRNA decapping enzyme DCPS causes syndromic intellectual disability with neuromuscular defects. Hum. Molec. Genet. 24: 3163-3171, 2015.
33. Sang, L., Miller, J. J., Corbit, K. C., Giles, R. H., Brauer, M. J., Otto, E. A., Baye, L. M., Wen, X., Scales, S. J., Kwong, M., Huntzicker, E. G., Stakianos, M. K., and 20 others. Mapping the NPHP-JBTS-MKS protein network reveals ciliopathy disease genes and pathways. Cell 145: 513-528, 2011.
34. Onat, O. E., Gulsuner, S., Bilguvar, K., Nazli Basak, A., Topaloglu, H., Tan, M., Tan, U., Gunel, M., Ozcelik, T. Missense mutation in the ATPase, aminophospholipid transporter protein ATP8A2 is associated with cerebellar atrophy and quadrupedal locomotion. Eur. J. Hum. Genet. 21: 281-285, 2013.
35. Matthijs, G., Schollen, E., Pardon, E., Veiga-Da-Cunha, M., Jaeken, J., Cassiman, J.-J., Van Schaftingen, E. Mutations in PMM2, a phosphomannomutase gene on chromosome 16p13, in carbohydrate-deficient glycoprotein type I syndrome (Jaeken syndrome). Nature Genet. 16: 88-92, 1997.
36. Celis, K., Shuldiner, S., Haverfield, E. V., Cappell, J., Yang, R., Gong, D.-W., Chung, W. K. Loss of function mutation in glutamic pyruvate transaminase 2 (GPT2) causes developmental encephalopathy. J. Inherit. Metab. Dis. 38: 941-948, 2015.
37. Piao, X., Hill, R. S., Bodell, A., Chang, B. S., Basel-Vanagaite, L., Straussberg, R., Dobyns, W. B., Qasrawi, B., Winter, R. M., Innes, A. M., Voit, T., Ross, M. E., Michaud, J. L., Descarie, J.-C., Barkovich, A. J., Walsh, C. A. G protein-coupled receptor-dependent development of human frontal cortex. Science 303: 2033-2036, 2004.
38. Mir, A., Sritharan, K., Mittal, K., Vasli, N., Araujo, C., Jamil, T., Rafiq, M. A., Anwar, Z., Mikhailov, A., Rauf, S., Mahmood, H., Shakoor, A., Ali, S., So, J., Naeem, F., Ayub, M., Vincent, J. B. Truncation of the E3 ubiquitin ligase component FBXO31 causes non-syndromic autosomal recessive intellectual disability in a Pakistani family. Hum. Genet. 133: 975-984, 2014.
39. Thevenon, J., Milh, M., Feillet, F., St-Onge, J., Duffourd, Y., Juge, C., Roubertie, A., Heron, D., Mignot, C., Raffo, E., Isidor, B., Wahlen, S., and 18 others. Mutations in SLC13A5 cause autosomal-recessive epileptic encephalopathy with seizure onset in the first days of life. Am. J. Hum. Genet. 95: 113-120, 2014.
40. Sillen, A., Anton-Lamprecht, I., Braun-Quentin, C., Kraus, C. S., Sayli, B. S., Ayuso, C., Jagell, S., Kuster, W., Wadelius, C. Spectrum of mutations and sequence variants in the FALDH gene in patients with Sjogren-Larsson syndrome. Hum. Mutat. 12: 377-384, 1998.
41. Caldovic, L., Morizono, H., Pangiao, M. G., Cheng, S. F., Packman, S., Tuchman, M. Null mutations in the N-acetylglutamate synthase gene associated with acute neonatal disease and hyperammonemia. Hum. Genet. 112: 364-368, 2003.
42. Haberle, J., Schmidt, E., Pauli, S., Kreuder, J. G., Plecko, B., Galler, A., Wermuth, B., Harms, E., Koch, H. G. Mutation analysis in patients with N-acetylglutamate synthase deficiency. Hum. Mutat. 21: 593-597, 2003.

43. Bernkopf, M., Webersinke, G., Tongsook, C., Koyani, C. N., Rafiq, M. A., Ayaz, M., Muller, D., Enzinger, C., Aslam, M., Naeem, F., Schmidt, K., Gruber, K., Speicher, M. R., Malle, E., Macheroux, P., Ayub, M., Vincent, J. B., Windpassinger, C., Duba, H.-C. Disruption of the methyltransferase-like 23 gene METTL23 causes mild autosomal recessive intellectual disability. Hum. Molec. Genet. 23: 4015-4023, 2014.
44. Reiff, R. E., Ali, B. R., Baron, B., Yu, T. W., Ben-Salem, S., Coulter, M. E., Schubert, C. R., Hill, R. S., Akawi, N. A., Al-Younes, B., Kaya, N., Evrony, G. D., and 10 others. METTL23, a transcriptional partner of GABPA, is essential for human cognition. Hum. Molec. Genet. 23: 3456-3466, 2014.
45. Gibbons, R. J., Picketts, D. J., Villard, L., Higgs, D. R. Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). Cell 80: 837-845, 1995.
46. Kutsche, K., Yntema, H., Brandt, A., Jantke, I., Nothwang, H. G., Orth, U., Boavida, M. G., David, D., Chelly, J., Fryns, J.-P., Moraine, C., Ropers, H.-H., Hamel, B. C. J., van Bokhoven, H., Gal, A. Mutations in ARHGEF6, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation. Nature Genet. 26: 247-250, 2000.
47. Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., Zoghbi, H. Y. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nature Genet. 23: 185-188, 1999.
48. Meloni, I., Bruttini, M., Longo, I., Mari, F., Rizzolio, F., D'Adamo, P., Denvriendt, K., Fryns, J.-P., Toniolo, D., Renieri, A. A mutation in the Rett syndrome gene, MECP2, causes X-linked mental retardation and progressive spasticity in males. Am. J. Hum. Genet. 67: 982-985, 2000.
49. Timal, S., Hoischen, A., Lehle, L., Adamowicz, M., Huijben, K., Sykut-Cegielska, J., Paprocka, J., Jamroz, E., van Spronsen, F. J., Korner, C., Gilissen, C., Rodenburg, R. J., Eidhof, I., Van den Heuvel, L., Thiel, C., Wevers, R. A., Morava, E., Veltman, J., Lefeber, D. J. Gene identification in the congenital disorders of glycosylation type I by whole-exome sequencing. Hum. Molec. Genet. 21: 4151-4161, 2012.
50. de Ligt, J., Willemsen, M. H., van Bon, B. W. M., Kleefstra, T., Yntema, H. G., Kroes, T., Vulto-van Silfhout, A. T., Koolen, D. A., de Vries, P., Gilissen, C., del Rosario, M., Hoischen, A., Scheffer, H., de Vries, B. B. A., Brunner, H. G., Veltman, J. A., Vissers, L. E. L. M. Diagnostic exome sequencing in persons with severe intellectual disability. New Eng. J. Med. 367: 1921-1929, 2012.
51. des Portes, V., Pinard, J. M., Billuart, P., Vinet, M. C., Koulakoff, A., Carrie, A., Gelot, A., Dupuis, E., Motte, J., Berwald-Netter, Y., Catala, M., Kahn, A., Beldjord, C., Chelly, J. A novel CNS gene required for neuronal migration and involved in X-linked subcortical laminar heterotopia and lissencephaly syndrome. Cell 92: 51-61, 1998.
52. Gleeson, J. G., Allen, K. M., Fox, J. W., Lamperti, E. D., Berkovic, S., Scheffer, I., Cooper, E. C., Dobyns, W. B., Minnerath, S. R., Ross, M. E., Walsh, C. A. Doublecortin, a brain-specific gene mutated in human X-linked lissencephaly and double cortex syndrome, encodes a putative signaling protein. Cell 92: 63-72, 1998.
53. Claes, L., Del-Favero, J., Ceulemans, B., Lagae, L., Van Broeckhoven, C., De Jonghe, P. *De novo* mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy. Am. J. Hum. Genet. 68: 1327-1332, 2001.
54. Endele, S., Rosenberger, G., Geider, K., Popp, B., Tamer, C., Stefanova, I., Milh, M., Kortum, F., Fritsch, A., Pientka, F. K., Hellenbroich, Y., Kalscheuer, V. M., and 16 others. Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nature Genet. 42: 1021-1026, 2010.
55. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, Rieder MJ, Nickerson DA, Bernier R, Fisher SE, Shendure J, Eichler EE. Exome sequencing in sporadic autism spectrum disorders identifies severe *de novo* mutations. Nat Genet. 2011 Jun;43(6):585-9.
56. Fry AE, Rees E, Thompson R, Mantripragada K, Blake P, Jones G, Morgan S, Jose S, Mugalaasi H, Archer H, McCann E, Clarke A, Taylor C, Davies S, Gibbon F, Te Water Naude J, Hartley L, Thomas G, White C, Natarajan J, Thomas RH, Drew C, Chung SK, Rees MI, Holmans P, Owen MJ, Kirov G, Pilz DT, Kerr MP.

Pathogenic copy number variants and SCN1A mutations in patients with intellectual disability and childhood-onset epilepsy. BMC Med Genet. 2016 Apr 26;17(1):34.

Table A- 4 CNVs called as either pathogenic or as variants of unknown significance. **CNVs were also called using Genotyping Console (Affymetrix platform) and PennCNV (all arrays).**

CNV call	Family/members	hg19 coordinates	Size (Kb)	CN	Confidence	Genes	Comments	Overlap with DGV controls
Genome Suite: CNVPartition, All Illumina Core Exome arrays								
Pathogenic	PK17-45	chr7:94,631-2,752,932	2658	1	841	PARTIAL 7P MONOSOMY	apparently *de novo*	18%
Variant of unknown significance	PK68-1, 3 & 4	chrX:22,918,867-23,113,268	194	1	145	DDX53/PTCHD1AS	segregation not checked	100%
Variant of unknown significance	PK73-5	chr1:243471416-245431939	1961	3	159	many	1q44 trisomy	
Pathogenic	PK73-5	chr1:245,448,886-249,212,402	3764	1	2269	many	1q44 monosomy	21%
Pathogenic	PK73-8	chr9:119,646,682-119,803,933	157	1	117	ASTN2	exons 5-9	100%
Pathogenic	PK78-7	chr15:22,784,095-28,525,245	5741	3	6100	UBE3A; HERC2; GABR cluster; many others		19%
Pathogenic	PK93-4	chrX:3,555,302-154,780,283	151225	1	6548	many	TURNER'S SYNDROME	
Variant of unknown significance	PK95-7	chr15:63,494,159-63,544,106	50	0	143	RAB8B	exons 2 & 3	gains only
Pathogenic	PK117-5	chr2:116,583,565-124,954,598	8371	1	3053	GLI2; CNTNAP5	See Greally et	12%

							al, 2014	
Variant of unknown significance	PK117-3	chr2:30,670,973-31,244,635	574	3	620	LCLAT1; CAPN13; GALNT14		5%
Variant of unknown significance	AS66-3	chrX:83,128,103-83,130,202	2.1	0	509	CYLC1		0%
Pathogenic	ZA12-4	chr22:18,917,748-21,459,251	2542	1	702	many	22q11.21 DEL SYNDROME	
Pathogenic	AN17-10 & 11	chr2:50,228,332-50,786,770	558	1	142	NRXN1		14%
Pathogenic	IDSG30 VI:4	chr1:861349-1565891	705	3	315	many	1p36.33 dup	33%
Chromosome Analysis Suite (ChAS): All CytoScanHD arrays								
Pathogenic	PK28-2, 5 & 7 (males)	chrX:8127578-14862487	6735	2	8152	Many	In all 3 affecteds (male); maternal inheritance	10%
Genotyping Console: Affymetrix 250K NspI								
Pathogenic	PK8-1	chrX:99,545,384-100,148,449	603	1		PCDH19 and 6 others	EIEE9; MIM 300088	Vincent et al, 2012[a]

[a]Vincent AK, Noor A, Janson A, Minassian BA, Ayub M, Vincent JB, Morel CF. Identification of genomic deletions spanning the PCDH19 gene in two unrelated girls with intellectual disability and seizures. *Clin Genet*. 2012 82(6):540-545.

Table A- 5 Gene Ontology Pathway analysis

GO Term	Description	P-value	FDR q-value	Enrichment	GO Group Size	Input Genes Size	Count of overlapping genes
GO:0009225	nucleotide-sugar metabolic process	0.00000106	0.0152	26.98	33	103	5
GO:0006490	oligosaccharide-lipid intermediate biosynthetic process	0.00000122	0.00874	47.48	15	103	4
GO:0006488	dolichol-linked oligosaccharide biosynthetic process	0.00000122	0.00582	47.48	15	103	4
GO:1901137	carbohydrate derivative biosynthetic process	0.00000174	0.00622	5.54	386	103	12
GO:0044710	single-organism metabolic process	0.0000064	0.0183	1.86	4210	103	44
GO:0006486	protein glycosylation	0.00000729	0.0174	6.79	236	103	9
GO:0043413	macromolecule glycosylation	0.00000729	0.0149	6.79	236	103	9
GO:0070085	glycosylation	0.00000953	0.0171	6.57	244	103	9
GO:0043436	oxoacid metabolic process	0.0000131	0.0208	3.3	917	103	17
GO:0044723	single-organism carbohydrate metabolic process	0.0000143	0.0204	4.16	557	103	13
GO:0005975	carbohydrate metabolic process	0.0000158	0.0205	3.84	649	103	14
GO:0006082	organic acid metabolic process	0.0000166	0.0198	3.24	934	103	17
GO:0044699	single-organism process	0.0000217	0.0238	1.32	11051	103	82
GO:0006590	thyroid hormone generation	0.0000365	0.0373	44.52	12	103	3
GO:0044711	single-organism biosynthetic process	0.0000407	0.0388	2.79	1214	103	19
GO:0044281	small molecule metabolic process	0.0000426	0.0381	2.45	1672	103	23
GO:0006047	UDP-N-acetylglucosamine metabolic process	0.0000599	0.0504	38.16	14	103	3
GO:0042403	thyroid hormone metabolic process	0.0000913	0.0726	33.39	16	103	3

Table A- 6 Gene List for anatomic/temporal transcription analyses.

Gene	Category
ALG13	Glycosylation
ALG3	Glycosylation
DPAGT1	Glycosylation
GMPPA	Glycosylation
GMPPB	Glycosylation
LMAN2L	Glycosylation
MAN1B1	Glycosylation
PDE4D	Glycosylation
PGM3	Glycosylation
PMM2	Glycosylation
SLC39A8	Glycosylation
SRD5A3	Glycosylation
TMEM199	Glycosylation
XYLT1	Glycosylation
ACO2	Hormones and Metabolism
ACTB	Hormones and Metabolism
CACNA1D	Hormones and Metabolism
DUOX2	Hormones and Metabolism
GCDH	Hormones and Metabolism
GNE	Hormones and Metabolism
GPT2	Hormones and Metabolism
HGSNAT	Hormones and Metabolism
MMACHC	Hormones and Metabolism
NAGS	Hormones and Metabolism
NARS2	Hormones and Metabolism
PAH	Hormones and Metabolism
PCDH19	Hormones and Metabolism
PDHX	Hormones and Metabolism
PEX1	Hormones and Metabolism
PSPH	Hormones and Metabolism
SLC16A1	Hormones and Metabolism
TG	Hormones and Metabolism
TPO	Hormones and Metabolism
TUSC3	Hormones and Metabolism
WDR73	Hormones and Metabolism
ABI2	None
ADAT3	None
AFF3	None
ALDH3A2	None

ANK3	None
AP4M1	None
ARHGEF6	None
ATRX	None
BBS7	None
BDNF	None
C12ORF4	None
CAPS	None
CC2D2A	None
CCDC82	None
CNTN2	None
CRADD	None
CRBN	None
DCPS	None
DCX	None
DHCR7	None
DMBT1	None
EDC3	None
ERCC8	None
EXTL3	None
FBXO31	None
FBXO47	None
FMN2	None
GPR56	None
GPR64	None
GRIK2	None
HERC2	None
HNMT	None
IMPA1	None
KIAA1033	None
KPTN	None
LAMC1	None
LINS1	None
LRP2	None
MAGEA11	None
MAP3K7	None
MAPK8	None
MBOAT7	None
MECP2	None
MED23	None
METTL23	None
MPDZ	None

NDST1	None
NSUN2	None
PGAP1	None
PGAP2	None
PIDD1	None
PRSS12	None
PRSS48	None
SDK2	None
SLAIN1	None
SLC13A5	None
SLC6A17	None
SPATA13	None
ST3GAL3	None
SYNRG	None
TAF2	None
TBC1D23	None
TCTN2	None
TDP2	None
TECR	None
TET1	None
TMEM135	None
TRAPPC6B	None
TRAPPC9	None
TRMT1	None
TTI2	None
UBA7	None
USP44	None
VPS13B	None
VPS35	None
ZBTB11	None

Table A- 7 Top anatomical regions for ID gene expression

Organ Name	Annotated	Significant	Expected	Fold Enrichment	p-value	Bonferonni adjusted
frontal cortex	11864	84	60.16	1.4	0.0000000202	0.0000003434
pigmented layer of retina	11983	83	60.76	1.37	0.000000171	0.000002907
metanephros	11552	80	58.57	1.37	0.000000981	0.000016677
neocortex	11928	81	60.48	1.34	0.00000183	0.00003111
cerebral cortex	12766	84	64.73	1.3	0.00000242	0.00004114
liver	12015	81	60.92	1.33	0.00000281	0.00004777
forebrain	13943	88	70.7	1.24	0.00000291	0.00004947
brain	10255	73	52	1.4	0.00000524	0.00008908
placenta	13998	87	70.98	1.23	0.0000165	0.0002805
embryo	7118	56	36.09	1.55	0.0000249	0.0004233
chorionic villus	9112	64	46.2	1.39	0.000153	0.002601
cartilage tissue	12013	77	60.91	1.26	0.000201	0.003417
testis	16272	90	82.51	1.09	0.00751	0.12767
amniotic fluid	11724	71	59.45	1.19	0.00758	0.12886
female gonad	16534	90	83.83	1.07	0.0205	0.3485
spinal cord	6459	40	32.75	1.22	0.0729	1
lung	4111	26	20.84	1.25	0.125	1

7.2 Supplementary Material for Chapter 3

Table A- 8 Gene List for ASD

ABAT	ABCA10	ABCA13	ABCA7	ACE
ACHE	ACTN4	ACY1	ADA	ADAMTS18
ADARB1	ADCY3	ADCY5	ADK	ADNP
ADORA2A	ADORA3	ADRB2	ADSL	AFF2
AFF4	AGAP1	AGAP2	AGBL4	AGMO
AGO1	AGO3	AGO4	AGTR2	AHDC1
AHI1	AKAP9	ALDH1A3	ALDH5A1	ALG12
ALG6	ALOX5AP	AMBRA1	AMPD1	AMT
ANK2	ANK3	ANKRD11	ANKS1B	ANXA1
APBA2	APBB1	APC	APH1A	APP
AR	ARHGAP11B	ARHGAP15	ARHGAP24	ARHGAP32
ARHGAP33	ARHGAP5	ARHGEF10	ARHGEF9	ARID1B
ARNT2	ARVCF	ASAP2	ASH1L	ASMT
ASPM	ASS1	ASTN2	ASXL3	ATG7
ATP10A	ATP1A1	ATP1A3	ATP2B2	ATP6V0A2
ATP8A1	ATRNL1	ATRX	ATXN7	AUTS2
AVP	AVPR1A	AVPR1B	AZGP1	BAIAP2
BAZ2B	BBS4	BCAS1	BCKDK	BCL11A
BCL2	BDNF	BICDL1	BIRC6	BRAF
BRCA2	BRD4	BRINP1	BRPF1	BRWD3
BST1	BTAF1	BTRC	C12orf57	C12ORF57
C15orf62	C16orf13	C3orf58	C3ORF58	C4B
CA6	CACNA1A	CACNA1B	CACNA1C	CACNA1D
CACNA1E	CACNA1F	CACNA1G	CACNA1H	CACNA1I
CACNA2D1	CACNA2D3	CACNB2	CADM1	CADM2
CADPS	CADPS2	CAMK2A	CAMK4	CAMSAP2
CAMTA1	CAPN12	CAPRIN1	CARD11	CASC4
CASK	CASZ1	CBLN1	CBS	CC2D1A
CCDC88C	CCDC91	CCNG1	CCNK	CCSER1
CCT4	CD276	CD38	CD44	CD99L2
CDC42BPB	CDH10	CDH11	CDH13	CDH22
CDH8	CDH9	CDK8	CDKL5	CDKN1B
CECR2	CELF4	CELF6	CEP135	CEP290
CEP41	CGNL1	CHD1	CHD2	CHD3
CHD5	CHD7	CHD8	CHMP1A	CHRM3
CHRNA7	CHRNB3	CHST5	CIB2	CIC
CLASP1	CLK2	CLN8	CLSTN2	CLSTN3
CLTCL1	CMIP	CMPK2	CNGB3	CNKSR2

CNOT3	CNR1	CNR2	CNTN3	CNTN4
CNTN5	CNTN6	CNTNAP2	CNTNAP3	CNTNAP4
CNTNAP5	COL28A1	COL5A1	CPEB4	CPT2
CREBBP	CRHR2	CSMD1	CSNK1D	CSNK1E
CTCF	CTNNA2	CTNNA3	CTNNB1	CTNND2
CTTNBP2	CUL3	CUL7	CUX1	CUX2
CX3CR1	CXCR3	CYFIP1	CYLC2	CYP11B1
CYP27A1	DAB1	DAGLA	DAPK1	DAPP1
DCTN5	DCUN1D1	DCX	DDC	DDX11
DDX3X	DDX53	DEAF1	DENR	DGKK
DGKZ	DHX30	DIAPH3	DIP2A	DIP2C
DISC1	DIXDC1	DLG1	DLG4	DLGAP1
DLGAP2	DLGAP3	DLGAP4	DLX1	DLX2
DLX3	DLX6	DMD	DMXL2	DNAH10
DNAH17	DNAH3	DNAJC19	DNER	DNM1L
DNMT3A	DOCK1	DOCK10	DOCK4	DOCK8
DOLK	DPP10	DPP4	DPP6	DPYD
DPYSL2	DPYSL3	DRD1	DRD2	DRD3
DSCAM	DST	DUSP15	DUSP22	DVL1
DVL3	DYDC1	DYDC2	DYNC1H1	DYRK1A
EEF1A2	EFR3A	EGR2	EHMT1	EIF3G
EIF4E	EIF4EBP2	ELAVL2	ELAVL3	ELOVL2
ELP4	EML1	EMSY	EN2	EP300
EP400	EPC2	EPHA6	EPHB2	EPHB6
EPPK1	EPS8	ERBB4	ERBIN	ERG
ERMN	ESR1	ESR2	ESRRB	ETFB
ETHE1	EXOC3	EXOC5	EXOC6	EXOC6B
EXT1	F13A1	FABP3	FABP5	FABP7
FAM19A2	FAM19A3	FAM47A	FAM92B	FAM98C
FAN1	FAT1	FBN1	FBXO11	FBXO33
FBXO40	FCRL6	FEZF2	FGA	FGD1
FGFBP3	FGFR2	FHIT	FLT1	FOLH1
FOXP1	FOXP2	FRK	GABBR2	GABRA1
GABRA4	GABRA5	GABRB1	GABRB3	GABRG3
GABRQ	GAD1	GADD45B	GALNT13	GALNT14
GALNT8	GAMT	GAN	GAP43	GAS2
GATM	GDA	GGNBP2	GIGYF1	GIGYF2
GLIS1	GLO1	GLRA2	GNA14	GNAS
GNB1L	GPC4	GPC6	GPD2	GPHN
GPR139	GPR37	GPR85	GPX1	GRIA1
GRID1	GRID2	GRID2IP	GRIK2	GRIK3
GRIK4	GRIK5	GRIN1	GRIN2A	GRIN2B

GRIP1	GRM4	GRM5	GRM7	GRM8
GRPR	GSK3B	GSN	GSPT2	GSTM1
GTF2I	GUCY1A2	H2AFZ	HDAC3	HDAC6
HDAC8	HDC	HDLBP	HECTD4	HECW2
HEPACAM	HERC2	HIVEP2	HIVEP3	HLA-A
HLA-B	HLA-DRB1	HLA-G	HMGN1	HNRNPH2
HNRNPU	HOMER1	HOXA1	HOXB1	HRAS
HS3ST5	HSD11B1	HTR1B	HTR2A	HTR3A
HTR3C	HTR7	HYDIN	ICA1	IFNG
IFNGR1	IFT140	IL17A	IL17RA	IL1R2
IL1RAPL1	IL1RAPL2	IL6	ILF2	IMMP2L
INPP1	INTS6	IQGAP3	IQSEC2	IRF2BPL
ITGA4	ITGB3	ITGB7	ITPR1	JAKMIP1
JARID2	JMJD1C	KANK1	KAT2B	KAT6A
KATNAL1	KATNAL2	KCNC1	KCND2	KCND3
KCNH5	KCNJ10	KCNJ15	KCNJ2	KCNK7
KCNMA1	KCNQ2	KCNQ3	KCNS3	KCNT1
KCTD13	KCTD3	KDM4B	KDM4C	KDM5B
KDM5C	KDM6A	KDM6B	KHDRBS2	KHDRBS3
KIAA0442	KIAA1586	KIF13B	KIF14	KIF21B
KIF5C	KIRREL3	KIT	KLC2	KLF16
KMO	KMT2A	KMT2C	KMT2E	KMT5B
KRR1	KRT26	LAMA1	LAMB1	LAMC3
LAT	LEO1	LEP	LILRB2	LIN7B
LMX1B	LNPK	LRBA	LRFN2	LRFN5
LRP2	LRPPRC	LRRC1	LRRC4	LZTR1
LZTS2	MAB21L2	MACROD2	MAGED1	MAGEL2
MAL	MAOA	MAOB	MAP2	MAPK1
MAPK3	MAPK8IP2	MARK1	MBD1	MBD3
MBD4	MBD5	MBD6	MBOAT7	MC4R
MCC	MCM4	MCM6	MCPH1	MDGA2
MDM4	MECP2	MED12	MED13	MED13L
MEF2C	MEGF10	MEGF11	MEIS2	MET
MFRP	MIB1	MIM	MIR137	MKL2
MLANA	MLC1	MNT	MOCOS	MPP6
MSANTD2	MSN	MSNP1AS	MSR1	MTF1
MTHFR	MTOR	MTR	MTX2	MUC12
MUC4	MYH10	MYH4	MYO16	MYO1A
MYO1E	MYO5A	MYO5C	MYO9B	MYOZ1
MYT1L	NAA15	NAALADL2	NACC1	NAGLU
NAV2	NBEA	NCKAP1	NCKAP5	NCKAP5L
NCOR1	NDUFA5	NEFL	NEGR1	NEO1

NEXMIF	NFE2L3	NFIA	NFIB	NINL
NIPA1	NIPA2	NLGN1	NLGN2	NLGN3
NLGN4X	NLGN4Y	NLRP5	NMNAT1	NOS1
NOS1AP	NOS2	NOTCH2NL	NPAS2	NR1D1
NR1H2	NR1I3	NR2F1	NR3C2	NR4A2
NRCAM	NRG1	NRP2	NRXN1	NRXN2
NRXN3	NSMCE3	NTNG1	NTRK1	NTRK2
NTRK3	NUAK1	NUDCD2	NUP133	NXPH1
ODF3L2	OFD1	OGT	OPHN1	OPRM1
OR1C1	OR2M4	OR2T10	OR52M1	OTUD7A
OTX1	OXT	OXTR	P2RX4	P2RX5
P4HA2	PAFAH1B1	PAH	PAK2	PAPOLG
PARD3B	PATJ	PAX5	PAX6	PBX1
PCCA	PCDH10	PCDH11X	PCDH15	PCDH8
PCDH9	PCDHA1	PCDHA10	PCDHA11	PCDHA12
PCDHA13	PCDHA2	PCDHA3	PCDHA4	PCDHA5
PCDHA6	PCDHA7	PCDHA8	PCDHA9	PCDHAC1
PCDHAC2	PCDHGA11	PCM1	PDCD1	PDE1C
PDE4A	PDE4B	PDK2	PDZD4	PECR
PER1	PER2	PEX7	PGLYRP2	PHB
PHF2	PHF21A	PHF3	PHIP	PHRF1
PIK3CA	PIK3CG	PINX1	PITX1	PLAUR
PLCB1	PLCD1	PLN	PLPPR4	PLXNA3
PLXNA4	PLXNB1	PNPLA7	POGZ	POLA2
POMT1	PON1	POT1	POU3F2	PPFIA1
PPM1D	PPP1R1B	PPP1R3F	PPP2CA	PPP2R1B
PPP2R2C	PPP2R5D	PQBP1	PREX1	PRICKLE1
PRICKLE2	PRKAR1B	PRKCA	PRKCB	PRKDC
PRKN	PRODH	PRPF39	PRR12	PRUNE2
PSD3	PSMD10	PSMD12	PTBP2	PTCHD1
PTEN	PTGER3	PTGS2	PTK7	PTPN11
PTPRB	PTPRC	PTPRT	PVALB	PXDN
PYHIN1	QRICH1	RAB11FIP5	RAB19	RAB2A
RAB39B	RAB43	RAC1	RAD21	RAD21L1
RAI1	RALA	RALGAPB	RANBP17	RAPGEF4
RASD1	RASSF5	RAX	RB1CC1	RBFOX1
RBM27	RBM8A	RBMS3	REEP3	RELN
RERE	RFWD2	RFX3	RGS7	RHEB
RHOXF1	RIMS1	RIMS3	RIT2	RLIM
RNF135	RNF38	ROBO1	ROBO2	RP11-1407O15.2
RPL10	RPP25	RPS6KA2	RPS6KA3	SAE1
SALL1	SAMD11	SASH1	SATB2	SBF1

SCFD2	SCN11A	SCN1A	SCN2A	SCN4A
SCN8A	SCN9A	SCP2	SDC2	SEMA3F
SEMA5A	SERPINE1	SET	SETBP1	SETD1A
SETD1B	SETD2	SETD5	SETDB1	SETDB2
SEZ6L2	SF3B1	SGSH	SGSM3	SH3KBP1
SH3RF2	SHANK1	SHANK2	SHANK3	SHOX
SIN3A	SLC12A5	SLC12A6	SLC16A3	SLC16A7
SLC1A1	SLC22A15	SLC22A3	SLC22A9	SLC24A2
SLC25A12	SLC25A14	SLC25A24	SLC25A27	SLC25A39
SLC27A4	SLC29A4	SLC2A9	SLC30A3	SLC30A5
SLC33A1	SLC35A2	SLC35A3	SLC35B1	SLC38A10
SLC4A10	SLC6A1	SLC6A3	SLC6A4	SLC6A8
SLC7A3	SLC7A5	SLC7A7	SLC9A1	SLC9A9
SLCO1B3	SLITRK5	SMAD4	SMARCA2	SMARCA4
SMARCC2	SMC1A	SMC3	SMG6	SMURF1
SNAP25	SND1	SNRPN	SNTG2	SNX5
SOD1	SON	SOX10	SOX11	SPARCL1
SPAST	SPEN	SPP2	SRCAP	SRGAP2C
SRGAP3	SRRM4	SRSF11	SSPO	SSRP1
ST7	ST8SIA2	STAT1	STK39	STX1A
STXBP1	STXBP5	STYK1	SUCLG2	SUGCT
SYAP1	SYN1	SYN2	SYN3	SYNCRIP
SYNE1	SYNGAP1	SYNJ1	SYT17	SYT3
TAF1C	TAF1L	TAF6	TANC2	TAOK2
TBC1D23	TBC1D31	TBC1D5	TBC1D7	TBL1X
TBL1XR1	TBR1	TBX1	TCF12	TCF20
TCF4	TCF7L2	TDO2	TECTA	TERB2
TERF2	TERT	TET2	TGM3	TH
THBS1	THRA	TLK2	TM4SF19	TM4SF20
TMEM39B	TMLHE	TNIP2	TNK2	TNRC6B
TOMM20	TOP1	TOP2B	TOP3B	TPH2
TPO	TRAF7	TRAPPC6B	TRAPPC9	TRIM33
TRIO	TRIP12	TRP362ARG	TRPC6	TRPM1
TRPS1	TRRAP	TSC1	TSC2	TSHZ3
TSN	TSPAN17	TSPAN4	TSPAN7	TSPOAP1
TTC25	TTN	TUBGCP5	TXNRD2	TYR
UBA6	UBE2H	UBE3A	UBE3B	UBE3C
UBL7	UBN2	UBR5	UBR7	UCN3
UIMC1	UNC13A	UNC79	UNC80	UPB1
UPF2	UPF3B	USH2A	USP15	USP45
USP7	USP9Y	VASH1	VDR	VIL1
VIP	VLDLR	VRK3	VSIG4	WAC

WASF1	WDFY3	WDFY4	WDR93	WFS1
WNK3	WNT1	WNT2	WWOX	XIRP1
XPC	XPO1	YEATS2	YTHDC2	YWHAE
YWHAG	ZBTB16	ZBTB20	ZC3H4	ZIC1
ZMYND11	ZNF18	ZNF292	ZNF385B	ZNF407
ZNF462	ZNF517	ZNF548	ZNF559	ZNF626
ZNF713	ZNF774	ZNF8	ZNF804A	ZNF827
ZSWIM5	ZSWIM6	ZWILCH		

Table A- 9 Gene List for ID

AACS	ABAT	ABCA2	ABI2	ACACA
ACAT1	ACBD6	ACHE	ACO2	ACOX1
ACSL4	ACTB	ACTC1	ACTL6B	ADAM28
ADAP1	ADAT3	ADCK3	ADGRG1	ADIPOR1
ADK	ADNP	ADRA2B	ADSL	AFF3
AGAP2	AGBL2	AGK	AGO1	AGO3
AGO4	AGT	AHDC1	AHI1	AHSG
AIMP1	AK1	ALA269GLY	ALDH18A1	ALDH3A2
ALDH4A1	ALDH5A1	ALDH7A1	ALG11	ALG13
ALG3	ALG9	ALMS1	ALPK1	ALS2
ALX4	AMER1	AMER2	AMHR2	AMMECR1
AMPD2	AMT	AMZ2	ANAPC2	ANK2
ANK3	ANKRD11	ANP32A	AP1S1	AP1S2
AP3B1	AP3B2	AP4B1	AP4E1	AP4M1
AP4S1	APC2	APTX	AQP10	ARCN1
AREL1	ARFGEF2	ARFGEF3	ARG1644LYSFSTER22	ARG387CYS
ARG415TRP	ARG541TER	ARG599CYS	ARHGAP21	ARHGAP32
ARHGAP33	ARHGEF6	ARHGEF9	ARID1A	ARID1B
ARID2	ARL13B	ARL14EP	ARV1	ARX
ASAH1	ASCC3	ASCL1	ASH1L	ASH2L
ASN195LYSfsTER30	ASNS	ASP370TYR	ASPA	ASPH
ASPM	ASS1	ASTN1	ASXL1	ASXL3
ATF7IP	ATL1	ATN1	ATOH1	ATP13A1
ATP13A2	ATP1A2	ATP1A3	ATP2A1	ATP2B3
ATP2C2	ATP6AP2	ATP7A	ATP8A2	ATRX
AUTS2	ZSCAN25	AVPR2	B3GALNCAPN10	B3GALNT2
B4GALNT1	B4GAT1	BBS4	BBS7	BBS9
BCAS3	BCKDK	BCL11A	BCL11B	BCOR
BCS1L	BDH1	BDNF	BIN1	BLM
BOD1	BPTF	BRAF	BRCA1	BRCA2
BRF1	BRPF1	BRWD3	BTF3	BTN2A2
BTN3A2	C10orf2	C12orf4	C12ORF4	C12orf57
C12ORF57	C12orf65	C12ORF65	C15orf38-AP3S2	C16orf90
C19orf12	C21orf2	C22orf31	C22orf33	C2orf42
C3orf17	C5orf42	C5ORF42	C9orf114	CA2
CA8	CABP7	CACNA1A	CACNA1C	CACNA1D
CACNA1E	CACNA1G	CACNA1H	CACNA2D1	CACNG2
CADPS	CAMK2A	CAMK2B	CAMK2G	CAMTA1
CAPN10	CAPN12	CAPN15	CAPN3	CAPN9
CAPRIN1	CAPS	CAPS2	CASK	CASP2
CASZ1	CBL	CBS	CC2D1A	CC2D2A

CCAR2	CCDC22	CCDC82	CCDC88C	CCNA2
CCNT2	CD163L1	CDC42	CDC42BPB	CDCA5
CDH11	CDH15	CDH4	CDK10	CDK13
CDK19	CDK5RAP2	CDK9	CDKL5	CDKN2AIP
CDON	CELSR2	CENPF	CENPJ	CEP104
CEP290	CEP41	CEP57	CEP76	CEP83
CEP89	CEP97	CFAP44	CHAF1B	CHAMP1
CHD1L	CHD2	CHD3	CHD4	CHD7
CHD8	CHL1	CHMP1A	CHRNA4	CHRNA7
CIC	CINP	CKAP2L	CKAP5	CLASP1
CLCN4	CLHC1	CLIC2	CLIP1	CLMN
CLMP	CLN3	CLN6	CLP1	CLPB
CLPTM1	CLTC	CLTCL1	CNKSR1	CNKSR2
CNNM2	CNOT3	CNPY3	CNTN2	CNTNAP2
COG5	COG6	COL18A1	COL4A1	COL4A3BP
COL4A5	COL5A1	COL9A3	Copyright	COQ5
COX7B	CPLX1	CPNE6	CRADD	CRBN
CREBBP	CRIPT	CSNK2A1	CSNK2B	CSPP1
CSRP2BP	CTCF	CTNNA2	CTNNB1	CTNND2
CTSA	CTSD	CTU2	CUL3	CUL4B
CWC27	CYB5R3	CYP27A1	CYP2U1	DARS
DARS2	DCAF17	DCC	DCHS1	DCPS
DCTN2	DCX	DDHD2	DDX3x	DDX3X
DDX50	DEAF1	DENND2	DENND5A	DEPDC5
DHCR24	DHCR7	DHX30	DHX37	DIAPH1
DICER1	DIP2B	DLG3	DLG4	DLX3
DLX6	DMBT1	DMBX1	DMD	DMXL2
DNA2	DNAH3	DNAJC3	DNAJC6	DNHD1
DNM1	DNM1L	DNMT3A	DOCK3	DOCK6
DOCK8	DPAGT1	DPH1	DPP6	DRD2
DSCAM	DSCAML1	DUOX1	DUOX2	DUP
DVL3	DYM	DYNC1H1	DYRK1A	DYRK4
EBF3	EBP	ECHS1	EDC3	EEF1A2
EEF1B2	EEF1D	EFTUD2	EGLN2	EHMT1
EIF2B1	EIF2S3	EIF4A2	ELOVL4	ELP2
EMC8	EML3	ENO2	ENO3	ENTPD1
EOGT	EP300	EPB41L1	EPB41L4A	EPM2A
ERCC8	ERLIN2	EVC2	EXOC4	EXOSC2
EXOSC3	EXOSC5	EXTL3	EZH1	EZR
FAM104A	FAM120AOS	FAM126A	FAM160B1	FAM177A1
FAM200A	FAM200B	FAM20C	FAM222A	FAM234B
FANCG	FAR1	FARS2	FASN	FAT4

FAXDC2	FBN1	FBN2	FBXL4	FBXO11
FBXO31	FBXO47	FCN3	FDPS	FEZF2
FGD1	FGF12	FGF14	FGFR2	FGFR3
FKRP	FKTN	FMN2	FMOD	FMR1
FNDC3A	FOSL2	FOXG1	FOXP1	FOXP2
FOXRED1	FRAS1	FRMPD4	FRRS1L	FRY
FSCN1	FUCA1	FXR1	FXR2	FZD3
GABRA1	GABRB2	GABRB3	GABRG2	GALE
GALNT2	GALT	GAMT	GAN	GATAD2B
GCC2	GCDH	GCK	GCN1	GDI1
GEMIN4	GGN	GJA1	GJB2	GJC2
GK	GLB1	GLDC	GLI2	GLI3
GLN590HIS	GLN87FS	GLRA1	GLRA2	GLS
GM2A	GMPPA	GMPPB	GNAI1	GNAO1
GNAS	GNB1	GNB5	GNE	GON4L
GPAA1	GPR56	GPR64	GPT2	GRAMD1B
GRIA1	GRIA3	GRIK2	GRIN1	GRIN2A
GRIN2B	GRM1	GRM7	GSS	GTF2I
GTF3C1	GTF3C3	GTPBP3	GUF1	GUSB
HACE1	HACL1	HADH	HAX1	HBB
HCFC1	HDAC4	HDAC8	HECW2	HELZ
HEMK1	HEPACAM	HERC1	HERC2	HEXA
HEXB	HGSNAT	HGSNAT	HIBCH	HIRA
HIST1H1C	HIST1H1E	HIST1H2AC	HIST1H4B	HIST3H3
HIVEP2	HIVEP3	HMG20A	HMGB3	HMGCLL1
HMGXB3	HMNT	HNMT	HNRNPD	HNRNPK
HNRNPL	HNRNPU	HPD	HPRT1	HRAS
HSD17B10	HSPD1	HSPG2	HTR2A	HUWE1
HYAL2	IDS	IFIH1	IFT172	IFT81
IGF1	IGFBP4	IL1RAPL1	ILF3	IMPA1
INA	INIP	INO80	INPP4A	INPP5K
INSR	IPP	IQSEC2	IRF2BPL	IRX5
ISCA1	ISCA2	ITGA7	ITGAV	ITGB6
ITPR1	ITSN1	JMJD1C	KANK1	KANSL1
KAT6A	KAT6B	KCNA2	KCNB1	KCNC1
KCNC3	KCND3	KCNH1	KCNH2	KCNJ10
KCNJ6	KCNK6	KCNK9	KCNQ2	KCNQ3
KCNQ5	KCNS3	KCNT1	KCTD18	KCTD3
KCTD7	KDM5A	KDM5B	KDM5C	KDM6A
KDM6B	KIAA0196	KIAA0442	KIAA0586	KIAA1033
KIAA1109	KIAA1210	KIAA2022	KIDINS220	KIF11
KIF1A	KIF21A	KIF23	KIF4A	KIF5A

KIF5C	KIF7	KIRREL3	KLHL24	KLHL33
KLHL41	KMT2A	KMT2C	KMT2D	KMT5B
KPNA7	KPTN	KRAS	L1CAM	L2HGDH
LAMA1	LAMA2	LAMC1	LAMP2	LARGE1
LARP4B	LARP7	LARS2	LBR	LDB3
LENG8	LEO1	LEU221VAL	LEU282PRO	LHFPL5
LHX3	LINS	LINS1	LMAN2L	LOXHD1
LPIN3	LRCH3	LRP1	LRP2	LRP5
LRRIQ3	LTN1	LYS276fs	LYST	LZTFL1
MADD	MAF	MAG	MAGEA11	MAGEL2
MAGI2	MAMDC2	MAN1B1	MAN2B1	MAN2B2
MANBA	MAOA	MAP2K1	MAP3K7	MAPK3
MAPK8	MAPT	MATN4	MBD5	MBL2
MBNL3	MBOAT7	MBTPS2	MCCC1	MCCC2
MCM3AP	MCPH1	MDGA2	MECP2	MED12
MED13	MED13L	MED23	MED25	MEF2C
MEGF9	MEIS2	ZNF668	METTL23	METTL4
METTL5	MFF	MFN2	MFSD2A	MGAT2
MGME1	MICU1	MIM	MKKS	MMACHC
MMADHC	MPDZ	MPLKIP	MPP4	MRI1
MRPL10	MSL3	MSS51	MT-ATP6	MT-CO1
MT-CO3	MT-CYB	MT-ND1	MT-ND3	MT-TE
MT-TL1	MT-TN	MT-TS1	MT-TW	MTF2
MTFMT	MTHFR	MTOR	MTR	MTSS1L
MUTATION	MXRA8	MYCN	MYH10	MYO18B
MYO1E	MYOCD	MYT1L	NAA10	NAA15
NACC1	NAGA	NAGLU	NAGS	NALCN
NAPB	NARG2	NARS	NARS2	NAT10
NBN	NCAPD2	NCL	NDP	NDST1
NDUFAF1	NDUFS1	NECAP1	NEK9	NEMF
NEURL4	NEXMIF	NF1	NFE2L3	NFIB
NFIX	NFXL1	NGLY1	NHLRC2	NHP2
NHS	NID1	NLGN4X	NONO	NPC2
NPHP4	NR2F1	NR4A2	NRXN1	NSD1
NSDHL	NSUN2	NTN1	NTNG2	NUDT2
NUP107	NUS1	OCLN	OCM2	ODC1
OFD1	OGDHL	OGT	OPA1	OPHN1
OR2A12	ORC1	OTUD6B	OTX2	P4HA3
PACS1	PACS2	PAFAH1B1	PAH	PAK3
PAN2	PANK2	PANX1	PAPOLG	PARP1
PARTIAL	PAX6	PBX1	PCDH19	PCDHB4
PCGF2	PDE4D	PDHA1	PDHX	PDK2

PDPR	PDSS1	PECR	PEPD	PEX1
PEX11B	PEX5	PEX6	PGAM5	PGAP1
PGAP2	PGAP3	PGK1	PGM3	PHC1
PHF12	PHF21A	PHF3	PHF8	PHIP
PHLPP1	PIANP	PIDD1	PIGA	PIGG
PIGL	PIGN	PIGO	PIGQ	PIGV
PIGW	PIGY	PIK3CA	PIK3R2	PIP5K1A
PIWIL1	PLA2G6	PLAC8L1	PLCB1	PLEKHA4
PLEKHG2	PLIN1	PLK2	PLK5	PLP1
PMM2	PMPCA	PNKP	PNPLA6	POC1A
POC1B	POGZ	POLG	POLR1C	POLR3A
POLR3B	POLRMT	POMGNT1	POMGNT2	POMT1
POMT2	POU1F1	POU3F3	PPFIA1	PPM1D
PPOX	PPP1CB	PPP1R15B	PPP1R35	PPP2R1A
PPP2R2C	PPP2R5B	PPP2R5C	PPP2R5D	PPRC1
PQBP1	PREP	PRICKLE2	PRKACA	PRKAR1A
PRKAR2B	PRKCA	PRKCG	PRKD1	PRKRA
PRMT10	PRMT7	PROCA1	PRPF18	PRPS1
PRR12	PRRT2	PRSS12	PRSS48	PRUNE
PSMA7	PSMG4	PSPH	PTCHD1	PTDSS1
PTEN	PTPN11	PTPN23	PTPRT	PTRH2
PTRHD1	PUF60	PURA	PUS7	PYCR1
QRFPR	QRICH1	RAB11A	RAB39B	RAB3GAP1
RABL6	RAC1	RAD50	RAF1	RAI1
RALGAPA1	RALGDS	RARB	RARS2	RBBP8
RBFOX1	RCAN1	RDH11	RERE	REST
RFT1	RFX8	RGL1	RGR	RGS6
RGS7	RIT1	RLIM	RNASEH2	RNASEH2C
RNF125	RNF216	RNF5	RNFT2	ROGDI
ROS1	RPGRIP1L	RPL10	RPL26	RPLP1
RPS6KA3	RRP8	RTN4IP1	RTTN	RUBCN
RUSC2	RXRB	RXYLT1	RYR1	RYR2
SACS	SAMHD1	SASS6	SATB2	SBDS
SBF1	SCAPER	SCG2	SCGN	SCN1A
SCN1B	SCN2A	SCN8A	SDCCAG8	SDK2
SEC23IP	SEC24D	SEPSECS	10-Sep	SERAC1
SET	SETBP1	SETD1B	SETD2	SETD5
SETX	SF3B1	SGCA	SGMS1	SGSH
SHANK2	SHANK3	SHISA6	SHOC2	SHOX
SHROOM4	SIK1	SIL1	SIM2	SIN3A
SIX4	SKIDA1	SLAIN1	SLC13A5	SLC16A1
SLC16A2	SLC18A2	SLC1A4	SLC20A2	SLC25A15

SLC25A23	SLC2A1	SLC31A1	SLC35A2	SLC35A3
SLC35C1	SLC39A14	SLC39A8	SLC44A1	SLC45A1
SLC46A1	SLC4A10	SLC4A4	SLC5A2	SLC6A1
SLC6A17	SLC6A8	SLC7A10	SLC9A1	SLC9A6
SLC9A9	SLITRK5	SMAD4	SMARCA2	SMARCA4
SMARCB1	SMARCD1	SMC1A	SMC3	SMS
SMURF2	SMYD5	SNAP25	SNAPC5	SNAPIN
SNX14	SNX5	SOBP	SON	SOX11
SOX2	SOX5	SP2	SPAST	SPATA13
SPATA5	SPDL1	ZNF804	SPEN	SPG11
SPG20	SPLICE	SPP1	SPR	SPRY2
SPTAN1	SPTBN2	SPTLC1	SRCAP	SRD5A3
SRGAP2	SRGAP2C	SRGAP3	SRPX2	SRRM2
SRSF11	ST3GAL3	ST3GAL5	ST5	ST7
STAG1	STARD9	STC1	STIL	STRA6
STX1A	STX1B	STX3	STXBP1	STXBP5
SUCLA2	SUCLG1	SUMF1	SUMF2	SURF1
SUSD4	SUV420H1	SUZ12	SV2C	SVIL
SYCN	SYDE2	SYNCRIP	SYNE1	SYNGAP1
SYNRG	SYP	SYT1	SZT2	TAB2
TACO1	TADA1	TAF	TAF1	TAF13
TAF2	TAF6	TANGO2	TAOK1	TBC1D23
TBC1D24	TBC1D7	TBC1D8	TBCK	TBL1XR1
TBR1	TBRG1	TBX1	TBX3	TCF12
TCF20	TCF4	TCF7L2	TCTN1	TCTN2
TDP2	TECR	TELO2	TET1	TG
TGIF1	TH	THG1L	THOC2	THOC6
THR157GLNFSTER4	THRB	TIMM8A	TLK2	TM2D3
TMCO1	TMED7-TICAM2	TMEM132D	TMEM135	TMEM147
TMEM178A	TMEM199	TMEM216	TMEM222	TMEM231
TMEM237	TMEM42	TMEM67	TMEM92	TMEM94
TMTC3	TNIK	TNN	TNPO2	TNPO3
TNRC6B	TP53TG5	TPO	TPP1	TRA2B
TRAF7	TRAIP	TRAP1	TRAPPC11	TRAPPC6B
TRAPPC9	TRERF1	TRIM47	TRIO	TRIP12
TRMT1	TRMT10A	TRP129TER	TRP214TER	TRP267TER
TRP302TER	TRP362ARG	TRP473TER	TRP548ARG	TRPM6
TRRAP	TSC1	TSC2	TSEN15	TSEN54
TSHR	TSPAN18	TSPAN7	TTC1	TTC28
TTC5	TTI1	TTI2	TUB	TUBA1A
TUBA3E	TUBA8	TUBAL3	TUBB	TUBB2B
TUBB3	TUBB4A	TUBG1	TUBGCP	TUBGCP6

TUSC3	TUT1	U2AF2	UBA5	UBA7
UBE2A	UBE2J2	UBE3B	UBE3B	UBE4A
UBQLN1	UBR4	UBR7	UFC1	ULK2
UNC80	UPB1	UPF3B	ZNF711	UROC1
USP2	USP27X	USP44	USP9X	VAMP2
VARS	VEZF1	VLDLR	VPS11	VPS13B
VPS13C	VPS35	VRK1	WAC	WASHC4
WDFY3	WDFY4	WDR26	WDR35	WDR45
WDR45B	WDR62	WDR73	WDR81	WDR87
WDR93	WFDC1	WHSC1	WT1	WWOX
XIRP1	XPOT	XRCC4	XRRA1	XYLT1
YAP1	YARS	YTHDF1	YTHDF3	YWHAG
YY1	ZBTB11	ZBTB16	ZBTB18	ZBTB20
ZBTB24	ZBTB40	ZBTB7A	ZC3H14	ZC4H2
ZCCHC8	ZDHHC9	ZEB	ZEB2	ZFAND2B
ZFYVE26	ZIC1	ZMYM5	ZMYND11	ZNF148
ZNF292	ZNF33	ZNF335	ZNF385B	ZNF462
ZNF526	ZNF582			

Table A- 10 Gene List for All

AACS	ABAT	ABCA10	ABCA13	ABCA2
ABCA7	ABI2	ACACA	ACAT1	ACBD6
ACE	ACHE	ACO2	ACOX1	ACSL4
ACTB	ACTC1	ACTL6B	ACTN4	ACTRT1
ACY1	ADA	ADAM28	ADAMTS18	ADAP1
ADARB1	ADAT3	ADCK3	ADCY3	ADCY5
ADGRF2	ADGRG1	ADIPOR1	ADK	ADNP
ADORA2A	ADORA3	ADRA2B	ADRB2	ADSL
AEBP1	AFF2	AFF3	AFF4	AGA
AGAP1	AGAP2	AGBL2	AGBL4	AGK
AGMO	AGO1	AGO3	AGO4	AGT
AGTR2	AHDC1	AHI1	AHSG	AIFM1
AIMP1	AK1	AKAP9	ALA269GLY	ALDH18A1
ALDH1A3	ALDH3A2	ALDH4A1	ALDH5A1	ALDH7A1
ALG11	ALG12	ALG13	ALG3	ALG6
ALG9	ALMS1	ALOX5AP	ALPK1	ALS2
ALX4	AMBRA1	AMER1	AMER2	AMHR2
AMMECR1	AMPD1	AMPD2	AMT	AMZ2
ANAPC2	ANK2	ANK3	ANKRD11	ANKS1B
ANP32A	ANXA1	AP1S1	AP1S2	AP3B1
AP3B2	AP4B1	AP4E1	AP4M1	AP4S1
APBA2	APBB1	APC	APC2	APH1A
APP	APTX	AQP10	AR	ARCN1
AREL1	ARFGAP2	ARFGEF2	ARFGEF3	ARG1644LYSFSTER22
ARG387CYS	ARG415TRP	ARG541TER	ARG599CYS	ARHGAP11B
ARHGAP15	ARHGAP21	ARHGAP23	ARHGAP24	ARHGAP32
ARHGAP33	ARHGAP5	ARHGEF10	ARHGEF6	ARHGEF9
ARID1A	ARID1B	ARID2	ARL13B	ARL14EP
ARNT2	ARV1	ARVCF	ARX	ASAH1
ASAP2	ASCC3	ASCL1	ASH1L	ASH2L
ASL	ASMT	ASN195LYSfsTER30	ASNS	ASP370TYR
ASPA	ASPH	ASPM	ASS1	ASTN1
ASTN2	ASXL1	ASXL3	ATF7IP	ATG7
ATL1	ATN1	ATOH1	ATP10A	ATP13A1
ATP13A2	ATP1A1	ATP1A2	ATP1A3	ATP2A1
ATP2B1	ATP2B2	ATP2B3	ATP2C2	ATP6AP2
ATP6V0A2	ATP7A	ATP8A1	ATP8A2	ATRNL1
ATRX	ATXN7		AUTS2	AVP
AVPR1A	AVPR1B	AVPR2	AZGP1	B3GALNCAPN10
B3GALNT2	B4GALNT1	B4GAT1	BAIAP2	BAIAP2L2
BAZ2B	BBS4	BBS7	BBS9	BCAS1

BCAS3	BCKDK	BCL11A	BCL11B	BCL2
BCOR	BCS1L	BDH1	BDNF	BICDL1
BIN1	BIRC6	BLM	BOD1	BPTF
BRAF	BRCA1	BRCA2	BRD4	BRF1
BRINP1	BRPF1	BRWD3	BST1	BTAF1
BTF3	BTN2A2	BTN3A2	BTRC	C10orf2
C10orf25	C12ORF4	C12ORF57	C12ORF65	C15orf38-AP3S2
C15orf62	C16orf13	C16orf90	C19orf12	C21orf2
C22orf31	C22orf33	C2orf42	C3orf17	C3ORF58
C4B	C5ORF42	C9orf114	CA2	CA6
CA8	CABP2	CABP7	CACNA1A	CACNA1B
CACNA1C	CACNA1D	CACNA1E	CACNA1F	CACNA1G
CACNA1H	CACNA1I	CACNA2D1	CACNA2D3	CACNB2
CACNG2	CADM1	CADM2	CADPS	CADPS2
CAMK2A	CAMK2B	CAMK2G	CAMK4	CAMSAP2
CAMTA1	CAPN10	CAPN12	CAPN15	CAPN3
CAPN9	CAPRIN1	CAPS	CAPS2	CAPSL
CARD11	CASC4	CASK	CASP2	CASZ1
CBL	CBLN1	CBS	CC2D1A	CC2D2A
CCAR2	CCDC150	CCDC152	CCDC22	CCDC82
CCDC88C	CCDC91	CCNA2	CCNG1	CCNK
CCNT2	CCSER1	CCT4	CD163L1	CD276
CD38	CD44	CD99L2	CDC42	CDC42BPB
CDCA5	CDH10	CDH11	CDH13	CDH15
CDH20	CDH22	CDH4	CDH8	CDH9
CDK10	CDK13	CDK19	CDK5RAP2	CDK8
CDK9	CDKL5	CDKN1B	CDKN2AIP	CDON
CECR2	CELA3A	CELF4	CELF6	CELSR1
CELSR2	CENPF	CENPJ	CEP104	CEP135
CEP290	CEP41	CEP57	CEP76	CEP83
CEP89	CEP97	CFAP44	CGNL1	CHAF1B
CHAMP1	CHD1	CHD1L	CHD2	CHD3
CHD4	CHD5	CHD7	CHD8	CHL1
CHMP1A	CHRM3	CHRNA4	CHRNA7	CHRNB3
CHST5	CIB2	CIC	CINP	CKAP2L
CKAP5	CLASP1	CLCA4	CLCN4	CLEC4E
CLHC1	CLIC2	CLIP1	CLK2	CLMN
CLMP	CLN3	CLN6	CLN8	CLP1
CLPB	CLPTM1	CLSTN2	CLSTN3	CLTC
CLTCL1	CMIP	CMPK2	CNGA2	CNGB3
CNKSR1	CNKSR2	CNNM2	CNOT3	CNPY3
CNPY4	CNR1	CNR2	CNTN2	CNTN3

CNTN4	CNTN5	CNTN6	CNTNAP2	CNTNAP3
CNTNAP4	CNTNAP5	COG5	COG6	COL18A1
COL28A1	COL4A1	COL4A3BP	COL4A5	COL5A1
COL9A3	Copyright	COQ5	COX7B	CPEB4
CPLX1	CPM	CPNE6	CPT2	CRADD
CRBN	CREBBP	CRHR2	CRIPT	CSMD1
CSNK1D	CSNK1E	CSNK2A1	CSNK2B	CSPP1
CSRP2BP	CTCF	CTNNA2	CTNNA3	CTNNB1
CTNND2	CTR9	CTSA	CTSD	CTTNBP2
CTU2	CUL3	CUL4B	CUL7	CUX1
CUX2	CWC27	CX3CR1	CXCR3	CYB5R3
CYFIP1	CYLC2	CYP11B1	CYP27A1	CYP2U1
CYP39A1	DAB1	DAGLA	DAPK1	DAPP1
DARS	DARS2	DCAF17	DCC	DCHS1
DCPS	DCTN2	DCTN5	DCUN1D1	DCX
DDC	DDHD2	DDX11	DDX3X	DDX50
DDX53	DEAF1	DENND1B	DENND2	DENND4B
DENND5A	DENR	DEPDC4	DEPDC5	DGKK
DGKZ	DHCR24	DHCR7	DHX30	DHX37
DIAPH1	DIAPH3	DICER1	DIP2A	DIP2B
DIP2C	DISC1	DIXDC1	DLG1	DLG3
DLG4	DLGAP1	DLGAP2	DLGAP3	DLGAP4
DLX1	DLX2	DLX3	DLX6	DMBT1
DMBX1	DMD	DMXL2	DNA2	DNAH10
DNAH14	DNAH17	DNAH3	DNAH8	DNAJC19
DNAJC3	DNAJC6	DNER	DNHD1	DNM1
DNM1L	DNM3	DNMT3A	DOCK1	DOCK10
DOCK3	DOCK4	DOCK6	DOCK8	DOLK
DPAGT1	DPH1	DPP10	DPP4	DPP6
DPY19L3	DPYD	DPYSL2	DPYSL3	DRD1
DRD2	DRD3	DRP2	DSCAM	DSCAML1
DST	DUOX1	DUOX2	DUP	DUSP15
DUSP22	DVL1	DVL3	DYDC1	DYDC2
DYM	DYNC1H1	DYRK1A	DYRK4	DZANK1
EBF3	EBP	ECHS1	ECM1	EDC3
EEF1A2	EEF1B2	EEF1D	EFCAB8	EFR3A
EFR3B	EFTUD2	EGFL6	EGLN2	EGR2
EHMT1	EIF2B1	EIF2S3	EIF3G	EIF4A2
EIF4E	EIF4EBP2	ELAVL2	ELAVL3	ELOVL2
ELOVL4	ELP2	ELP4	EMC8	EML1
EML3	EMSY	EN2	ENO2	ENO3
ENPP6	ENTPD1	EOGT	EP300	EP400

EPB41L1	EPB41L4A	EPC2	EPHA6	EPHB1
EPHB2	EPHB6	EPM2A	EPPK1	EPS8
ERBB4	ERBIN	ERCC8	ERG	ERLIN2
ERMN	ESR1	ESR2	ESRRB	ETFB
ETHE1	EVC2	EXOC3	EXOC4	EXOC5
EXOC6	EXOC6B	EXOSC2	EXOSC3	EXOSC5
EXT1	EXTL3	EZH1	EZR	F13A1
FABP3	FABP5	FABP7	FAM104A	FAM120AOS
FAM126A	FAM160B1	FAM177A1	FAM19A2	FAM19A3
FAM200A	FAM200B	FAM20C	FAM222A	FAM234B
FAM47A	FAM53C	FAM92B	FAM98C	FAN1
FANCG	FAR1	FARS2	FASN	FAT1
FAT4	FAXDC2	FBN1	FBN2	FBXL4
FBXO11	FBXO31	FBXO33	FBXO40	FBXO47
FCN3	FCRL6	FDPS	FEZF2	FGA
FGD1	FGF12	FGF14	FGFBP3	FGFR2
FGFR3	FHIT	FKRP	FKTN	FLT1
FMN2	FMOD	FMR1	FNDC3A	FOLH1
FOSL2	FOXG1	FOXP1	FOXP2	FOXRED1
FRAS1	FRG1	FRK	FRMPD4	FRRS1L
FRY	FSCN1	FUCA1	FXR1	FXR2
FZD3	GABBR2	GABRA1	GABRA4	GABRA5
GABRB1	GABRB2	GABRB3	GABRG2	GABRG3
GABRQ	GAD1	GADD45B	GALE	GALNT13
GALNT14	GALNT2	GALNT8	GALT	GAMT
GAN	GAP43	GAS2	GATAD2B	GATM
GCC2	GCDH	GCK	GCN1	GDA
GDI1	GEMIN4	GFI1B	GGN	GGNBP2
GIGYF1	GIGYF2	GIMAP8	GJA1	GJB2
GJC2	GK	GLB1	GLDC	GLI2
GLI3	GLIS1	GLN590HIS	GLN87FS	GLO1
GLRA1	GLRA2	GLS	GM2A	GMPPA
GMPPB	GNA14	GNAI1	GNAO1	GNAS
GNB1	GNB1L	GNB5	GNE	GON4L
GPAA1	GPC4	GPC6	GPD2	GPHN
GPR139	GPR180	GPR37	GPR56	GPR64
GPR85	GPT2	GPX1	GRAMD1B	GRIA1
GRIA3	GRID1	GRID2	GRID2IP	GRIK2
GRIK3	GRIK4	GRIK5	GRIN1	GRIN2A
GRIN2B	GRIP1	GRM1	GRM2	GRM4
GRM5	GRM7	GRM8	GRPR	GSE1
GSK3B	GSN	GSPT2	GSS	GSTM1

GTF2H1	GTF2I	GTF3C1	GTF3C3	GTPBP3
GUCY1A2	GUCY2C	GUCY2F	GUF1	GUSB
H2AFZ	HACE1	HACL1	HADH	HAX1
HBB	HCFC1	HDAC3	HDAC4	HDAC6
HDAC8	HDC	HDHD5	HDLBP	HECTD4
HECW2	HELZ	HEMK1	HEPACAM	HERC1
HERC2	HEXA	HEXB	HGSNAT	HGSNAT
HIBCH	HIRA	HIST1H1C	HIST1H1E	HIST1H2AC
HIST1H4B	HIST3H3	HIVEP2	HIVEP3	HLA-A
HLA-B	HLA-DRB1	HLA-G	HMG20A	HMGB3
HMGCLL1	HMGN1	HMGXB3	HMNT	HNMT
HNRNPD	HNRNPH2	HNRNPK	HNRNPL	HNRNPU
HOMER1	HOXA1	HOXB1	HPD	HPRT1
HRAS	HS3ST5	HSD11B1	HSD17B10	HSPD1
HSPG2	HTR1B	HTR2A	HTR3A	HTR3C
HTR7	HTRA2	HUWE1	HYAL2	HYDIN
ICA1	IDS	IFIH1	IFNG	IFNGR1
IFT140	IFT172	IFT81	IGF1	IGFBP4
IL17A	IL17RA	IL1R2	IL1RAPL1	IL1RAPL2
IL6	ILDR2	ILF2	ILF3	IMMP2L
IMPA1	INA	INIP	INO80	INPP1
INPP4A	INPP5K	INSR	INTS6	IPP
IQGAP3	IQSEC2	IRF2BPL	IRX5	ISCA1
ISCA2	ITGA4	ITGA7	ITGAV	ITGB3
ITGB6	ITGB7	ITPR1	ITSN1	JAKMIP1
JARID2	JMJD1C	KANK1	KANSL1	KAT2B
KAT6A	KAT6B	KAT7	KATNAL1	KATNAL2
KCNA2	KCNB1	KCNC1	KCNC3	KCND2
KCND3	KCNH1	KCNH2	KCNH5	KCNJ10
KCNJ15	KCNJ2	KCNJ6	KCNK6	KCNK7
KCNK9	KCNMA1	KCNQ2	KCNQ3	KCNQ5
KCNS3	KCNT1	KCTD13	KCTD18	KCTD3
KCTD7	KDM4B	KDM4C	KDM5A	KDM5B
KDM5C	KDM6A	KDM6B	KHDRBS2	KHDRBS3
KIAA0040	KIAA0196	KIAA0442	KIAA0586	KIAA1033
KIAA1109	KIAA1143	KIAA1210	KIAA1586	KIAA2022
KIDINS220	KIF11	KIF13B	KIF14	KIF16B
KIF1A	KIF21A	KIF21B	KIF23	KIF4A
KIF5A	KIF5C	KIF7	KIRREL3	KIT
KLC2	KLF16	KLHL24	KLHL33	KLHL41
KLK3	KMO	KMT2A	KMT2C	KMT2D
KMT2E	KMT5B	KPNA7	KPTN	KRAS

KRR1	KRT26	KTN1	L1CAM	L2HGDH
LAMA1	LAMA2	LAMB1	LAMC1	LAMC3
LAMP2	LANCL3	LAP3	LARGE1	LARP4B
LARP7	LARS2	LAT	LBR	LCE4A
LCTL	LDB3	LENG8	LEO1	LEP
LEU221VAL	LEU282PRO	LGMN	LHFPL5	LHX3
LILRB2	LIN7B	LINS	LINS1	LMAN2L
LMX1B	LNP1	LNPK	LOC100129697	LOXHD1
LPIN3	LRBA	LRCH3	LRFN2	LRFN5
LRP1	LRP2	LRP5	LRPPRC	LRRC1
LRRC34	LRRC4	LRRIQ3	LTN1	LYS276fs
LYST	LZTFL1	LZTR1	LZTS2	MAB21L2
MACROD2	MADD	MAF	MAFA	MAG
MAGEA11	MAGEB2	MAGED1	MAGEL2	MAGI2
MAL	MAMDC2	MAN1B1	MAN2B1	MAN2B2
MANBA	MAOA	MAOB	MAP2	MAP2K1
MAP3K7	MAPK1	MAPK3	MAPK8	MAPK8IP2
MAPKBP1	MAPT	MARK1	MATN4	MBD1
MBD3	MBD4	MBD5	MBD6	MBL2
MBNL3	MBOAT7	MBTPS2	MC4R	MCC
MCCC1	MCCC2	MCM3AP	MCM4	MCM6
MCPH1	MDGA2	MDM4	MECP2	MED12
MED13	MED13L	MED23	MED25	MEF2C
MEGF10	MEGF11	MEGF9	MEIS2	ZSWIM5
MET	METTL23	METTL4	METTL5	MFF
MFN2	MFRP	MFSD2A	MGAT2	MGME1
MIB1	MICU1	MIM	MIR137	MKKS
MKL2	MLANA	MLC1	MMACHC	MMADHC
MNT	MOCOS	MORC1	MPDZ	MPLKIP
MPP4	MPP6	MRI1	MRPL10	MRPL4
MSANTD2	MSL3	MSN	MSNP1AS	MSR1
MSS51	MT-ATP6	MT-CO1	MT-CO3	MT-CYB
MT-ND1	MT-ND3	MT-TE	MT-TL1	MT-TN
MT-TS1	MT-TW	MTF1	MTF2	MTFMT
MTHFR	MTMR8	MTOR	MTR	MTSS1L
MTX2	MUC12	MUC4	MUTATION	MXRA8
MYCN	MYH10	MYH11	MYH15	MYH4
MYO16	MYO18B	MYO1A	MYO1E	MYO1F
MYO5A	MYO5C	MYO9B	MYOCD	MYOZ1
MYT1L	NAA10	NAA15	NAALADL2	NACC1
NAGA	NAGLU	NAGS	NALCN	NAPB
NARG2	NARS	NARS2	NAT10	NAV2

NBEA	NBN	NCAPD2	NCKAP1	NCKAP5
NCKAP5L	NCL	NCOR1	NDP	NDST1
NDUFA5	NDUFAF1	NDUFS1	NECAP1	NEFL
NEGR1	NEK9	NEMF	NEO1	NEURL4
NEXMIF	NF1	NFE2L3	NFIA	NFIB
NFIC	NFIX	NFXL1	NGLY1	NHLRC2
NHP2	NHS	NID1	NINL	NIPA1
NIPA2	NLGN1	NLGN2	NLGN3	NLGN4X
NLGN4Y	NLRP5	NMNAT1	NONO	NOS1
NOS1AP	NOS2	NOTCH2NL	NPAS2	NPC2
NPHP4	NR1D1	NR1H2	NR1I3	NR2F1
NR3C2	NR4A2	NRCAM	NRG1	NRK
NRP2	NRXN1	NRXN2	NRXN3	NSD1
NSDHL	NSMCE3	NSUN2	NTN1	NTNG1
NTNG2	NTRK1	NTRK2	NTRK3	NUAK1
NUDCD2	NUDT2	NUP107	NUP133	NUS1
NVL	NXPH1	OCLN	OCM2	ODC1
ODF3L2	OFD1	OGDHL	OGT	OPA1
OPHN1	OPRM1	OR1C1	OR2A12	OR2M4
OR2T10	OR52M1	ORC1	OSBPL3	OTUD6B
OTUD7A	OTX1	OTX2	OXT	OXTR
P2RX4	P2RX5	P4HA2	P4HA3	PACS1
PACS2	PAFAH1B1	PAH	PAK2	PAK3
PAN2	PANK2	PANX1	PAPOLG	PARD3B
PARP1	PARTIAL	PATJ	PAX5	PAX6
PBX1	PCCA	PCDH10	PCDH11X	PCDH15
PCDH19	PCDH8	PCDH9	PCDHA1	PCDHA10
PCDHA11	PCDHA12	PCDHA13	PCDHA2	PCDHA3
PCDHA4	PCDHA5	PCDHA6	PCDHA7	PCDHA8
PCDHA9	PCDHAC1	PCDHAC2	PCDHB4	PCDHGA11
PCGF2	PCM1	PDCD1	PDE1C	PDE4A
PDE4B	PDE4D	PDHA1	PDHX	PDK2
PDPR	PDSS1	PDZD4	PECR	PEPD
PER1	PER2	PEX1	PEX11B	PEX5
PEX6	PEX7	PGAM5	PGAP1	PGAP2
PGAP3	PGK1	PGLYRP2	PGM3	PHB
PHC1	PHF12	PHF2	PHF21A	PHF3
PHF8	PHIP	PHKA1	PHLPP1	PHRF1
PIANP	PIDD1	PIGA	PIGG	PIGL
PIGN	PIGO	PIGQ	PIGV	PIGW
PIGY	PIK3CA	PIK3CG	PIK3R2	PINX1
PIP5K1A	PITX1	PIWIL1	PKD1L1	PLA2G6

PLAC8L1	PLAUR	PLCB1	PLCD1	PLCL2
PLEKHA4	PLEKHG2	PLIN1	PLK2	PLK5
PLN	PLP1	PLPPR4	PLXNA3	PLXNA4
PLXNB1	PM20D1	PMM2	PMPCA	PNKP
PNPLA6	PNPLA7	POC1A	POC1B	POGZ
POLA2	POLG	POLR1C	POLR3A	POLR3B
POLRMT	POMGNT1	POMGNT2	POMT1	POMT2
PON1	POT1	POU1F1	POU3F2	POU3F3
PPFIA1	PPIL2	PPM1D	PPOX	PPP1CB
PPP1R15B	PPP1R1B	PPP1R35	PPP1R36	PPP1R3F
PPP2CA	PPP2R1A	PPP2R1B	PPP2R2C	PPP2R5B
PPP2R5C	PPP2R5D	PPRC1	PQBP1	PRB2
PREP	PREX1	PRICKLE1	PRICKLE2	PRKACA
PRKAR1A	PRKAR1B	PRKAR2B	PRKCA	PRKCB
PRKCG	PRKD1	PRKDC	PRKN	PRKRA
PRMT10	PRMT7	PROCA1	PRODH	PRPF18
PRPF39	PRPS1	PRR12	PRRT2	PRRT3
PRSS12	PRSS3	PRSS48	PRUNE	PRUNE2
PSD3	PSMA7	PSMD10	PSMD12	PSMG4
PSPH	PTBP2	PTCHD1	PTDSS1	PTEN
PTGER3	PTGS2	PTK7	PTPN11	PTPN2
PTPN23	PTPRB	PTPRC	PTPRT	PTRH2
PTRHD1	PUF60	PURA	PUS7	PVALB
PXDN	PYCR1	PYHIN1	QRFPR	QRICH1
RAB11A	RAB11FIP5	RAB19	RAB2A	RAB39B
RAB3GAP1	RAB3GAP2	RAB42	RAB43	RABL6
RAC1	RAD21	RAD21L1	RAD50	RAF1
RAI1	RALA	RALGAPA1	RALGAPB	RALGDS
RANBP17	RANBP9	RAPGEF4	RARB	RARS2
RASAL2	RASD1	RASSF5	RASSF6	RAX
RB1CC1	RBBP8	RBFOX1	RBM27	RBM8A
RBMS3	RCAN1	RDH11	REEP3	RELN
RERE	REST	RETN	RFT1	RFWD2
RFX3	RFX8	RGL1	RGR	RGS6
RGS7	RHEB	RHOXF1	RIMS1	RIMS3
RIT1	RIT2	RLIM	RNASEH2	RNASEH2C
RNF125	RNF135	RNF216	RNF38	RNF5
RNFT2	ROBO1	ROBO2	ROGDI	ROS1
RP11-1407O15.2	RP1L1	RPGRIP1L	RPL10	RPL26
RPLP1	RPP25	RPS6KA2	RPS6KA3	RRAGB
RRP8	RRS1	RSRC1	RTN4IP1	RTTN
RUBCN	RUSC2	RXRB	RXYLT1	RYR1

RYR2	SACS	SAE1	SAGE1	SALL1
SAMD11	SAMD14	SAMHD1	SASH1	SASS6
SATB2	SBDS	SBF1	SBF2	SCAPER
SCFD2	SCG2	SCGN	SCN10A	SCN11A
SCN1A	SCN1B	SCN2A	SCN4A	SCN8A
SCN9A	SCP2	SDC2	SDCCAG8	SDK2
SEC23IP	SEC24D	SEMA3F	SEMA5A	SEPSECS
SEPT10	SERAC1	SERPINE1	SERPINF2	SET
SETBP1	SETD1A	SETD1B	SETD2	SETD5
SETDB1	SETDB2	SETX	SEZ6L2	SF3B1
SGCA	SGMS1	SGSH	SGSM3	SH3KBP1
SH3RF2	SHANK1	SHANK2	SHANK3	SHH
SHISA6	SHOC2	SHOX	SHROOM4	SIK1
SIL1	SIM2	SIN3A	SIX4	SKIDA1
SLAIN1	SLAMF7	SLC12A5	SLC12A6	SLC13A5
SLC16A1	SLC16A2	SLC16A3	SLC16A7	SLC18A2
SLC1A1	SLC1A4	SLC20A2	SLC22A15	SLC22A3
SLC22A9	SLC24A2	SLC25A12	SLC25A14	SLC25A15
SLC25A23	SLC25A24	SLC25A27	SLC25A39	SLC27A4
SLC29A4	SLC2A1	SLC2A9	SLC30A3	SLC30A5
SLC31A1	SLC33A1	SLC35A2	SLC35A3	SLC35B1
SLC35C1	SLC36A1	SLC38A10	SLC39A14	SLC39A8
SLC44A1	SLC45A1	SLC46A1	SLC4A10	SLC4A4
SLC5A2	SLC6A1	SLC6A17	SLC6A3	SLC6A4
SLC6A8	SLC7A10	SLC7A3	SLC7A5	SLC7A7
SLC9A1	SLC9A6	SLC9A9	SLCO1B3	SLITRK5
SMAD4	SMARCA2	SMARCA4	SMARCB1	SMARCC2
SMARCD1	SMC1A	SMC3	SMG6	SMS
SMURF1	SMURF2	SMYD5	SNAP25	SNAPC5
SNAPIN	SND1	SNRPN	SNTG2	SNX14
SNX5	SOBP	SOD1	SON	SOX10
SOX11	SOX2	SOX5	SP2	SPARCL1
SPAST	SPATA13	SPATA18	SPATA20	SPATA5
SPATS2	SPDL1	SPECC1	ZSWIM6	SPEN
SPG11	SPG20	SPLICE	SPP1	SPP2
SPR	SPRY2	SPTAN1	SPTBN2	SPTLC1
SRCAP	SRD5A3	SRGAP2	SRGAP2C	SRGAP3
SRPX2	SRRM2	SRRM4	SRSF11	SSPO
SSRP1	ST3GAL3	ST3GAL5	ST5	ST7
ST8SIA2	STAG1	STARD9	STAT1	STC1
STIL	STK39	STRA6	STX1A	STX1B
STX3	STXBP1	STXBP5	STYK1	SUCLA2

SUCLG1	SUCLG2	SUGCT	SUMF1	SUMF2
SURF1	SUSD4	SUV420H1	SUZ12	SV2C
SVIL	SYAP1	SYCN	SYDE2	SYN1
SYN2	SYN3	SYNCRIP	SYNE1	SYNGAP1
SYNJ1	SYNRG	SYP	SYT1	SYT17
SYT3	SZT2	TAB2	TACO1	TADA1
TAF	TAF1	TAF13	TAF1C	TAF1L
TAF2	TAF6	TANC2	TANGO2	TAOK1
TAOK2	TBC1D23	TBC1D24	TBC1D31	TBC1D5
TBC1D7	TBC1D8	TBCB	TBCK	TBL1X
TBL1XR1	TBR1	TBRG1	TBX1	TBX3
TCF12	TCF20	TCF4	TCF7L2	TCTN1
TCTN2	TDO2	TDP2	TECPR2	TECR
TECTA	TELO2	TERB2	TERF2	TERT
TET1	TET2	TG	TGIF1	TGM3
TH	THBS1	THG1L	THOC2	THOC6
THR157GLNFSTER4	THRA	THRB	TIMM8A	TIMP4
TLK2	TM2D3	TM4SF19	TM4SF20	TMCO1
TMED7-TICAM2	TMEM132D	TMEM135	TMEM147	TMEM170A
TMEM178A	TMEM199	TMEM210	TMEM216	TMEM222
TMEM231	TMEM237	TMEM25	TMEM27	TMEM39B
TMEM42	TMEM67	TMEM92	TMEM94	TMIGD3
TMLHE	TMTC3	TNIK	TNIP2	TNK2
TNN	TNPO2	TNPO3	TNRC6B	TOMM20
TOMM34	TOP1	TOP2B	TOP3B	TP53TG5
TPH2	TPO	TPP1	TRA2B	TRAF7
TRAIP	TRAP1	TRAPPC11	TRAPPC6B	TRAPPC9
TRERF1	TRIM33	TRIM47	TRIM55	TRIO
TRIP12	TRMT1	TRMT10A	TRP129TER	TRP214TER
TRP267TER	TRP302TER	TRP362ARG	TRP473TER	TRP548ARG
TRPC5	TRPC6	TRPM1	TRPM6	TRPS1
TRRAP	TSC1	TSC2	TSEN15	TSEN54
TSHR	TSHZ3	TSN	TSPAN17	TSPAN18
TSPAN4	TSPAN7	TSPOAP1	TTC1	TTC13
TTC25	TTC28	TTC5	TTI1	TTI2
TTN	TUB	TUBA1A	TUBA3E	TUBA8
TUBAL3	TUBB	TUBB2B	TUBB3	TUBB4A
TUBG1	TUBGCP	TUBGCP5	TUBGCP6	TUSC3
TUT1	TXNRD2	TYR	U2AF2	UBA5
UBA6	UBA7	UBE2A	UBE2H	UBE2J2
UBE3A	UBE3B	UBE3C	UBE4A	UBL7
UBN2	UBQLN1	UBR4	UBR5	UBR7

UCN3	UFC1	UHRF1BP1	UHRF2	UIMC1
ULK2	UNC13A	UNC79	UNC80	UPB1
UPF2	UPF3B	ZWILCH	UROC1	USH2A
USP15	USP2	USP27X	USP44	USP45
USP7	USP9X	USP9Y	VAMP2	VARS
VASH1	VDR	VEZF1	VIL1	VIP
VKORC1L1	VLDLR	VPS11	VPS13B	VPS13C
VPS16	VPS35	VRK1	VRK3	VSIG4
WAC	WASF1	WASHC4	WDFY3	WDFY4
WDR26	WDR27	WDR35	WDR45	WDR45B
WDR62	WDR73	WDR81	WDR87	WDR90
WDR93	WFDC1	WFS1	WHSC1	WNK3
WNT1	WNT2	WT1	WWOX	XIRP1
XKR5	XKR9	XPC	XPO1	XPOT
XRCC4	XRRA1	XYLT1	YAP1	YARS
YEATS2	YTHDC2	YTHDF1	YTHDF3	YWHAE
YWHAG	YY1	ZAR1	ZBTB11	ZBTB16
ZBTB18	ZBTB20	ZBTB24	ZBTB40	ZBTB7A
ZC3H14	ZC3H4	ZC4H2	ZCCHC8	ZDHHC9
ZEB	ZEB2	ZFAND2B	ZFYVE26	ZIC1
ZMYM5	ZMYND11	ZNF112	ZNF121	ZNF148
ZNF18	ZNF185	ZNF292	ZNF33	ZNF335
ZNF385B	ZNF407	ZNF462	ZNF517	ZNF526
ZNF548	ZNF559	ZNF582	ZNF626	ZNF668
ZNF711	ZNF713	ZNF774	ZNF8	ZNF804
ZNF804A	ZNF827	ZNF888	ZSCAN25	

7.3 Supplementary Material for Chapter 4

7.3.1 Supplemental Materials and Methods

ES methods for GeneDx (17-003, 17-007, 17-008, 17-010, 17-002, 17-028, 17-019): Using genomic DNA from the proband and parent(s), when available, the exonic regions and flanking splice junctions of the genome were captured using the Clinical Research Exome kit (Agilent Technologies, Santa Clara, CA) or the IDT xGen Exome Research Panel v1.0. Massively parallel (NextGen) sequencing was done on an Illumina system with 100bp or greater paired-end reads. Reads were aligned to human genome build GRCh37/UCSC hg19 and analyzed for sequence variants using a custom-developed analysis tool. Additional sequencing technology and variant interpretation protocol has been previously described.(Retterer et al., 2016)
The general assertion criteria for variant classification are publicly available on the GeneDx ClinVar submission page (http://www.ncbi.nlm.nih.gov/clinvar/submitters/26957/). Exome data analysis was child-parent trio based, when both parents were available (for individuals 17-003, 17-007, 17-008, 17-010, 17-002, and 17-019). One child (17-028) underwent child-parent (duo) analysis as a paternal sample was not submitted for analysis.

ES methods for Ambry (17-009, 17-027, 18-003, 18-004, 18-005): Library preparation, DES, bioinformatics and data analysis were performed as previously described at Ambry Genetics (Aliso Viejo, California, USA)(Farwell et al., 2015), with the exception of the capture reagent which was IDT xGen Exome Research Panel V.1.0. Approximately 98% of the exomes of both patients and their parents were sequenced at a depth of at least 20X. Bidirectional Sanger sequencing was used to confirm the *ZNF292* variants in patients and their parents. Minor allele frequency (MAF) thresholds of 0.1% and 1% were used, respectively, for variants displaying dominant and recessive modes of inheritance in known Mendelian disease genes. The MAF thresholds for variants found in genes presently not known to cause any Mendelian disease were 0.1% and 0.2% for dominant and recessive modes, respectively. The population databases used included 1000 g phase 3, the Exome Sequencing Project and the Exome as well as Genome Aggregation Consortia (ExAC and gnomAD, respectively). The dominant inheritance model comprised *de novo* variants whereas the recessive model comprised rare homozygous and compound heterozygous variants. If one or both parents were affected with some or all of the

clinical features of the proband then we also assessed all rare variants inherited from that parent in the AD model. We also assessed a third 'incomplete penetrance' model containing variants inherited from either parent that were (1) listed in the Human Gene Mutation Database (HGMD) or internally classified as pathogenic or likely pathogenic in our clinical laboratory, (2) predicted to cause loss of function such as nonsense, frameshifts and canonical splice sites, and (3) found in genes that cause imprinted disorders inherited from either parent. Both patients and their parents were provided with a complementary analysis of secondary findings in the minimum list of genes recommended by the American College of Medical Genetics and Genomics. There were no reportable secondary findings in this cohort.

ES methods for 17-014: Institutional Research Ethics Board approval was received through the Centre for Addiction and Mental Health (CAMH) and other institutional recruiting sites. Parents from Iranian families filled out the Social Communication Questionnaire (SCQ) and children who met the specified cut-off were then assessed using the Autism Diagnostic Interview-Revised (ADI-R) and some with the Autism Diagnostic Observation Schedule (ADOS). Pakistani ASD trios were assessed using the Childhood Autism Rating Scale (CARS). The diagnosis of ASD was confirmed in all subjects using the ADI-R and ADOS and 3DI. ES was performed using the Thruplex DNA-Seq (Rubicon Genomics) Library Preparation Kit with the Agilent SureSelect V5 Exome Capture system for the affected child and both parents. Sequencing was performed on the Illumina Hiseq 2500. Variants were initially prioritized based on their read depth (20X). In-house control exome data was also used to assess the likelihood of variants being common in Iranian and Pakistani populations as well as gnomAD and the Greater Middle Eastern Variome Server. Classic algorithms will be used to predict the functional impact of variants on pathogenicity. Cross-referencing other ASD genetics studies was also used to support a decision whether a variant is implicated in ASD. All variants were Sanger-confirmed.

ES methods for 17-026: Written informed consent was obtained through an institutional review board (IRB)-approved research study at the Institute for Genomic Medicine at Columbia University (protocol AAAO6702). DNA was extracted from proband samples, only. The exome was sequenced on a NovaSeq 6000 with the Kapa Biosystem's Hyper Library Preparation Kit, and whole-exome captured with IDT xGen Exome Research Panel V1.0.

7.3.1.1 Supplemental Clinical Data

A. MASTER COHORT OF INDIVIDUALS WITH *ZNF292* MUTATIONS

17-001 (p25)

This is a 17-year-old boy of Northern European ancestry. He is the second of four children. All three siblings are reported as healthy with no family history of intellectual disability. He was born at 36 weeks of gestation with normal growth measurements. Birth weight was 3.4 kg (–0.26 SD), birth length 42 cm (–3.05 SD) and birth OFC 35 cm (–0.4 SD). The child has mild speech and gross motor delays. He was diagnosed with visuospatial dyspraxia at four years of age. A baseline brain MRI was performed and did not reveal any abnormalities.

A *de novo ZNF292* variant (c. 265C>T, p. Arg89*) was identified through a targeted NGS panel. No additional genetic findings were noted.

17-027

This is a seven-year-old male of Northern European ancestry born at 38 weeks of gestation. His birth weight was 3.675 kg (+0.23 SD) and birth length was reported as 'normal', with no specific measurements available. He first walked at around two years of age and said his first words close to four years of age. A brain MRI performed at 10 months of age was overall normal with mild plagiocephaly. An EEG at four year of age was reported as normal. The patient was last assessed at age seven years. His height was 119.9 cm (–0.5 SD), weight was 24.3 kg (+0.25 SD), and head circumference was 54.3 cm (+1.67 SD). The patient has global developmental delays, severe expressive language delays, hypotonia, autism spectrum disorders, and a tremor. Eye examination revealed myopia, astigmatism and strabismus. Medical history is also remarkable for tracheomalacia and eczema. Physical examination revealed mild dysmorphic facial features including a thin upper lip, smooth philtrum, upturned nasal tip, synophrys, and sparse but long eyebrows.

Previous laboratory tests include normal urine organic acids, mildly elevated urinary creatine, normal SNP microarray, normal Angelman/Prader-Willi methylation studies, normal Fragile X syndrome testing, and normal screening tests for congenital disorders of glycosylation (CDG). A

targeted multi-gene panel of 140 genes (IDNext; Ambry Genetics) was completed and detected a maternally inherited *MBD5* variant (p. Ser512Phe), classified as likely benign. A maternally inherited *FGD1* variant (p. Pro40Leu) was also detected and interpreted as a variant of uncertain significance (VUS). Exome sequencing was performed and detected a *de novo* variant in *ZNF292* gene (c. 433delA, p. Ser145Alafs*7), interpreted as likely pathogenic.

19-010

This 15 month old girl is the third child of healthy, non-consanguineous parents of German origin. She has a healthy sister and a developmentally delayed 2.5 years old brother.

She was born at a gestational age of 38 weeks by cesarean section with a weight of 2465 g (-1.86SD), a length of 46 cm (-1.91 SD) and an OFC of 35cm (0.54 SD). She started to sit at age 14 months and could not yet walk at age 15 months. At that time she started to speak first words. At age 15 months her height was 72 cm (-2.3SD), her weight was 6.4 kg (-3.54 SD) and her OFC was 46 cm (-0.49SD). Plagiocephaly and peripheral muscular hypertonia were noted.

Exome sequencing initially revealed the heterozygous variant c.1277G>A, p.(Trp426*) in KDM5C which was also identified heterozygously in the mother and hemizygously in the affected brother.

Re-analysis of the exome data for candidate variants revealed a frameshift variant in ZNF292 (c.1160delG, p.(Arg387Leufs*5)) (SureSelect Human All Exon V6 kit (Agilent Technologies, Santa Clara, USA), sequenced with 125 bp paired-end reads on an Illumina Hiseq 2500 system (Illumina, Inc., San Diego, USA)). Sanger sequencing (primers 5'-TAACTTCCAAGTTTGCCTAGCCA-3' and 5'-CAGAATTCTGGATCAAAGGGCCA-3') confirmed the variant in the individual.

19-011

This three-year and six-month old boy is the second child of healthy, non-consanguineous parents of German origin. The family history was unremarkable for intellectual disability or other neurologic issues. His birth was at a gestational age of 41 weeks with a weight of 3600 g (+0.1 SD), length 52 cm (-0.2 SD) and OFC 36.0 cm (+0.3 SD). Walking age was 19 months and he was able to speak first words at the age of 2 years. Evaluated at the age of three years and six-month he had mild to moderate developmental delay. His length was 102 cm (-0.7 SD), his

weight was 16.0 kg (+0.9 SD), and his OFC was 50 cm (-1.1 SD). Gross motor development was delayed with problems in coordination while fine motor development was near normal. He was able to speak single words, but pronunciation was not clear. Facial features were nonspecific with a trigonocephaly, facial asymmetry, short nose, long philtrum, hypertelorism, broad nasal tip, thin upper lip, wide intermammillary distance and mild syndactyly II/III of the toes.

Genetic testing for Beckwith-Wiedeman syndrome was unremarkable, as was a standard karyotype analysis. A *de novo* nonsense variant in ZNF292 (c.1360C>T, p.(Arg454*)) was identified by clinical exome sequencing (TruSight One Expanded Panel (Illumina, Inc., San Diego, CA); ~6,700 genes). Segregation testing by Sanger sequencing (primers 5'-GAACTGTGTGTAAAGGCTCTTCG-3' and 5'-GTAGGCTTGCCAATTCCTAAACC-3') confirmed the and showed that it is absent in both parents' DNA. Together with the results of the DNA-based fingerprinting (PowerPlex assay; Promega Corporation, Madison, USA) this confirms that the variant arose *de novo*.

17-003

This is a five-year old boy of Northern European ancestry. He has a healthy brother. He was born at 40 weeks of gestation. Birth weight was 3.8 kg (+0.49 SD), length 53 cm (+1.08 SD), and OFC 37 cm (+ 0.68 SD). On last assessment at age five years, his growth measurements were as follows: weight 19.6 kg (+ 0.39 SD), and height 109.0 cm (0 SD). He has global developmental delays with developmental regression noted at age three years. He had particular regression with verbal and language development, as he was able to speak full (though rather scripted) sentences early on and would sign, but subsequently lost all language/verbal communication. At age six years, he does not verbalize and uses a single sign (hand clapping for "yes" or "more"). He relies currently on a talker for simple single words (that consist mostly food requests). He was formally diagnosed with autism spectrum disorder (ASD) at 23 months of age, with ADHD. Facial examination is remarkable for mild dysmorphic facial features including a high arched palate, and low-set square-shaped ears. Neurological examination is remarkable for hypotonia. Brain MRIs performed at three and five years of age showed mild ventriculomegaly. Medical history is also remarkable for left esotropia that resolved by age 2.5 years, significant sleep dysregulation including delayed sleep onset and frequent awakenings, constipation, and possible restless leg syndrome. He was also identified to have low Ferritin levels and has been placed on iron supplements.

A *de novo ZNF292* variant (c.1567C>T, p.Gln523*) was identified by clinical exome sequencing. No additional genetic findings were reported.

19-013

This is a 4-year-old boy of European (Dutch) ancestry. He was born after an uneventful pregnancy at 38 weeks of gestation with a birth weight of 2780 grams (–1.25 SD). He had surgery for an umbilical hernia at age 1 year. He has developmental delay; he started walking at 20 months and spoke his first words between 9 and 16 months. He received a formal ADHD diagnosis at the age of 5 years. At the age of 6 years, a disharmonic IQ profile was noted (WPSSI-III-NL VIq 108, PIQ 89, VS 76). He is currently enrolled in special education. The child has significant behavioral issues, characterized by attention deficit, temper tantrums, and autistic traits (there is no formal autism diagnosis). Methylphenidate was prescribed, but was stopped due to side effects and he is now treated with an antipsychotic. A brain MRI showed an atypical (widened) configuration of the intraparietal sulcus on the right with diminished volume of the superior parietal lobule , further normal MRI configuration of the brain. He has subtle dysmorphic facial features.

A heterozygous *de novo ZNF292* variant (Chr6(GRCh37):g.87965517del; NM_015021.2:c.2170del (p.(Cys724fs)) was identified by exome sequencing.

17-022

This is a 3-year 8-month old male of Chinese ethnicity born full-term. He began babbling at eight months of age, spoke his first words at 24 months of age, and walked independently at 18 months of age. When last assessed at age three years, he had mild intellectual disability with speech delays, ASD, and ADHD. He was found to have paroxysmal high potential activity on EEG.

A *de novo ZNF292* variant (c. 2494_2495insTCACA, p. Ser832Ilefs*28) was detected by a targeted molecular inversion probe (MIP) panel. This variant was confirmed by Sanger sequencing.

18-007

This is a five-year old girl of French ancestry. The family history is remarkable for a half-sister with multiple congenital anomalies (with limited clinical data on this individual). She was born at 41 weeks of gestation following an unremarkable pregnancy. Her birth weight was 3.17 kg (– 2 SD), length 48 cm (– 2 SD) and OFC 35 cm (0 SD). On last assessment at age 3.5 years, her weight was 12.7 kg (– 2 SD), height 89 cm (– 2 SD) and OFC 50 cm (+1 SD). She had moderate developmental delays. She walked independently at 23 months of age and spoke her first words at 3.5 years with subsequently poor speech development, as she was unable to speak in full sentences at five years of age. Neurological exam was remarkable for diffuse hypotonia. She also had several behavioral issues including anxiety, emotional disturbances and sleep dysregulation. She does not have autistic features. Physical examination is also remarkable for several dysmorphic features inducing epicanthus, upslanted palpebral fissures, prominent forehead, small ears, large mouth and bulbous nose. Medical history is also remarkable for an atrial shunt, partial agenesis of the lacrimal ducts and ichthyosis. A baseline brain MRI was performed and did not reveal any abnormalities.

A *de novo ZNF292* variant (c.2814_2821dupCATTAAGG, p. p.Val941fs) was detected by a targeted NGS panel for 272 ID genes. An array CGH was previously performed and did not reveal any copy number abnormalities.

17-005

This four-year and one-month old boy is the first-born child for healthy, non-consanguineous German parents. The family history was unremarkable for intellectual disability or other neurologic issues. The boy was delivered by forceps after an abnormal CTG at a gestational age of 38 weeks. Birth weight was 2.76 kg (–1.28 SD), length 49 cm (–0.44 SD) and OFC 34.0 cm (+0.83 SD). He sat independently at 8-9 months of age, walked at 21 months of age and spoke his first words at 18 months of age. He was last evaluated at age four years and one month. His length was 94.5 cm (–1.8 SD), his weight was 12.0 kg (–2.6 SD), and his OFC was 50.8 cm (+0.23 SD). He had mild to moderate developmental delays, with delayed fine motor skills and speech comprehension, muscular hypotonia, delayed motor skills, and delayed speech comprehension. Facial features were nonspecific with a broad forehead, deep set eyes, relatively small mouth and large ears. He had a mild pectus excavatum deformity, simian crease of the

right hand and left foot, and fetal finger pads. He also had behavioral issues, anxiety and distractedness, hypothyreosis and frequent obstipation.

Genetic testing for Fragile X syndrome was unremarkable, as was a high-resolution microarray analysis. X-inactivation studies performed in the healthy mother showed non-skewed X-inactivation. A *de novo* frameshift variant in *ZNF292* (c.3066_3069del, p.Glu1022Aspfs*3) was identified by exome sequencing. This patient was previously published as part of a large series of individuals with neurodevelopmental disorders.(Popp et al., 2017)

17-006

This is a 10-month-old boy of Northern European Ancestry. He was born at 40 weeks of gestation. Birth weight was 3.32 kg (0 SD), length 53.5 cm (+1.3 SD) and OFC 34.5 cm (–0.6 SD). On last assessment at age 10 months, his weight was 9.5 kg (–0.4 SD), length 71.1 cm (–1.05 SD) and OFC 47 cm (+0.8 SD). He has global developmental delays and has not walked independently yet. Neurological exam is remarkable for hypotonia. A brain MRI performed at five months of age was normal. Eye exam is remarkable for nystagmus. He also has micrognathia, high arched palate, nuchal nevus flammeus and torticollis. Abdominal ultrasound was normal.

The child was identified to have a duplication of the X chromosome, including the *PAR2* gene that was interpreted as a copy number variant of uncertain significance. A *de novo ZNF292* variant (c. 0366_3069delAAGA, p.Glu1022Aspfs*3) was identified by exome sequencing.

19-012

This 19-year-old female individual is the second child of healthy, consanguineous parents of Syrian origin. She has a brother who also suffers from intellectual disability and growth retardation, while two other siblings are healthy.

She was born at term after an uneventful pregnancy with a weight of 2250 g (150 g below the 3. percentile). Early infantile development was unremarkable, but her motor development was delayed. She could sit the age of eight months and walked at the age of 22 months. She attended regular school up to the 5th class and then she attended a special school. At the time of examination, she presented with mild intellectual disability (estimated IQ 68). She could read

and write but had difficulties with math (knowledge of first or second class). She had growth retardation (142 cm tall, 14 cm below the 3. percentile) and microcephaly (47 cm, 6 cm below the 5. Percentile). Growth hormone was decreased (0.15 ng/ml; normal range: up to 7.00 ng/ml). Clinical examination revealed no other organic malformations. She had no epilepsy and a brain MRI was unremarkable.

Copy number analysis (SNP Chi; Illumina, Inc., San Diego, USA) excluded any submicroscopic aberrations and screening for fragile X syndrome was unremarkable. Homozygosity mapping identified three candidate regions on chromosomes 3, 5, and 12 of 11 Mb, 7.9 Mb, and 5.1 Mb, respectively. Exome sequencing (SureSelect Human All Exon V5 kit (Agilent Technologies, Santa Clara, USA), sequenced with 125bp paired-end reads on an Illumina HiSeq 2500 system (Illumina, Inc., San Diego, USA)) identified one candidate variant for which proof of relevance could not be achieved.

Re-analysis of the exome data for candidate variants revealed a frameshift variant in ZNF292 (c.3460_3463del, p.(Val1154Ilefs*7)). Sanger sequencing (primers 5'-ATCAGTGCCTCCAAAAGCTCC-3' and 5'-AAATGAGCTGGAAAAATGGGTGG-3') confirmed the variant in the individual.

17-007

This is a 24-year-old male of German, Italian and Hungarian ancestry. Family history is remarkable for a sister who was a ventricular septal defect (VSD), and a family history of cerebrovascular accidents. Birth growth measurements are not available. On his last assessment at age 24 years, his height was 160 cm (–2.3 SD) and his weight was 66.23 kg (–0.34 SD). He had mild-moderate delays in all areas of development, in addition to issues with memory and forgetfulness, with concerns regarding possible developmental regression. Neurologic exam was remarkable for several abnormalities including neuropathy, frontal release signs, dysarthria, unsteady gait, tandem walking and positive Rhomberg sign. His speech was noted to be dysarthric, with loss of spoken words reported on a weekly basis. He was formally diagnosed with ASD at age eight years. He also has subtle retinal pigmentation on formal ophthalmologic assessments. He has a history of fatigue with dyspnea and exertion. Exercise studies were abnormal suggesting respiratory muscle weakness. Echocardiogram and EKGs were normal. Medical history is also remarkable for constipation. Brain MRI performed at 23 years of age

showed atypical gyri within the posterior superior right parietal lobe without mass effect (images unavailable). A second brain MRI was performed at 24 years of age and was normal by report.

A *de novo ZNF292* variant (c. 3709dupA, p. Thr1237Asnfs*20) was detected by clinical exome sequencing. No additional genetic changes were reported.

18-003

This is a two-year old female of Hispanic ancestry. Family history is remarkable for her mother who has intellectual disability. Mother reports being diagnosed with "mental retardation" and was in special education throughout school. She could not obtain a job. However, she is able to take care of her daughter with the help of a home health nurse. Other medical issues in the mother include arthritis since 12th grade, with pain and swelling in her wrists and legs, low blood sugar and asthma. A maternal uncle has dyslexia. Mother was 27 years old at the time of the pregnancy. Prenatal history was remarkable for premature birth at 25 weeks of gestation due to a bulging amniotic sac. Birth weight was 1.8 kg (+1.7 SD), and birth length 34 cm (0 SD). The child was born prematurely at 2 weeks and had a complicated postnatal course. She spent 143 days in the neonatal intensive care unit (NICU) due to seizures, patent ductus arteriosus, and respiratory disease. Her post NICU course was complicated by multiple hospital admissions due to recurrent fevers, upper respiratory tract infections with chronic lung disease, acute otitis media and shunt malfunction.

On her last assessment at age two years, her weight was 13.3 kg (+1 SD), length 85 cm (0 SD) and head OFC 40 cm (– 5.5 SD). Developmental assessments at age 23 months revealed a developmental age of nine months. The child was not rolling, sitting, crawling, cruising or walking. Neonatal seizures occurred with episodes initially consisting of absence seizures, and tonic clonic activity. These episodes were controlled with phenobarbital and levetiracetam. Neurological exam was also remarkable for mixed tone (axial hypotonia and appendicular hypertonia). It is unknown whether the child has ASD or behavioral issues. Brain MRIs were remarkable for post-hemorrhagic hydrocephalus with focal cystic encephalomalacia in the right frontal lobe. A repeat brain MRI at age 1 year and 9 months showed stable right temporal horn and fourth ventricles with stable ventriculomegaly.

A maternally inherited *ZNF292* variant (c.3724delC, p.Gln1242Lysfs*5) was identified by exome sequencing. This variant was confirmed by Sanger sequencing.

17-008

This is a two year and 10 months old boy of Northern European ancestry. The child was the father's 7th and mother's 3rd child. Together, parents have two children including another boy who is healthy. Mother also has a 3-1/2-year-old boy from another partner who is healthy. Father has five children from another partner who range in age who include two girls (ages 13 and 15 years) and three boys (ages 6, 7 and 11 years). There is a strong family history of ADHD and learning problems, mood disorders, anxiety, and alcoholism. There is no family history of similar concerns as the proband. He was born full term via a repeat C-section. Pregnancy was complicated by gestational diabetes that was managed by NovoLog. No alcohol or other exposures were reported. Birth weight was 3.12 kg (–0.71 SD) and length 52.07 cm (+0.73 SD). On last assessment at age two years and 10 months, his height was 94 cm (+1.92 SD) and OFC 50 cm (+ 0.96 SD). He has mild ID. Neurological exam is remarkable for mixed tone (axial hypotonia, appendicular hypertonia). Brain MRI scan performed at age two years and six months showed a lacunar insult in the left subcortical white matter with diffusely decreased (likely secondary to an intra-uterine insult). There was no known prenatal history of any concern for intrauterine insult, and the only maternal disease was gestational diabetes. He was formally diagnosed with ASD and has self-injurious behavior. Eye examination is remarkable for nystagmus. The child was described to have dysmorphic features. Medical history is also remarkable for sensorineural hearing loss due to a paternally inherited variant in *GJB2* (c.224G>A, p. Arg75Gln).

A *de novo ZNF292* variant (c. 3812C>A, p. Ser1271*) was identified by clinical exome sequencing. Exome sequencing also detected the paternally inherited variant in *GJB2* (c.224G>A, p. Arg75Gln). This pathogenic variant has been previously associated with autosomal dominant progressive sensorineural hearing loss with both presence and absence of palmoplantar keratoderma. Functional studies of this *GJB2* variant show the resultant protein prevents the formation of functional channels, therefore this child's hearing loss is attributed to the *GJB2* mutation.

17-023

This is a six-year-old male of Chinese Han ethnicity who was born full-term. Birth weight was 3.5 kg (–0.1 SD), with a length of 52 cm (+0.7 SD). When assessed at age four years, his height was 102 cm (+0.28 SD) and his weight was 16 kg (–1.85 SD). Head circumference at six years of age was 49 cm (+1.47 SD). The patient was described to have normal intelligence, speech delays, autism, feeding issues, and sleeping problems. An EEG at age four years showed sharp waves.

A *de novo ZNF292* variant (c. 4416_4417insT, p. Ser1473Phefs*) was detected by a targeted MIP panel. This variant was confirmed by Sanger sequencing.

17-009

This is a three-year-old girl of mixed Caucasian-Hispanic ancestry. She has three healthy siblings. She was born at 40 weeks of gestation. Birth weight was 2.78 kg (–1.23 SD), length 47 cm (–1.1 SD) and OFC 34.5 cm (–0.2 SD). On last assessment at age three years, weight was 8.8 kg (–3.7 SD), length 76 cm (–4.78 SD) and OFC 46 cm (–1.64 SD). She has moderate developmental delays. She walked independently at 15 months of age and spoke her first words at 18 months of age. At age three years, her vocabulary consisted of 20 words with two-word phrases. She could follow two to three step commands. Facial examination was remarkable for flat nasal bridge with a bulbous nose, mild ocular hypertelorism, and bilateral fifth finger clinodactyly. Neurological exam showed moderate-severe hypotonia. A baseline Brain MRI scan at age 15 months showed diffuse cerebellar hypoplasia with cerebellar asymmetry with possible clefts in the cerebellum. Medical history is also remarkable for feeding difficulties, in addition to esotropia and strabismus status post-surgical correction. She has not had any seizures, to date.

A *de novo ZNF292* variant (c. 4897A>T, p. Lys1633*) was identified by exome sequencing. A *de novo* variant in *DHX30* (c.997G>T, 43, p.Ala333Ser) was also reported and predicted to be benign by in silico prediction programs.

17-010

This is a five year and three months old boy of Irish, Scottish and English ancestry. He has a twin sister, and an older maternal half-brother. His birth weight was 2.25 kg (–2.1 SD). He has situs

inversus, history of feeding difficulties and failure to thrive. On last assessment at age five years and three months, his weight was 15.8 kg (–1.35 SD), height 107 cm (–0.4 SD) and OFC 50 cm (–0.72 SD). He has moderate developmental delays, with speech and language delays. He walked independently at 11 months of age and spoke his first words at 18 months of age. Neurological examination is remarkable for a cortical thumb. Currently, he is overall making good progress with activities of daily living.

A *de novo ZNF292* variant (c.518C>T, p. Gln1792*) was identified by clinical exome sequencing. Exome sequencing also identified compound heterozygous variants in *COG1*. The first variant (c.2084_2085ins17, p.Gly696Leufs*5) was paternally inherited and interpreted to be likely pathogenic. The second variant (c.1580T>C, p.Val527Ala) is maternally inherited and interpreted as a VUS. Follow-up testing, including carbohydrate deficient transferrin analysis, protein C, thrombin time, antithrombin III, protein S and CMP, were normal following consultation with CDG. Therefore, there is insufficient evidence that these variants are pathogenic.

18-004

This is a nine-year old boy of Northern European and Ashkenazi Jewish Ancestry. The family history is overall unremarkable. The child was born at 39 weeks of gestation. Birth weight was 2.9 kg (– 1 SD). On last assessment at age 9 years and 1 month, his weight was 31.8 kg (– 1.75 SD), height 121 cm (–2 SD) and OFC 54 cm (+1 SD). He has global developmental delays and speech delays with possible seizures at age three years. Neurologic exam was remarkable for hypotonia. He also has ASD, ADHD and multiple behavioral issues including impulsivity and aggressive behaviors. Medical history is also remarkable for ocular refractive errors requiring glasses and subtle dysmorphic facial features including a long face with low-set ears. He also had an inguinal hernia that was surgically repaired at age one year.

A *de novo ZNF292* variant (c.5515_5569del55, p.Glu1839Hisfs*6) was identified by trio-clinical exome sequencing. This variant was Sanger confirmed.

17-011

This is a seven-year-old girl of Northern European ancestry. Family history is overall unremarkable. She was born full-term. Birth weight was 2.45 kg (–1.86 SD), length 49 cm (–0.2 SD) and OFC 34.5 cm (–0.2 SD). On last assessment at age seven years, her weight was 18 kg (–1.75 SD), height 112 cm (–1.8 SD) and OFC 50 cm (–1.14 SD). She has mild to moderate developmental delay and a long-standing history of feeding difficulties. She first walked at 23 months of age and spoke her first words at age three years. She attended school in a mainstream class up until the age of six years. She was then enrolled in specialized education due to difficulties with reading. Facial examination was remarkable for cup-shaped ears. Echocardiogram, abdominal ultrasounds, and brain MRI were all normal. She is described to have a friendly personality without any obvious behavioral issues. Eye examination was remarkable for strabismus.

Karyotype was normal. Floating-Harbor syndrome was initially clinically suspected. However, *SRCAP* gene sequencing did not detect any mutations. A multi-gene panel of 272 genes associated with neurodevelopmental disorders was performed and detected a *de novo* variant in *ZNF292* (c.5959_5965dup, p.Gly1989Alafs*9).

17-012 (p21)

This is a six year and three-month-old boy of Northern European ancestry. He has a healthy younger sister. He was born at 41 weeks of gestation. Birth weight was 2.62 kg (–1.5 SD), length 50 cm (–0.06 SD) and OFC 36 cm (+0.06 SD). At age three years and three months, his weight was 15.3 kg (+0.5 SD), height 103 cm (+1.79 SD) and OFC 48cm (– 0.99 SD). He has delayed fine motor and speech development, without any intellectual disability. He independently walked at 14 months of age and said his first words between 12-15 months of age. A brain MRI was performed at 16 months of age and was normal besides a non-specific cyst of the left ventricle. He was formally diagnosed with ASD and ADHD. Medical history is also remarkable for dilated cardiomyopathy, feeding issues, and a pectus adductus deformity. Dysmorphic facial features include a prominent forehead, mild enophthalmos, small chin, thin upper lip, and full cheeks.

A targeted NGS panel of 456 genes associated with neurodevelopmental disorders detected a *de novo* variant in *ZNF292* (c. 6142_6145del, p.Lys2048Valfs*11).

17-013 (p1)

This is an 18-year-old man of Northern European ancestry. Family history is overall unremarkable. Birth weight was 3.2 kg (–0.58 SD), length 50 cm (–0.06 SD), and OFC 34 cm (–0.83 SD). On last assessment at 18 years of age, weight was 52 kg (–1.67 SD), height 176 cm (+0.15 SD) and OFC 53 cm (–1.41 SD). He has severe global developmental delays, with regression of acquired skills first noted at age 14 months. He walked independently at 14 years of age. He has severe speech and language delays and is non-verbal, with autistic features, aggressive behaviors, and sleep dysregulation. Neurological examination is remarkable for hypotonia. He also has abnormal skin pigmentation on both arms with hyperpigmented and hypopigmented streaks. Clinical photos were significant for laterally prominent ears, wide thick lips, tented upper lip, short philtrum, prominent eyebrows, and deep-set eyes. A baseline brain MRI scan was performed and did not reveal any abnormalities.

Targeted NGS of 456 genes associated with neurodevelopmental disorders detected a *de novo* variant in *ZNF292* (c. 6160_6161del, p. Glu2054Lysfs*14).

17-014

This is a six-year seven-month old male from a non-consanguineous family. He has a healthy sister. His parents first became concerned regarding his development at age 10 months. His first words were at 18 months and his first phrase was at two years of age. At age four years, he spoke only a few phrases. He was diagnosed with high-functioning ASD at age six years and seven months. Based on his ADI-R and ADOS assessments, he did not have sufficient eye contact, but was able to use facial expression to communicate with others. There was no evidence of pretend play, but he was interested on other children. He could talk, nod and shake his head for "yes" and "no". Although he could talk, he had problems with conversations and had neologism, inappropriate talk and questions, repetitive words, pre-occupation and restricted behaviors. He did not show obsessive behaviors or hand mannerism at the time of the test. He was described to have mild developmental delays. A baseline EEG was performed at age six year and did not reveal any abnormalities.

7.3.1.1.1 Exome sequencing identified a *de novo* variant in *ZNF292* (c.6160_6161del, p. Glu2054Lysfs*14).

19-001

This 18-year old male was born following an uneventful pregnancy at 40 weeks of gestation. He was born to an outbred family from the Punjab region, and both parents belonged to the same cultural clan. Birth weight and height was not recorded but were normal by report. During the early neonatal period, he was unable to attain early normal milestones such as lifting his head, among others. He could not roll, sit, crawl, cruise or walk. His growth was delayed partially due to significant constipation, and inadequate feeding. Motor abilities were also delayed. Eventually, he was diagnosed with moderate intellectual disability (ID). His speech was delayed and dysarthric. He could only speak a few words at age 18 years. A brain CT scan was performed and did not reveal any abnormalities. Past medical history is also remarkable for early-onset hypotonia. On last assessment, he had no cardiac, skeletal, or gastrointestinal abnormalities. Vision was normal. He was socially withdrawn with minimal eye contact. He had multiple stereotypies, as well as behavioral issues including aggression and lack of cooperation. He is described as emotionally attached to his mother. He also has difficulties with attention and restricted communicative behavior. All of these issues have prevented school placement. His IQ has consistently been low since childhood.

7.3.1.1.2 Exome sequencing identified a ZNF292 variant (c.6160_6161del, p. Glu2054Lysfs*14), which was confirmed by Sanger sequencing.

17-021

This is a six-year-old female of Puerto-Rican, Irish, Greek, Swedish, and Lebanese ancestry. She was born at 38 weeks of gestation. She is the only child to her parents. She began walking at 19 months of age and spoke her first words around age two years. She has a history of early onset hypotonia that normalized over time. On last assessment at age six years, weight was 18 kg (–0.71 SD), height 113.4 cm (–0.26 SD), and head circumference 48 cm (–1.66 SD.) She had moderate intellectual disability and ASD, with no other behavioral issues. A previously performed brain MRI scan was normal. Formal developmental assessment through WPPISI-IV reported the following scores: Visual spatial 70, fluid reasoning 94, working memory 87 and

processing speed 64. Adaptive Behavior Assessment reported a general adaptive composite of 67. Beery-Buktenica Test of Visual and Motor Integration score was 88. Dysmorphic facial features were noted including metopic prominence, broad nose, short philtrum, full lips, large everted ears, and a high arched palate.

Trio exome sequencing detected a *de novo ZNF292* variant (c. 6160_6161delGA, p. Glu2054Lysfs*14). Testing also detected two variants in the *MPDZ* gene. The first variant (p. Lys1198Arg) is maternally-inherited and the second variant (p. Thr257Met) is *de novo*. Both variants were interpreted as VUS. Bi-allelic pathogenic variants in this gene have been reported in association with non-syndromic hydrocephalus, which was not present in this patient.

17-024

This is a three-year 11-month old male of Chinese Han ethnicity who was born at 40 weeks of gestation. Birth weight was 3.5 kg (–0.1 SD) and length 50 cm (–0.06 SD). The child's first words were at age eight months and he first walked at age 18 months. When last assessed at age three years, height was 94 cm (–0.24 SD), weight was 12 kg (–0.13 SD), and head circumference was 47.5cm (–1.28 SD). This child was noted to have severe intellectual disability, speech delay, ASD, ADHD, and sleeping problems.

Targeted sequencing through a MIP panel detected a *de novo ZNF292* variant *(*c. 6279_6280insA, p. Arg2094Thrfs*10), which was confirmed by Sanger sequencing.

18-006

This is a two-year-old female of Hispanic ancestry. Family history is remarkable for an older brother who had speech delays that reportedly improved with speech therapy and is currently in a regular classroom setting. The child was born full term. Birth weight was 3.65 kg (+0.5 SD). On last assessment at age 36 months, weight was 10.3 kg (–2.92 SD), height 84.2 cm (–2.5 SD) and OFC 46.5 cm (–2 SD). She had mild gross and fine motor delays with more marked speech delays. She walked independently at 15 months and spoke her first word "mama" at 12 months of age. However, at age 3 years, she only spoke 2-3 words in total, and understood simple commands (in a bilingual setting). Hearing assessment was normal. Neurological exam was notable for mild hypotonia, without other abnormalities. She had multiple stereotypies with hand

flapping behaviors. Formal Neuropsychiatric evaluations at age 13 months and 20 months identified her to be at "high risk for autism". A formal ASD diagnosis was not given due to her young age, however she was assessed to be on the ASD spectrum. She was also noted to be quite hyperactive with multiple ADHD features. Physical examination was also remarkable for several dysmorphic facial features including a broad nasal bridge, bulbous tip, short palpebral fissures, short philtrum, large ears, cupped teeth that are also widely spaced, and micrognathia. Medical history is also remarkable for feeding issues. She has never had any seizures and a brain MRI has not been performed, to date.

Trio-based exome sequencing identified a *de novo ZNF292* variant (c.6343C>T, p.Arg2115*). She was also identified to have a single variant in *CEP152* (c.4127delA, p.Lys1376Serfs*22), with no additional *CEP152* variants.

17-015

This is a two-year old boy of Northern European ancestry. He was born at 41 weeks of gestation. Birth weight was 4.4 kg (+ 1.7 SD). Length and OFC measurements were within the normal range by report. On last assessment at age two years, his weight was 12.5 kg (–0.14 SD), height 94.8 cm (+2.15 SD) and OFC 50.8 cm (+1.53 SD). He was noted to have early developmental delays. At age 21 months, his fine motor development was assessed to be at 15 months, and his gross motor development at 19 months of age. He walked independently at 20 months and spoke his first words at age one year. He subsequently had loss of speech development, and at age 2.5 years, only spoke a few words. He also has autistic features but has not undergone a formal ASD evaluation. He has significant feeding issues from early on in life with refusal to eat or drink for long periods of time. He was recently hospitalized for dehydration due to refusal to drink. Eye examination was remarkable for esotropia and amblyopia of the right eye, with bilateral mild myopia and astigmatism. Facial examination was remarkable for broad forehead, large ears, pointed chin and prominent eyes. He had bilateral simian creases. He also has hypermobility of the ankle joints. He had recurrent upper respiratory and ear infections. He was also found to have unexplained normocytic anemia (Hemoglobin level 6.5 mmol/L; N 7.4-8.7). A brain MRI was performed although results were not reported.

Exome sequencing identified a *de novo* variant in *ZNF292* (c.6541C>T, p. Arg2181*). He was also identified to have a maternally inherited pathogenic mutation in *CC2D2A* (p.Val1097*). Disease causing mutations in this gene are associated with Joubert syndrome inherited in an autosomal recessive fashion. The child did not have features of Joubert syndrome, and a second variant in this gene has not been identified. Therefore, he is assessed to be a carrier.

18-005

This is a two-year old female born to an African American family. Family history was overall unremarkable. The child was born at 42 weeks of gestation. Birth weight was 3.5 kg (0 SD). On last assessment at age two years and one month, weight was 12.25 kg (+0.13 SD), height 91.44 cm (+1.63 SD) and OFC 45.72 cm (–1.23 SD). The child had developmental delays with speech regression. She stood independently before 12 months of age. At age two years, she was not walking, only cruising occasionally and had no speech development. She had babbled in infancy and said "mama" at 12 months but stopped speaking at age 15 months. She had staring spells and hypotonia, as well as ASD with multiple stereotypic behaviors. Neurological examination was also notable for ataxia. A brain MRI showed abnormal T2 signal intensities of unclear clinical significance.

7.3.1.1.3 Clinical exome sequencing identified a *de novo ZNF292* variant (c.6661_6664delTTAG, p. Leu2221Serfs*10).

B. CLINICAL DATA OF INDIVIDUALS WITH ADDITIONAL VARIANTS IN *ZNF292*

17-002a/a2

This family consists of two affected twin sisters, as well as four older siblings (three brothers, one sister) who are reported healthy. The affected twins are one year and one-month old girls. Family is of Scottish, German and Swedish Ancestry. The twins were born at 29 weeks of gestation following premature rupture of membranes at 21 weeks of gestation. Prenatal history was remarkable for twin-twin transfusion (a1 recipient; a2 donor), and intraventricular hemorrhage type (IVH grade II). Birth growth measurements are unavailable.

17-002 a1: This girl has a history of global developmental delays, infantile spasms, and cortical visual impairment. She also had feeding issues, with refusal to bottle feed, and concerns

regarding aspiration. When last assessed at age one year and one month, length was 70 cm (–1.59 SD), weight 9.3 kg (–0.49 SD), and OFC 41 cm (–3.42 SD). She has. She had not walked independently and did not speak any words. Age of onset of epilepsy was seven months, with more than 300 seizure episodes occurring daily initially. Seizure frequency subsequently decreased to 100 episodes per day. EEGs showed hypsarrhythmia. She has been treated with multiple anti-seizure medications including phenobarbital and zonisamide, as well the ketogenic diet. A baseline brain MRI performed at less than one year of age showed pontine and cerebellar hypoplasia, with a preserved vermis, and diffuse cerebral atrophy.

17-002 a2: This girl also has a history of global developmental delays, infantile spasms, chronic lung disease, and IVH grade II. She had regression of developmental skills with onset of epilepsy, with more than 600 spasms a day in clusters. EEGs showed hypsarrhythmia. She has feeding issues and does not take solids well. Ophthalmology evaluation was normal. A brain MRI scan showed pontine and cerebellar hypoplasia with a preserved vermis, thin corpus callosum and evidence of left-sided IVH.

A *de novo ZNF292* variant (c. 1408A>G, p. Ile470Val) was identified on clinical exome sequencing. She was also identified to have a maternally inherited variant in *EXOSC8* (c.815G>C, p.Ser272Thr) that was interpreted as a VUS, deletion-duplication analysis of this gene was normal.

17-028

This is a 12-year-old male of Northern European ethnicity. Family history is remarkable for a history of mental health issues in his mother. The child was born with penile chordee and ankyloglossia. When last assessed at age 12 years, height was 156 cm (+1.6 SD), weight was 48 kg (+0.63 SD), and head circumference was 56 cm (+1.56 SD). The child was described to have moderate intellectual disability with speech delays, ASD and self-mutilation. Dysmorphic facial features included a long face, epicanthal folds, and prominent ears.

Exome sequencing detected a *ZNF292* variant (c. 1858G>T, p. Glu620*). This variant was not paternally inherited; however, the mother was not available for testing.

17-025

This is a five-year 11-month old male of Chinese Han ethnicity born at 36 weeks of gestation. Birth weight was 2.9 kg (–1.06 SD), length 50 cm (–0.06 SD), and head circumference 40 cm (+2.54 SD). He spoke his first words at 8 months of age and first walked at 15 months of age. When last assessed at five years of age, his height was 110 cm (+ 0.03 SD), weight was 22.5 kg (+1.31 SD), and head circumference 80 cm (+2 SD). He was reported to have intellectual disability, speech delays with language regression, and ASD. Notably, his father has behavioral issues and possibly mild ID (not formally assessed).

Targeted sequencing through a MIP panel detected a paternally inherited *ZNF292* variant (c. 1897C>T, p. Arg633*) in the *ZNF292* gene. This variant was confirmed by Sanger sequencing.

17-026

This is a 13-year-old female of German, Irish, and Jewish ethnicity. She was born at 34 weeks of gestation. Birth weight was 2.46 kg (–1.83 SD). Family history is significant for her father who has Sturge-Weber syndrome, seizures since birth, arteriovenous malformation (AVM), and cardiac issues. The child's mother has compulsive movements and undiagnosed mental issues. The child was described to have early motor and speech delays. She first babbled at eight months of age and first walked at age one year. She was found to have mild mitral valve regurgitation and trivial aortic insufficiency. Her last cardiac evaluation showed trivial mitral regurgitation and no obvious aortic insufficiency. Medical history was also remarkable for vesicoureteral reflux requiring surgical intervention. When last assessed at age 13 years, height was 153 cm (–0.39 SD) and weight 48kg (+ 0.19 SD). Head circumference was 54 cm at age 12 years (+0.14 SD). The child has learning disabilities and attends a special school with 1:5 class ratio. She required physical and occupational therapy until 11 years of age.

She also had clinical seizures shortly after birth that consisted of stiffening of the right side of the body with arching of the neck. She was subsequently formally diagnosed with a seizure disorder at four years of age, then described as cessation of activity and rapid eye blinking. Seizures were managed with ethosuximide with good response. A few months later, she developed myoclonic movements involving the hands (unilateral and alternating either left or right side) initially occurring before going to sleep, then later upon waking up. She currently has generalized

epilepsy with myoclonic and absence seizures. Current seizures are characterized by arm jerks, arm rash, dizziness, excessive fatigue during day, sudden cessation of activity, very rapid eyelid blinking, myoclonic movements of the extremities, mostly in the arms bilaterally. These episodes occur immediately prior to going to sleep, with left arm and leg shaking, post-ictal confusion. Seizures are refractory to current treatment. A baseline brain MRI was performed and did not reveal any abnormalities. Initial EEGs did not demonstrate any epileptiform abnormalities. Follow-up EEGs, however, showed rare fragments of irregularly generalized spike and slow wave discharges that also activate with photic stimulation, generalized spike and wave discharges with some left frontal predominance. She also has ADHD, a tic disorder, ataxia, dysmetria, and dysarthria. Additional clinical features include excessive fatigue, dizziness, and a birthmark on the sacral area.

Previous testing includes a normal chromosome microarray. Research exome sequencing detected a *ZNF292* variant (c. 1905delA, p. Lys637Argfs*10). This variant was confirmed by Sanger sequencing. Inheritance of this variant is unknown.

17-004 (p22)

This is a nine-year-old boy of Northern European ancestry. He is the only child to his parents. He was born at 39+5 weeks of gestation. Birth weight was 3.91 kg (+0.71 SD), length 53 cm (+1.08 SD), and OFC 36 cm (+0.06 SD). On last assessment at age nine years, his weight was 26 kg (–0.7 SD), height 132 cm (–0.24 SD), and OFC 54 cm (+1.09 SD). He has mild ID, with normal speech. He walked at 17 months of age and spoke his first words at age two years. Neurological exam is remarkable for axial hypotonia with abnormal gait control. He was also found to have ADHD. Eye exam is remarkable for nystagmus. Clinical history is also remarkable for unilateral cryptorchidism. Facial examination revealed several dysmorphic facial features including a relatively large mouth with thick lips, prominent alae nasi with a mildly broad nasal tip and mildly down-slanting palpebral fissures.

A targeted NGS panel identified a *ZNF292* variant (c. 3053dup, p. Gln1019Thrfs*31). This variant was not maternally inherited. The father was not available for testing. No additional genetic findings were reported.

18-001

This is a 12-year-old female of Northern European ancestry. The family history is significant for a mother with a carnitine deficiency of unknown etiology. The proband's sister also has Crohn's disease. The child was born at 39 weeks of gestation. Birth weight was 3.4 kg (0 SD) and length 48.2 cm (–1 SD). Head size assessed to be large, but OFC measurements are not available. The child first walked at age 12 months. When last assessed at 12 years of age, her height was 137cm (–1.89 SD), weight was 35.38 kg (–1 SD), and OFC was 55 cm (+1.4 SD). She had short stature treated with growth hormone (GH) therapy. She was described to have mildly dysmorphic facial features including downslanting palpebral fissures and a flat nasal bridge. She has an atrial septal defect, dysmotility of the GI tract, and fructose intolerance. Development was delayed, particularly speech. She spoke her first words at ~3 years of age and used sign language while in preschool.

Exome sequencing detected a *ZNF292* variant (c. 5374_5378delAACAT, p. Asn1792Serfs*11). Inheritance of this variant is unknown. Exome sequencing also detected a *KMT2E* variant (c. 877T>G, p. Val296Gly) of unclear significance.

17-016

This is a four-year old girl with no significant family history. She was born full term at 40 weeks following an uncomplicated pregnancy by spontaneous vaginal delivery. Birth weight was 2.815 kg, length 48.26 cm and head circumference 33 cm. The child had infantile spasms at age 3.5 months with seizures subsequently evolving to Lennox-Gastaut syndrome (LGS). She has profound global developmental delays, with no period of normal development noted. She is non-verbal, non-ambulatory, and does not sit or roll over. She feeds by mouth and takes formula, thickened liquids and pureed foods due to history of aspiration. Seizures have been poorly responsive to the ketogenic diet, a vagal nerve stimulator, and multiple anti-epileptic medications. Neurological examination was significant for central hypotonia and spastic diplegia, with intermittent dystonic posturing of the legs. She has been diagnosed with cortical visual impairment without retinopathy. EEGs have shown multifocal epileptiform discharges and hypsarrhythmia. Brain MRI showed dysgenesis of the corpus callosum and severe paucity of the white matter considered sequelae of periventricular leukomalacia with enlarged ventricles. Exam at 3 years and 11 months was notable for dysmorphic features including long eyelashes, arched

eyebrows, bilateral epicanthic folds, long palpebral fissures, synophrys, high arched palate, small uvula, prominent alveolar ridges, broad nasal bridge, micrognathia, bifrontal narrowing of the forehead, thickened auricular helices as well as microcephaly, mild pectus carinatum, scoliosis, tapered fingers, clinodactyly of the 4th toes bilaterally and small 5th toenails. Most recent measurements at age 5 years and 7 months were height 100 cm (–2.55 SD), and length 15.6 kg (–1.69 SD). Head circumference was 46 cm (–2 SD) at 5 years and 5 months.

7.3.1.1.4 Exome sequencing detected a *de novo ZNF292* variant (c. 6578A>C, p.Tyr2193Ser). No additional genetic findings were reported.

17-017 Family (p23)

This is a family of Northern European ancestry that consists of two affected siblings, an affected mother, an affected maternal aunt, and an affected maternal cousin (Figure A- 6**A**). The first affected child (17-017a1) is a nine-year-old boy. His birth weight was 2.86 kg (–1 SD), length 49 cm (–0.4 SD) and OFC 32.5 cm (–1.49 SD). On last assessment at age nine years, his weight was 19.2 kg (–2 SD), height 119 cm (–2 SD) and OFC 48.5 cm (– 3 SD). He has intellectual disability and epilepsy. He has recurrent febrile seizures, beginning at age 22 months. A brain MRI scan was performed and revealed partial agenesis of the corpus callosum. Eye examination was remarkable for nystagmus and strabismus. Medical history is also remarkable for atopic dermatitis and TCD4 lymphocytes with B lymphopenia. Dysmorphic features noted include sparse eyebrows, short nasal bridge, and mildly thin upper lip. Targeted NGS was performed and detected a *ZNF292* variant (c. 6661_6664 del, p. Leu2221Serfs*10). This genetic change was also identified in the patient's mother, who also had mild intellectual disability and does not work. The clinical symptoms are similar, but less severe for the older brother (17-017a2) who has no growth retardation and his OFC is just under 2 SD below the mean. He tested negative for the familial *ZNF292* mutation and additional genetic investigations are currently pending. No brain MRI was performed for the older brother. The father also required some specialized education, presented with convulsive seizures, but is currently employed in a normal environment. A maternal aunt and her daughter also present with intellectual disabilities and do not carry this *ZNF292* gene variant. All in all, close examination of the clinical and molecular features of this family suggest that other environmental and/or genetic factors (in addition to the *ZNF292* variant) might be responsible for the cognitive issues in this family.

17-018 Family (a1/a2)

This family consists of two boys with developmental delays. Exome sequencing detected a *ZNF292* variant (c. 6935dupA, p. Ser2313Valfs*2) in both siblings, with very limited clinical and molecular data.

17-020 Family (a1/a2/a3)

This is a family of Northern European ancestry that consists of three affected siblings and their clinically unaffected father (Figure A- 6 **B**). The oldest affected sibling (17-020a1) is a 17-year-old boy. He was born at 40 weeks of gestation. Birth weight was 3.08 kg (–0.88 SD), length 52 cm (+0.7 SD) and OFC 34 cm (–0.83 SD). On last assessment at age 17 years, his weight was 72 kg (+0.6 SD), height 175 cm (+0.25 SD) and OFC 52 cm (–2 SD). He first walked at 32 months of age. He had mild intellectual disability and speech delays. He also exhibits aggressive behavior. Medical history is also remarkable for joint hypermobility. A baseline brain MRI scan was performed and found to be normal.

The second affected child (17-020a2) is a 12-year-old boy. He was born at 38 weeks of gestation. Birth weight was 2.58 kg (–1.57 SD), length 46.5 cm (–1.37 SD) and OFC 33.5 cm (–1.05 SD). On last assessment at age 12 years, his weight was 41.6 kg (+0.1 SD), height 179 cm (+3.9 SD) and OFC 54 cm (+0.1 SD). He first walked at 10 months of age. He has severe intellectual disability and profound speech delay. He also had autistic features, with aggressive behavior. Brain MRI scan revealed periventricular nodular heterotopia. Medical history is also remarkable for joint hypermobility and mild dysmorphic facial features including a high forehead, sparse hair in the temporal region, thin lips, upturned mouth, and small chin (notably, these features were also present in the oldest affected brother; 17-021a1, as well).

The third affected child (17-020a3) is a nine-year-old girl. She was born at 39 weeks of gestation. Her birth weight was 3.84 kg (+0.91 SD), length 45.5 cm (–1.74 SD) and OFC 32 cm (–1.77 SD). On last assessment at age nine years, her weight was 26 kg (–0.8 SD), height 129 cm (–0.6 SD) and OFC 49 cm (–2.3 SD). She walked at age 10 months. She has mild intellectual disability with mild speech delays. Medical history is also remarkable for joint hypermobility. A brain MRI scan was performed and did not identify any abnormalities.

The *ZNF292* variant (c. 7304_7308del, p. Val2435Glyfs*4) was detected on a targeted NGS and found to be paternally inherited from a clinically unaffected father. The father has no intellectual disability or behavioral concerns, to our knowledge. The variant was present in the three affected siblings and absent in the unaffected brother. Patients 17-020a2 and 17-020a3 also had a 9p24.1 deletion measuring 275-371 kb in size [(arr 9p24.1(6,055,497-6,330,726) (hg19)x1]. The 9p24.1 deletion was inherited from the unaffected father, but absent in the affected brother (17-020a1) and the unaffected brother. This history suggests that other environmental and genetic factors (in addition to the *ZNF292* variant) might be responsible for the cognitive issues in this family.

17-019

This is a 10-year-old boy born to a mother with gestational diabetes that was diet controlled. He weighed 4.36 kg (+1.63 SD) when delivered at 36 weeks of gestation. He had testicular torsion as an infant which required removal of the involved testicle. When last assessed at age 10 years, height and weight were at +2 SD above the mean. He has a history of mild ASD, ADHD, but no intellectual disability with a normal IQ. From a developmental standpoint, he walked at one year of age and spoke his first words at nine months of age. There were no early delays, but there were concerns for some unusual behaviors. When playing with cars, he tended to line up the cars and liked to play with the same things repetitively. On his most recent assessment, he had just finished 4th grade in a regular classroom setting, with good grades. He is described as socially immature and is currently receiving behavioral therapy. He required occupational therapy for sensory issues early on in life as well. He has astigmatism and myopia. He has motor tics. Facial examination shows hypertelorism.

Clinical exome sequencing detected a *de novo ZNF292* variant (c. 7062_7064delTAA, p.Asn2354del). No additional genetic findings were reported.

Table A- 11 Additional families identified with Variants in ZNF292 (N = 12 families, 15 individuals)

ID	**17-002**	**17-028**	**17-025***	**17-026**	**17-004**	**18-001**	**17-016**	**17-017***	**17-018a1/ a2**	**17-020***	**17-019**
***ZN9292* exon**	8	8	8	8	8	8	8	8	8	8	8
Genomic coordinates	6: 87964755	6: 87965205	6: 87965244	6: 87965252	6: 87966400dup	6: 87968721_87968725delAACAT	6: 87969925	6: 87970008_87970011del	6: 87970282dupA	6: 87970651_87970655	6: 87970409_87970411delTAA
cDNA change	c.1408A>G	c.1858G>T	c.1897C>T	c.1905delA	c.3053dup	c.5374_5378delAACAT	c. 6578A>C	c.6661_6664del	c.6935dupA	c.7304_7308del	c.7062_7064delTAA
Predicted protein alteration	p.Ile470Val	p.Glu620*	p.Arg633*	p. Lys637Argfs*10	p.Gln1019Thrfs*31	p.Asn1792Serfs*11	p.Tyr2193Ser	p.Leu2221Serfs*10	p.Ser2313Valfs*2	p.Val2435Glyfs*4	p.Asn2354del
Inheritance	DN	Not paternally-inherited	Paternally-inherited	ND	Not maternally-inherited	ND	DN	Familial*	ND	Familial*	DN
Method of detection	ES (GeneDx)	ES (GeneDx)	MIP targeted sequencing	Research ES	Targeted NGS (456 NDD genes)	ES	ES (UDN)	Targeted NGS (456 NDD genes)	ES	Targeted NGS (456 NDD genes)	ES (GeneDx)
Notes/ rationale for inclusion here	Missense variant	Mother not tested, paternal testing incomplete	Targeted testing rather than exome sequencing, paternally inherited with limited clinical data on the father	Paternal testing incomplete	Paternal testing incomplete	Paternal testing incomplete	Missense variant	Targeted testing rather than exome sequencing, variant does not segregate with phenotype (Figure A- 6)	Paternal testing incomplete	Targeted testing rather than exome sequencing, variant does not segregate with phenotype (**See figure S2B**)	In=frame deletion, of unclear significance

Abbreviations: DN, *de novo*; ES, exome sequencing; ND, no data; VUS, variant of uncertain significance; NDD, neurodevelopmental disorders; UDN, Undiagnosed Disease Network.

7.3.2 Supplemental Figures

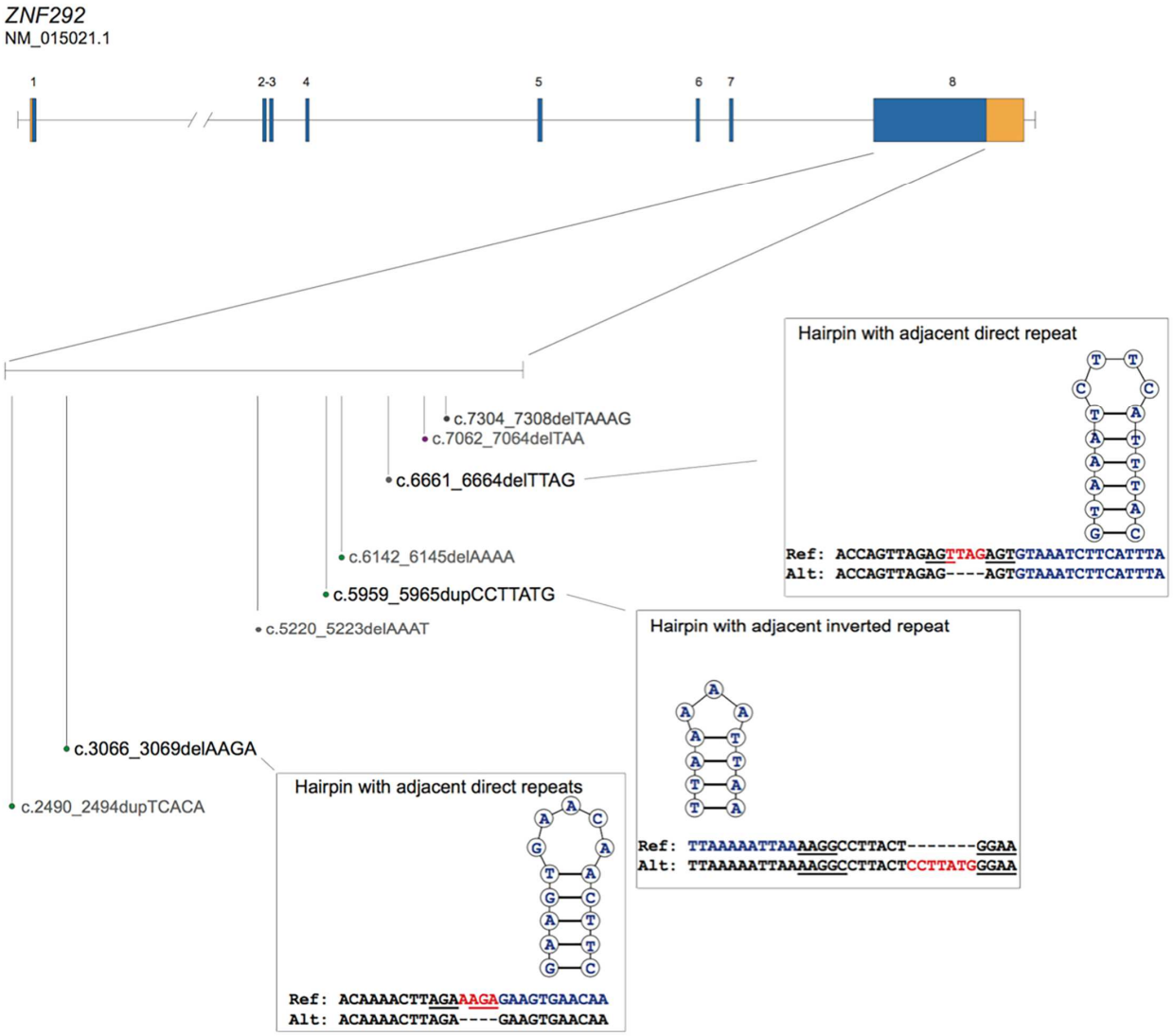


Figure A- 5 Sequence complexity of ZNF292. Gene diagram showing a subset of variants with complex insertion/deletion events in ZNF292. Nearby small palindromic nucleotide sequences are depicted in blue, red nucleotides indicate deleted sequences and underlined nucleotides represent direct or inverted repeats spanning the deletion breakpoints.

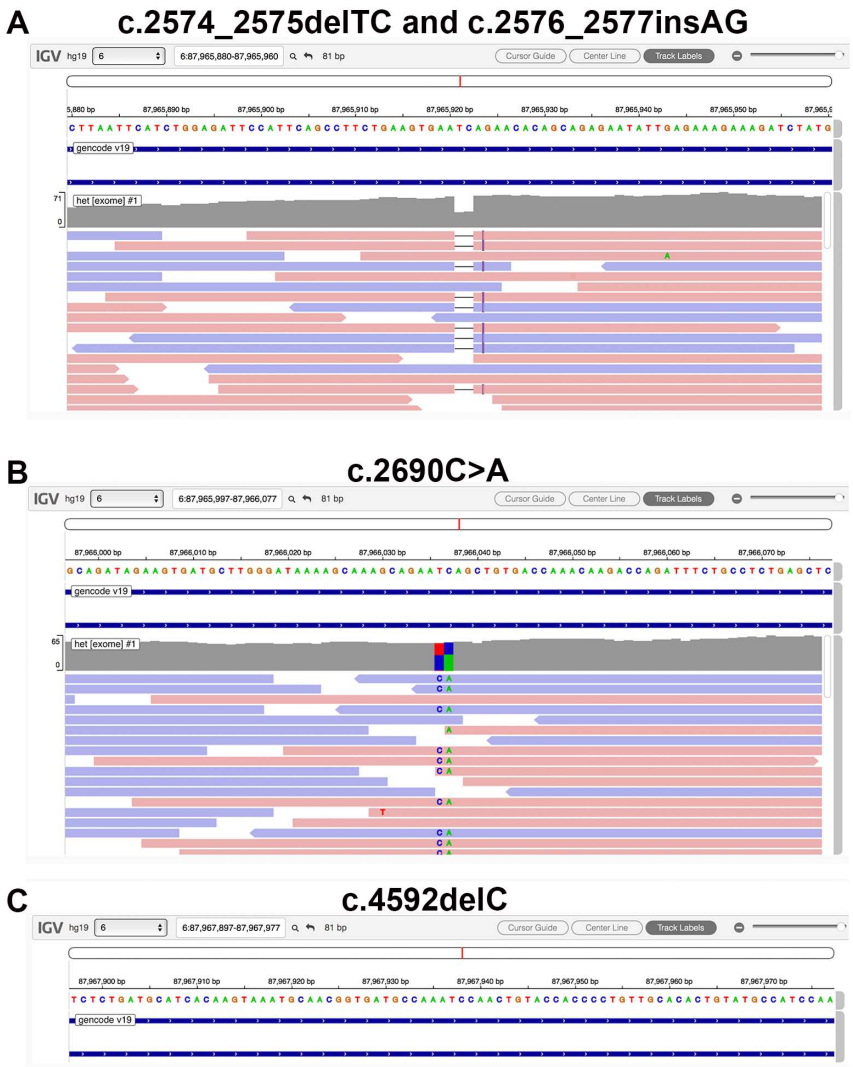


Figure A- 6: Integrative Genome Viewer (IGV) images of select ZNF292 variants in gnomAD. IGV images of 3 select *ZNF292* variants from gnomAD that likely represent mis-calls. **A**, two adjacent frameshifts, c.2574_2575delTC and c.2576_2577insAG, each observed only once and represent a single miscalled variant. **B**, a nonsense variant, c.2690C>A, is a dinucleotide change and should have been annotated as a missense variant. **C**, a c.4592delC variant was listed but not called in any individuals.

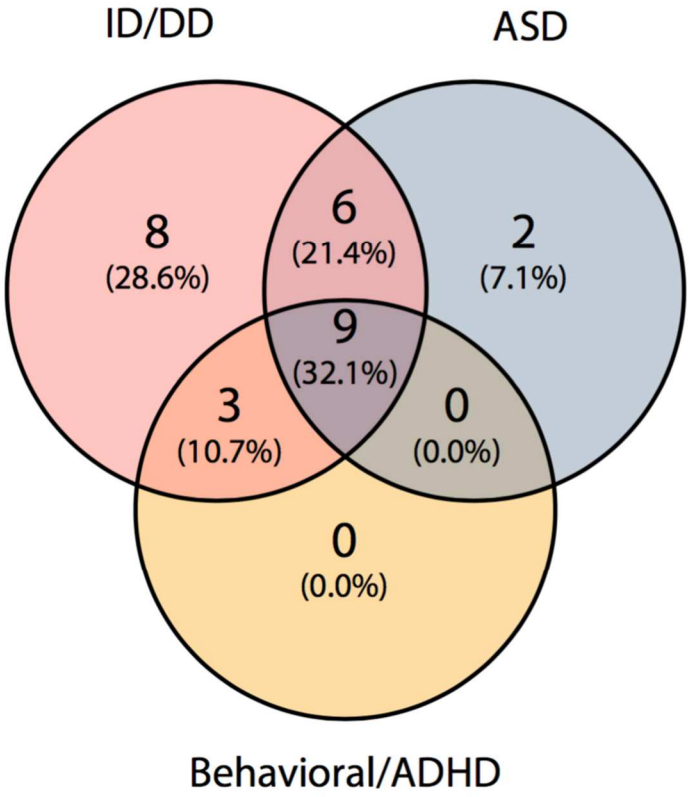


Figure A- 7: Venn diagram display of the distribution of ZNF292 mutation-positive individuals who presented with ID/DD, ASD, and Behavioral issues/ADHD features. Almost all individuals (26 out of 28 on whom clinical data were available) presented with intellectual disability/developmental delays, whereas only two individuals presented with ASD alone without ID. Notably, no individuals had isolated behavioral issues without ID or ASD (likely due to the fact that individuals with behavioral issues alone are less likely to be tested).

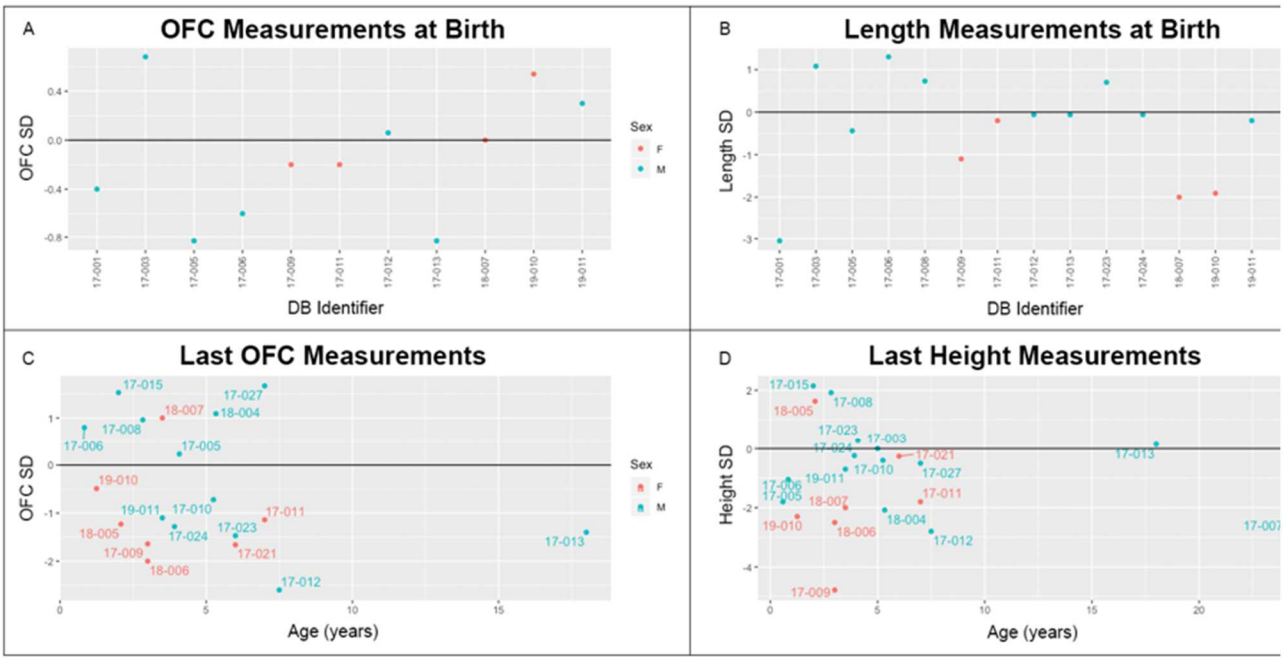


Figure A- 8: Growth (OFC and length) measurements on the *ZNF292* cohort (shown as standard deviations, SD). A, Occipitofrontal circumference (OFC) measurements of individuals with *ZNF292* variants at birth. B, length measurements of individuals with *ZNF292* variants at birth. C, OFC measurements of individuals with *ZNF292* variants at last assessment showing the majority of individuals had OFCs below 0 SD. D, length measurements of individuals with *ZNF292* variants at last assessment showing that the majority of individuals had a stature below 0 SD.

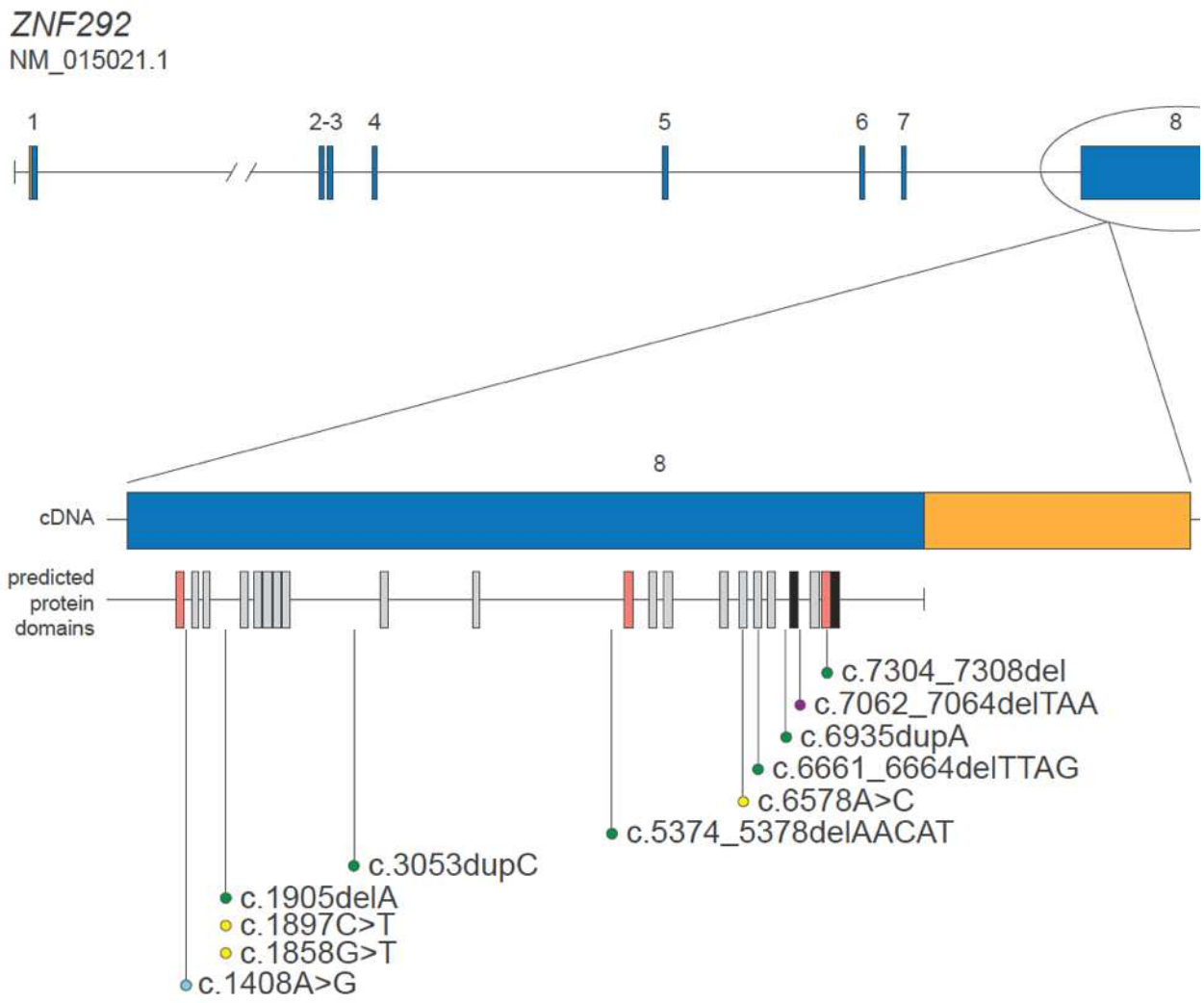


Figure A- 9: 1 Additional ZNF292 variants. *ZNF292* gene diagram showing additional variants identified in the gene of uncertain significance due to limited clinical or molecular data. The individual variants are shown in **Table A- 11**. The cDNA panel shows the coding and non-coding regions of the gene (in blue and yellow, respectively). The bottom panel shows the predicted protein domains including the zinc finger (C2H2 type) regions (shown in gray), the coiled coil domain (in pink) and the nuclear localization signal (in black). ZNF292 variants in the main cohort are shown, color-coded by type with nonsense variants shown in yellow, frameshift variants in green, missense variants in blue and in-frame deletions in purple.

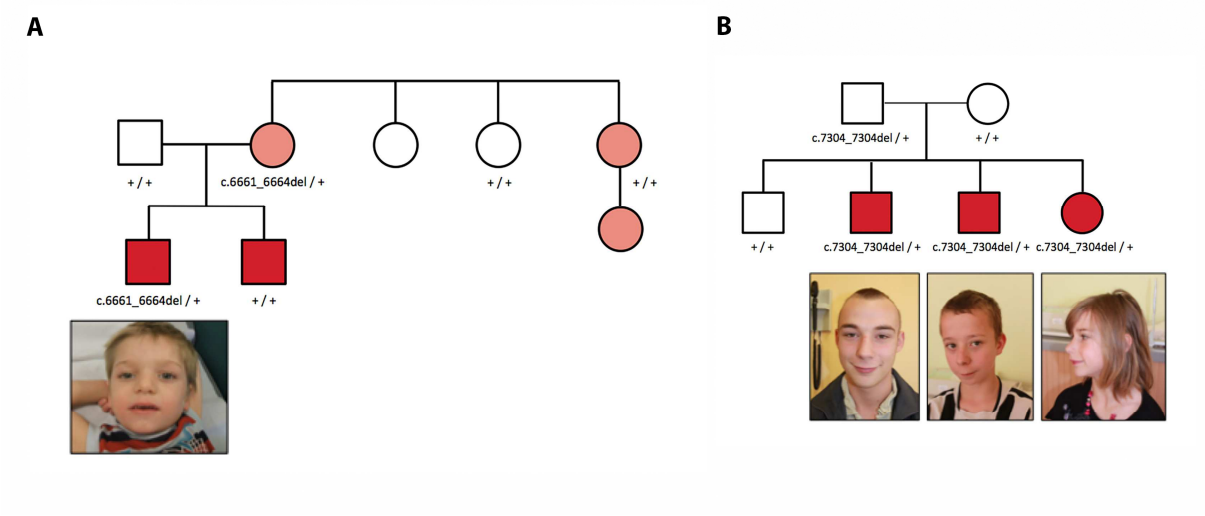


Figure A- 10 Pedigrees of families with additional ZNF292 variants. A, pedigree of family 17-017 showing a photograph of the proband who inherited a *ZNF292* variant (p.Leu2221Serfs*10) from his mother who also has mild intellectual disability. The proband's brother, maternal aunt and first cousin also have "developmental issues" but tested negative for the *ZNF292* variant; with incomplete clinical data on several individuals. B, Pedigree of family 17-020 that consists of three affected children, all of whom harbor a *ZNF292* variant, p. Val2435Glyfs*4. However, this variant was inherited from their putatively unaffected father. The variant was absent in their unaffected sibling. No other candidate variants were identified in either family, but screening was performed via targeted NGS of 456 genes associated with neurodevelopmental disabilities rather than exome sequencing.

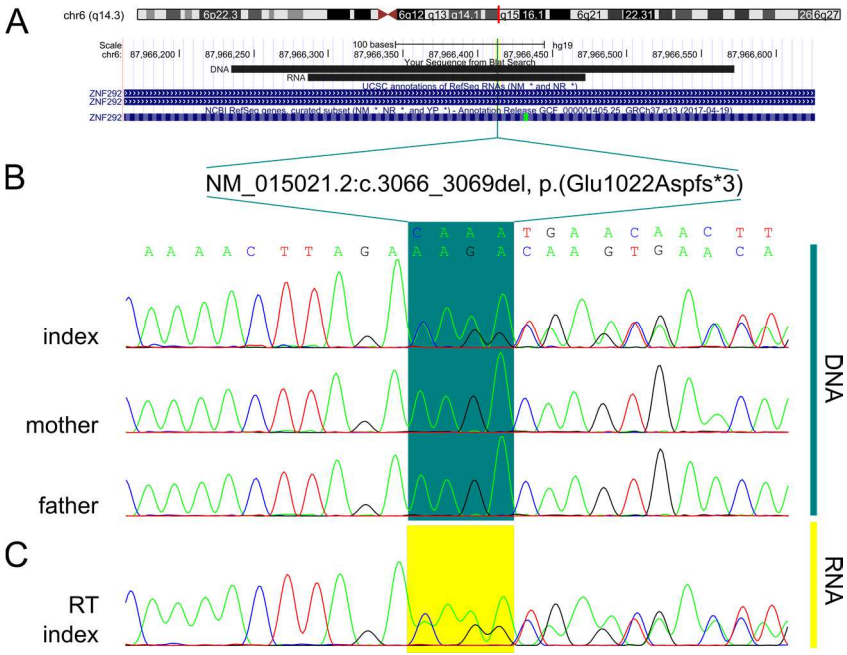


Figure A- 11 RT-PCR analysis on individual 17-005 (p. Glu1022Aspfs*3). (A) UCSC Browser snapshot (hg19: chr6:87,966,164-87,966,626) showing the position of amplicons used for segregation testing ("DNA" amplicon) and RT-PCR analysis ("RNA" amplicon) in individual 17-005 and his parents. The vertical bar indicates the position of the *ZNF292* variant c.3066_3069del, p.(Glu1022Aspfs*3). (B) Segregation testing by Sanger sequencing (primers 5'-TGAAGCACTGGTCACAGACTTAC-3' and 5'-TGGAACAAATGCAATACTTTCTAATG-3') confirms the variant in the index individual and shows that this variant is absent in both parents' DNA. Together with the results of the DNA-based fingerprinting (PowerPlex assay; Promega Corporation, Madison, USA) this confirms that the variant arose *de novo*. (C) Sanger sequencing of RT PCR products (primers: 5′- TTTGTGTCATCCAGGTTTCC-3′ and 5′-TTGTTTTGAAATGCTGCCTG-3′) generated after cDNA synthesis (Superscript II Reverse Transcriptase Kit; Invitrogen, Carlsbad, CA) from total RNA (PAXgene; Becton Dickinson, Franklin Lakes, NJ). Due to the position of the variant we could not use exon spanning primers, but instead performed DNase digestion before cDNA synthesis. RT-PCR analysis showed biallelic expression of the normal and frameshift transcript, indicating that the transcript with the frameshifting variant c.3066_3069del is not degraded by nonsense-mediated mRNA decay.

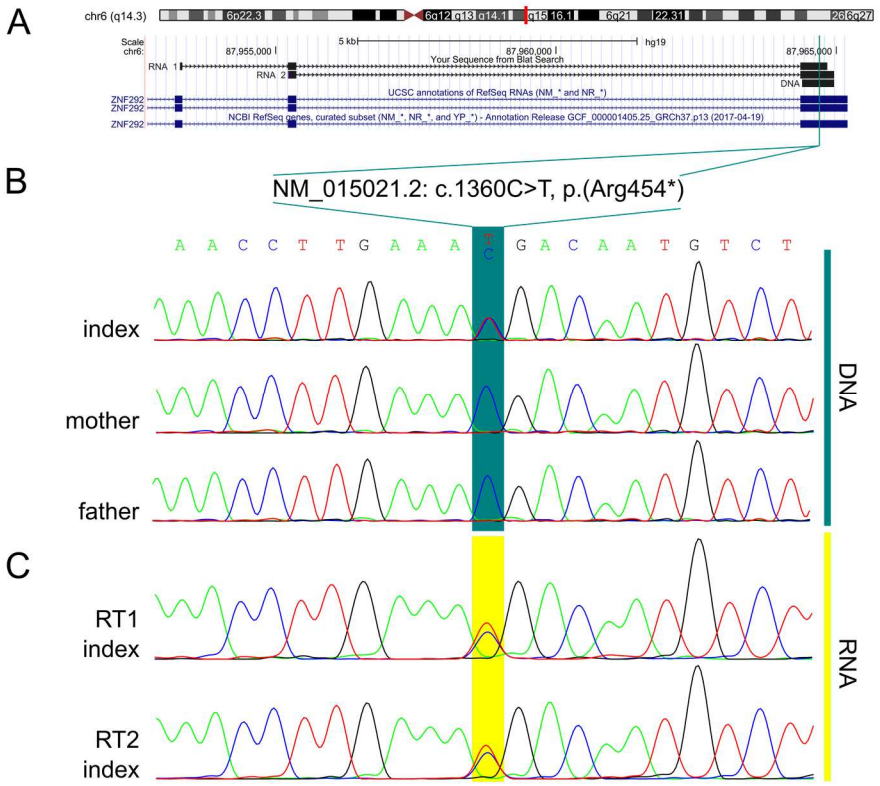


Figure A- 12 RT-PCR analysis on individual 19-005 (P.Arg454)*. (A) UCSC Browser snapshot (hg19: chr6:87,966,164-87,966,626) showing the position of amplicons used for segregation testing ("DNA" amplicon) and RT□PCR analysis ("RNA 1" and "RNA 2" amplicons) in this individual and his parents. The vertical bar indicates the position of the *ZNF292* variant c.1360C>T, p.(Arg454*). (B) Segregation testing by Sanger sequencing (primers 5'-GAACTGTGTGTAAAGGCTCTTCG-3' and 5'-GTAGGCTTGCCAATTCCTAAACC-3') confirms the variant in the index individual and shows that this variant is absent in both parents' DNA. Together with the results of the DNA-based fingerprinting (PowerPlex assay; Promega Corporation, Madison, USA) this confirms that the variant arose *de novo*. (C) Sanger sequencing of RT□PCR products using two different sets of exon spanning primers (RT1: 5′□TGTCACGTCAGCTCCAACAA□3′ and 5′□ACCAACTCCACCAGAAAGCC□3′; RT2: 5′□TGTACTGCGCTTGGGAACTT□3′ and 5′□ TACTGCATGTAGGCTTGCCA□3′) generated after cDNA synthesis (Superscript II Reverse Transcriptase Kit; Invitrogen, Carlsbad, CA) from total RNA (PAXgene; Becton Dickinson, Franklin Lakes, NJ). RT□PCR analysis showed biallelic expression of the normal and termination codon containing transcript, indicating that the transcript with the stop variant c.1360C>T is not degraded by nonsense□mediated mRNA decay.

7.4 Supplemental References

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013a). Predicting functional effect of human missense mutations using PolyPhen□2. *Current protocols in human genetics*, 7.20. 21-27.20. 41.

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013b). Predicting functional effect of human missense mutations using PolyPhen□2. *Current protocols in human genetics, 76*(1), 7.20. 21-27.20. 41.

Adzhubei, I., Schmidt, S., Peshkin, L., Ramensky, V., Gerasimova, A., & Bork, P. (2010). Catalog of Somatic Mutations in Cancer and bioinformatics mutation-prediction tools PolyPhen2. *Nat Methods, 7*, 248-249.

Ahmed, I., Buchert, R., Zhou, M., Jiao, X., Mittal, K., Sheikh, T. I., . . . Brohi, M. Q. (2015). Mutations in DCPS and EDC3 in autosomal recessive intellectual disability indicate a crucial role for mRNA decapping in neurodevelopment. *Human molecular genetics, 24*(11), 3172-3180.

Akbarian, S., Liu, C., Knowles, J. A., Vaccarino, F. M., Farnham, P. J., Crawford, G. E., . . . Geschwind, D. H. (2015). The psychencode project. *Nature neuroscience, 18*(12), 1707.

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature genetics, 23*(2), 185-188.

Anazi, S., Maddirevula, S., Faqeih, E., Alsedairy, H., Alzahrani, F., Shamseldin, H., . . . Abdulwahab, F. (2017). Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. *Molecular Psychiatry, 22*(4), 615-624.

Aoki, Y., Yoncheva, Y. N., Chen, B., Nath, T., Sharp, D., Lazar, M., . . . Di Martino, A. (2017). Association of white matter structure with autism spectrum disorder and attention-deficit/hyperactivity disorder. *JAMA psychiatry, 74*(11), 1120-1128.

Arloth, J., Bader, D. M., Röh, S., & Altmann, A. (2015). Re-Annotator: Annotation pipeline for microarray probe sequences. *PLoS One, 10*(10).

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Eppig, J. T. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics, 25*(1), 25-29.

Autry, A. E., & Monteggia, L. M. (2012). Brain-derived neurotrophic factor and neuropsychiatric disorders. *Pharmacological reviews, 64*(2), 238-258.

Baio, J. (2014). Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010.

Banerjee-Basu, S., & Packer, A. (2010). SFARI Gene: an evolving database for the autism research community. *Disease Models and Mechanisms, 3*(3-4), 133-135.

Barber, J. C., Maloney, V. K., Bewes, B., & Wakeling, E. (2006). Deletions of 2q14 that include the homeobox engrailed 1 (EN1) transcription factor are compatible with a normal phenotype. *European Journal of Human Genetics, 14*(6), 739-743.

Bastian, F., Parmentier, G., Roux, J., Moretti, S., Laudet, V., & Robinson-Rechavi, M. (2008). *Bgee: integrating and comparing heterogeneous transcriptome data among species.* Paper presented at the Data Integration in the Life Sciences.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological), 57*(1), 289-300.

Bernkopf, M., Webersinke, G., Tongsook, C., Koyani, C. N., Rafiq, M. A., Ayaz, M., . . . Naeem, F. (2014). Disruption of the methyltransferase-like 23 gene METTL23 causes mild autosomal recessive intellectual disability. *Human molecular genetics, 23*(15), 4015-4023.

Bittles, A. H., & Black, M. L. (2010). Consanguinity, human evolution, and complex diseases. *Proceedings of the National Academy of Sciences, 107*(suppl 1), 1779-1786.

Blackburn, P. R., Xu, Z., Tumelty, K. E., Zhao, R. W., Monis, W. J., Harris, K. G., . . . Mitkov, M. V. (2018). Bi-allelic alterations in AEBP1 lead to defective collagen assembly and connective tissue structure resulting in a variant of Ehlers-Danlos syndrome. *The American Journal of Human Genetics, 102*(4), 696-705.

Bolger, A. M., Lohse, M., & Usadel, B. J. B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *30*(15), 2114-2120.

Bradford, Y., Conlin, T., Dunn, N., Fashena, D., Frazer, K., Howe, D. G., . . . Moxon, S. A. (2011). ZFIN: enhancements and updates to the Zebrafish Model Organism Database. *Nucleic acids research, 39*(suppl 1), D822-D829.

Bradshaw, N. J., Trossbach, S. V., Köber, S., Walter, S., Prikulis, I., Weggen, S., & Korth, C. (2017). Disrupted in Schizophrenia 1 regulates the processing of reelin in the perinatal cortex. *Schizophrenia research*.

Bragin, E., Chatzimichali, E. A., Wright, C. F., Hurles, M. E., Firth, H. V., Bevan, A. P., & Swaminathan, G. J. (2013). DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic acids research*, gkt937.

Bruellman, R. J., Watanabe, Y., Ebrhim, R. S., Creech, M. K., Abdullah, M. A., Dumitrescu, A. M., . . . Weiss, R. E. (2020). Increased prevalence of TG and TPO mutations in Sudanese children with congenital hypothyroidism. *The Journal of Clinical Endocrinology & Metabolism, 105*(5), dgz297.

Buescher, A. V., Cidav, Z., Knapp, M., & Mandell, D. S. (2014). Costs of autism spectrum disorders in the United Kingdom and the United States. *JAMA pediatrics, 168*(8), 721-728.

Carlson, M. GO. db: A set of annotation maps describing the entire Gene Ontology [Internet]. 2016. In.

Celis, K., Shuldiner, S., Haverfield, E. V., Cappell, J., Yang, R., Gong, D.-W., & Chung, W. K. (2015). Loss of function mutation in glutamic pyruvate transaminase 2 (GPT2) causes developmental encephalopathy. *Journal of inherited metabolic disease, 38*(5), 941-948.

Chahrour, M. H., Timothy, W. Y., Lim, E. T., Ataman, B., Coulter, M. E., Hill, R. S., . . . Gabriel, S. B. (2012). Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS genetics, 8*(4).

Christensen, D. L., Maenner, M. J., Bilder, D., Constantino, J. N., Daniels, J., Durkin, M. S., . . . Robinson, C. (2019). Prevalence and characteristics of autism spectrum disorder among children aged 4 years—Early autism and developmental disabilities monitoring network, seven sites, United States, 2010, 2012, and 2014. *MMWR Surveillance Summaries, 68*(2), 1.

Clarke, T., Lupton, M., Fernandez-Pujals, A., Starr, J., Davies, G., Cox, S., . . . MacIntyre, D. (2015). Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Molecular Psychiatry*.

Coban-Akdemir, Z., White, J. J., Song, X., Jhangiani, S. N., Fatih, J. M., Gambin, T., . . . Punetha, J. J. T. A. J. o. H. G. (2018). Identifying genes whose mutant transcripts cause dominant disease traits by potential gain-of-function alleles. *103*(2), 171-187.

Colantuoni, C., Lipska, B. K., Ye, T., Hyde, T. M., Tao, R., Leek, J. T., . . . Weinberger, D. R. (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature, 478*(7370), 519-523.

Cook Jr, E. H., & Scherer, S. W. (2008). Copy-number variations associated with neuropsychiatric conditions. *Nature, 455*(7215), 919-923.

Craig, F., Margari, F., Legrottaglie, A. R., Palumbi, R., De Giambattista, C., & Margari, L. (2016). A review of executive function deficits in autism spectrum disorder and attention-deficit/hyperactivity disorder. *Neuropsychiatric disease and treatment, 12*, 1191.

Cuchillo-Ibáñez, I., Andreo-Lillo, P., Pastor-Ferrándiz, L., Carratalá-Marco, F., & Sáez-Valero, J. (2020). Elevated Plasma Reelin Levels in Children With Autism. *Frontiers in Psychiatry, 11*, 242.

Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., . . . Quake, S. R. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences, 112*(23), 7285-7290.

Darnell, J. C., Van Driesche, S. J., Zhang, C., Hung, K. Y. S., Mele, A., Fraser, C. E., . . . Chi, S. W. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell, 146*(2), 247-261.

David, M., Dzamba, M., Lister, D., Ilie, L., & Brudno, M. (2011). SHRiMP2: sensitive yet practical short read mapping. *Bioinformatics, 27*(7), 1011-1012.

Davies, G., Marioni, R. E., Liewald, D. C., Hill, W. D., Hagenaars, S. P., Harris, S. E., . . . Lyall, D. (2016). Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N= 112 151). *Molecular Psychiatry*.

Dawson, M., Mottron, L., & Gernsbacher, M. A. (2008). Learning in autism. *Learning and memory: A comprehensive reference, 2*, 759-772.

de Ligt, J., Willemsen, M. H., van Bon, B. W., Kleefstra, T., Yntema, H. G., Kroes, T., . . . Gilissen, C. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *New England Journal of Medicine, 367*(20), 1921-1929.

de Ligt, J., Willemsen, M. H., van Bon, B. W., Kleefstra, T., Yntema, H. G., Kroes, T., . . . Vissers, L. E. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med, 367*(20), 1921-1929. doi:10.1056/NEJMoa1206524

De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., . . . Walker, S. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature, 515*(7526), 209-215.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., . . . Hanna, M. J. N. g. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *43*(5), 491.

Doan, R. N., Lim, E. T., De Rubeis, S., Betancur, C., Cutler, D. J., Chiocchetti, A. G., . . . Freitag, C. M. (2019). Recessive gene disruptions in autism spectrum disorder. *Nature genetics, 51*(7), 1092.

Dobyns, W. B., & Mirzaa, G. M. (2019). *Megalencephaly syndromes associated with mutations of core components of the PI3K□AKT–MTOR pathway: PIK3CA, PIK3R2, AKT3, and MTOR.* Paper presented at the American Journal of Medical Genetics Part C: Seminars in Medical Genetics.

Dong, E., Gavin, D. P., Chen, Y., & Davis, J. (2012). Upregulation of TET1 and downregulation of APOBEC3A and APOBEC3C in the parietal cortex of psychotic patients. *Translational psychiatry, 2*(9), e159.

Dong, E., Ruzicka, W., Grayson, D., & Guidotti, A. (2015). DNA-methyltransferase1 (DNMT1) binding to CpG rich GABAergic and BDNF promoters is increased in the brain of schizophrenia and bipolar disorder patients. *Schizophrenia research, 167*(1), 35-41.

Durand, C. M., Betancur, C., Boeckers, T. M., Bockmann, J., Chaste, P., Fauchereau, F., . . . Anckarsäter, H. (2007). Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature genetics, 39*(1), 25-27.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC bioinformatics, 10*(1), 48.

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), 57-74.

Enns, G. M., Shashi, V., Bainbridge, M., Gambello, M. J., Zahir, F. R., Bast, T., . . . Cox, R. (2014). Mutations in NGLY1 cause an inherited disorder of the endoplasmic reticulum–associated degradation pathway. *Genetics in Medicine, 16*(10), 751-758.

Eppig, J. T., Blake, J. A., Bult, C. J., Kadin, J. A., Richardson, J. E., & Group, M. G. D. (2015). The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic acids research, 43*(D1), D726-D736.

Farwell, K. D., Shahmirzadi, L., El-Khechen, D., Powis, Z., Chao, E. C., Tippin Davis, B., . . . Tang, S. (2015). Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet Med, 17*(7), 578-586. doi:10.1038/gim.2014.154

Fatemi, S., Earle, J., & McMenomy, T. (2000). Reduction in Reelin immunoreactivity in hippocampus of subjects with schizophrenia, bipolar disorder and major depression. *Molecular Psychiatry, 5*(6), 654-663.

Fattahi, Z., Sheikh, T. I., Musante, L., Rasheed, M., Taskiran, I. I., Harripaul, R., . . . Hosseini, M. (2018). Biallelic missense variants in ZBTB11 can cause intellectual disability in humans. *Human molecular genetics, 27*(18), 3177-3188.

Fellmann, C., Gowen, B. G., Lin, P.-C., Doudna, J. A., & Corn, J. E. (2017). Cornerstones of CRISPR–Cas in drug discovery and therapy. *Nature Reviews Drug Discovery, 16*(2), 89-100.

Feng, G., Mellor, R. H., Bernstein, M., Keller-Peck, C., Nguyen, Q. T., Wallace, M., . . . Sanes, J. R. J. N. (2000). Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *28*(1), 41-51.

Fieremans, N., Bauters, M., Belet, S., Verbeeck, J., Jansen, A. C., Seneca, S., . . . Froyen, G. (2014). De novo MECP2 duplications in two females with intellectual disability and unfavorable complete skewed X-inactivation. *Human genetics, 133*(11), 1359-1367.

Figueiredo, T., Melo, U., Pessoa, A., Nobrega, P., Kitajima, J., Rusch, H., . . . Kok, F. (2015). A homozygous loss-of-function mutation in inositol monophosphatase 1 (IMPA1) causes severe intellectual disability. *Molecular Psychiatry*.

Fisher, R. A. (1925). 1954Statistical methods for research workers. *Edinburgh: Oliver and Boyd*.

Folsom, T. D., & Fatemi, S. H. (2013). The involvement of Reelin in neurodevelopmental disorders. *Neuropharmacology, 68*, 122-135.

Fombonne, E. (2005). The changing epidemiology of autism. *Journal of applied research in intellectual disabilities, 18*(4), 281-294.

Franić, S., Groen-Blokhuis, M. M., Dolan, C. V., Kattenberg, M. V., Pool, R., Xiao, X., . . . van der Sluis, S. (2015). Intelligence: shared genetic basis between Mendelian disorders and a polygenic trait. *European Journal of Human Genetics*.

Fromer, M., Moran, J. L., Chambert, K., Banks, E., Bergen, S. E., Ruderfer, D. M., . . . Owen, M. J. (2012). Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *The American Journal of Human Genetics, 91*(4), 597-607.

Fromer, M., Pocklington, A. J., Kavanagh, D. H., Williams, H. J., Dwyer, S., Gormley, P., . . . Ruderfer, D. M. (2014). De novo mutations in schizophrenia implicate synaptic networks. *Nature, 506*(7487), 179-184.

Fromer, M., & Purcell, S. M. (2014). Using XHMM Software to Detect Copy Number Variation in Whole□Exome Sequencing Data. *Current protocols in human genetics*, 7.23. 21-27.23. 21.

Frotscher, M. (1998). Cajal—Retzius cells, Reelin, and the formation of layers. *Current opinion in neurobiology, 8*(5), 570-575.

Gandal, M. J., Zhang, P., Hadjimichael, E., Walker, R. L., Chen, C., Liu, S., . . . Wang, Y. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science, 362*(6420), eaat8127.

Gardner, E. J., Prigmore, E., Gallone, G., Danecek, P., Samocha, K. E., Handsaker, J., . . . Sifrim, A. J. N. c. (2019). Contribution of retrotransposition to developmental disorders. *10*(1), 1-10.

Garshasbi, M., Hadavi, V., Habibi, H., Kahrizi, K., Kariminejad, R., Behjati, F., . . . Kuss, A. W. (2008). A defect in the TUSC3 gene is associated with autosomal recessive mental retardation. *The American Journal of Human Genetics, 82*(5), 1158-1164.

Gazal, S., Sahbatou, M., Babron, M.-C., Génin, E., & Leutenegger, A.-L. (2014). FSuite: exploiting inbreeding in dense SNP chip and exome data. *Bioinformatics, 30*(13), 1940-1941.

Gazal, S., Sahbatou, M., Babron, M. C., Genin, E., & Leutenegger, A. L. (2014). FSuite: exploiting inbreeding in dense SNP chip and exome data. *Bioinformatics, 30*(13), 1940-1941. doi:10.1093/bioinformatics/btu149

Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W., Willemsen, M. H., . . . Schenck, A. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature*.

Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W., Willemsen, M. H., . . . Veltman, J. A. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature, 511*(7509), 344-347. doi:10.1038/nature13394

Glessner, J. T., Wang, K., Cai, G., Korvatska, O., Kim, C. E., Wood, S., . . . Bradfield, J. P. (2009). Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature, 459*(7246), 569-573.

Gogolla, N. (2017). The insular cortex. *Current Biology, 27*(12), R580-R586.

Gordon, A., & Hannon, G. J. F. A. s.-r. p. t. h. h. c. e. f. t. (2010). Fastx-toolkit. *5*.

Gotham, K., Pickles, A., Lord, C. J. J. o. a., & disorders, d. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *39*(5), 693-705.

Griffiths, A. J., Hanson, A. H., Giannantonio, C. M., Mathur, S. K., Hyde, K., & Linstead, E. J. B. S. (2020). Developing Employment Environments Where Individuals with ASD Thrive: Using Machine Learning to Explore Employer Policies and Practices. *10*(9), 632.

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., . . . Anney, R. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics, 51*(3), 431-444.

Guo, H., Wang, T., Wu, H., Long, M., Coe, B. P., Li, H., . . . Duan, G. J. M. a. (2018). Inherited and multiple de novo mutations in autism/developmental delay risk genes suggest a multifactorial model. *9*(1), 64.

Haddad, L. A., Mingroni-Netto, R. C., Vianna-Morgante, A. M., & Pena, S. D. (1996). A PCR-based test suitable for screening for fragile X syndrome among mentally retarded males. *Human genetics, 97*(6), 808-812.

Hall, J., & Owen, M. J. (2015). Psychiatric classification–a developmental perspective. *The British Journal of Psychiatry, 207*(4), 281-282.

Hamamy, H. (2012). Consanguineous marriages. *Journal of community genetics, 3*(3), 185-192.

Hamdan, F. F., Gauthier, J., Araki, Y., Lin, D.-T., Yoshizawa, Y., Higashi, K., . . . Piton, A. (2011). Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *The American Journal of Human Genetics, 88*(3), 306-316.

Hamdan, F. F., Srour, M., Capo-Chichi, J. M., Daoud, H., Nassif, C., Patry, L., . . . Michaud, J. L. (2014). De novo mutations in moderate or severe intellectual disability. *PLoS Genet, 10*(10), e1004772. doi:10.1371/journal.pgen.1004772

Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., & McKusick, V. A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic acids research, 33*(suppl 1), D514-D517.

Harripaul, R., Vasli, N., Mikhailov, A., Rafiq, M. A., Mittal, K., Windpassinger, C., . . . Downey, S. (2016). Mapping Autosomal Recessive Intellectual Disability: Combined Microarray and Exome Sequencing Identifies 26 Novel Candidate Genes in 192 Consanguineous Families. *bioRxiv*, 092346.

Harripaul, R., Vasli, N., Mikhailov, A., Rafiq, M. A., Mittal, K., Windpassinger, C., . . . Downey, S. (2018). Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. *Molecular Psychiatry, 23*(4), 973-984.

Harripaul, R., Vasli, N., Mikhailov, A., Rafiq, M. A., Mittal, K., Windpassinger, C., . . . Vincent, J. B. (2017a). Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. *Molecular Psychiatry*. doi:10.1038/mp.2017.60

https://www.nature.com/articles/mp201760#supplementary-information

Harripaul, R., Vasli, N., Mikhailov, A., Rafiq, M. A., Mittal, K., Windpassinger, C., . . . Vincent, J. B. (2017b). Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. *Mol Psychiatry*. doi:10.1038/mp.2017.60

Heidari, A., Tongsook, C., Najafipour, R., Musante, L., Vasli, N., Garshasbi, M., . . . Sritharan, K. (2015). Mutations in the histamine N-methyltransferase gene, HNMT, are associated with nonsyndromic autosomal recessive intellectual disability. *Human molecular genetics, 24*(20), 5697-5710.

Heslop, P., Blair, P. S., Fleming, P., Hoghton, M., Marriott, A., & Russ, L. (2014). The Confidential Inquiry into premature deaths of people with intellectual disabilities in the UK: a population-based study. *The Lancet, 383*(9920), 889-895.

Howard, D., Negraes, P., Voineskos, A. N., Kaplan, A. S., Muotri, A., Duvvuri, V., & French, L. (2018). Molecular neuroanatomy of anorexia nervosa. *bioRxiv*, 440313.

Howard, H. C., Knoppers, B. M., Cornel, M. C., Clayton, E. W., Sénécal, K., & Borry, P. (2015). Whole-genome sequencing in newborn screening? A statement on the continued importance of targeted approaches in newborn screening programmes. *European Journal of Human Genetics*.

Howrigan, D. P., Simonson, M. A., Davies, G., Harris, S. E., Tenesa, A., Starr, J. M., . . . Wright, M. J. (2016). Genome-wide autozygosity is associated with lower general cognitive ability. *Molecular Psychiatry, 21*(6), 837-843.

Hsu, M.-K., Lin, H.-Y., & Chen, F.-C. J. P. o. (2017). NMD Classifier: A reliable and systematic classification tool for nonsense-mediated decay events. *12*(4).

Hu, B.-Y., Weick, J. P., Yu, J., Ma, L.-X., Zhang, X.-Q., Thomson, J. A., & Zhang, S.-C. J. P. o. t. N. A. o. S. (2010). Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *107*(9), 4335-4340.

Hu, H., Haas, S., Chelly, J., Van Esch, H., Raynaud, M., de Brouwer, A., . . . Laumonnier, F. (2016). X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. *Molecular Psychiatry, 21*(1), 133-148.

Hu, H., Haas, S. A., Chelly, J., Van Esch, H., Raynaud, M., de Brouwer, A. P., . . . Kalscheuer, V. M. (2015). X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. *Mol Psychiatry*. doi:10.1038/mp.2014.193

Hu, H., Kahrizi, K., Musante, L., Fattahi, Z., Herwig, R., Hosseini, M., . . . Larti, F. (2019). Genetics of intellectual disability in consanguineous families. *Molecular Psychiatry, 24*(7), 1027-1039.

Huang, J., Zhu, T., Qu, Y., & Mu, D. (2016). Prenatal, Perinatal and Neonatal Risk Factors for Intellectual Disability: A Systemic Review and Meta-Analysis. *PLoS One, 11*(4), e0153655.

Huang, L., Jolly, L. A., Willis-Owen, S., Gardner, A., Kumar, R., Douglas, E., . . . Cohen, M. (2012). A noncoding, regulatory mutation implicates HCFC1 in nonsyndromic intellectual disability. *The American Journal of Human Genetics, 91*(4), 694-702.

Hussain, R., & Bittles, A. (1998). The prevalence and demographic characteristics of consanguineous marriages in Pakistan. *Journal of biosocial science, 30*(02), 261-275.

Hussein, R., & Bittles, A. (1998). Consanguineous marriage in Pakistan. *Journal of biosocial science, 30*, 261-275.

Iqbal, Z., Neveling, K., Razzaq, A., Shahzad, M., Zahoor, M. Y., Qasim, M., . . . van Bokhoven, H. (2012). Targeted next generation sequencing reveals a novel intragenic deletion of the TPO gene in a family with intellectual disability. *Arch Med Res, 43*(4), 312-316. doi:10.1016/j.arcmed.2012.01.011

Ivanova, E. L., Mau-Them, F. T., Riazuddin, S., Kahrizi, K., Laugel, V., Schaefer, E., . . . Spitz, M.-A. (2017). Homozygous truncating variants in TBC1D23 cause pontocerebellar hypoplasia and alter cortical development. *The American Journal of Human Genetics, 101*(3), 428-440.

Jaffe, A. E., Gao, Y., Deep-Soboslay, A., Tao, R., Hyde, T. M., Weinberger, D. R., & Kleinman, J. E. (2015). Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nature neuroscience*.

Jagadeesh, K. A., Wenger, A. M., Berger, M. J., Guturu, H., Stenson, P. D., Cooper, D. N., . . . Bejerano, G. (2016). *M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity* (1061-4036). Retrieved from

Jaganathan, K., Panagiotopoulou, S. K., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., . . . Schwartz, G. B. (2019). Predicting splicing from primary sequence with deep learning. *Cell, 176*(3), 535-548. e524.

Jamra, R. A., Wohlfart, S., Zweier, M., Uebe, S., Priebe, L., Ekici, A., . . . Fakher, M. (2011). Homozygosity mapping in 64 Syrian consanguineous families with non-specific intellectual disability reveals 11 novel loci and high heterogeneity. *European Journal of Human Genetics, 19*(11), 1161-1166.

Jensen, L. R., Lenzner, S., Moser, B., Freude, K., Tzschach, A., Wei, C., . . . Moraine, C. (2007). X-linked mental retardation: a comprehensive molecular screen of 47 candidate genes from a 7.4 Mb interval in Xp11. *European Journal of Human Genetics, 15*(1), 68-75.

Jian, X., Boerwinkle, E., & Liu, X. (2014). In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic acids research, 42*(22), 13534-13544.

Johansen, A., Rosti, R. O., Musaev, D., Sticca, E., Harripaul, R., Zaki, M., . . . Froukh, T. (2016). Mutations in MBOAT7, Encoding Lysophosphatidylinositol Acyltransferase I, Lead to Intellectual Disability Accompanied by Epilepsy and Autistic Features. *The American Journal of Human Genetics*.

Kaiser, V. B., Svinti, V., Prendergast, J. G., Chau, Y.-Y., Campbell, A., Patarcic, I., . . . Miljkovic, A. (2015). Homozygous loss-of-function variants in European cosmopolitan and isolate populations. *Human molecular genetics, 24*(19), 5464-5474.

Kakar, N., Goebel, I., Daud, S., Nürnberg, G., Agha, N., Ahmad, A., . . . Borck, G. (2012). A homozygous splice site mutation in TRAPPC9 causes intellectual disability and microcephaly. *European journal of medical genetics, 55*(12), 727-731.

Karalunas, S. L., Hawkey, E., Gustafsson, H., Miller, M., Langhorst, M., Cordova, M., . . . Nigg, J. T. (2018). Overlapping and distinct cognitive impairments in attention-deficit/hyperactivity and autism spectrum disorder without intellectual disability. *Journal of abnormal child psychology, 46*(8), 1705-1716.

Karczewski, K., & Francioli, L. (2017). The Genome Aggregation Database (gnomAD). *MacArthur Lab*.

Kaufman, L., Ayub, M., & Vincent, J. B. (2010). The genetic basis of non-syndromic intellectual disability: a review. *Journal of neurodevelopmental disorders, 2*(4), 182-209.

Kerin, T., Ramanathan, A., Rivas, K., Grepo, N., Coetzee, G. A., & Campbell, D. B. (2012). A noncoding RNA antisense to moesin at 5p14. 1 in autism. *Science translational medicine, 4*(128), 128ra140-128ra140.

Kevelam, S. H., van Harssel, J. J., van der Zwaag, B., Smeets, H. J., Paulussen, A. D., & Lichtenbelt, K. D. (2012). A patient with a mild holoprosencephaly spectrum phenotype and heterotaxy and a 1.3 Mb deletion encompassing GLI2. *American Journal of Medical Genetics Part A, 158*(1), 166-173.

Khan, M. A., Rafiq, M. A., Noor, A., Ali, N., Ali, G., Vincent, J. B., & Ansar, M. (2011). A novel deletion mutation in the TUSC3 gene in a consanguineous Pakistani family with autosomal recessive nonsyndromic intellectual disability. *BMC medical genetics, 12*(1), 56.

Khan, M. A., Rafiq, M. A., Noor, A., Hussain, S., Flores, J. V., Rupp, V., . . . Khan, F. S. (2012). Mutation in NSUN2, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability. *The American Journal of Human Genetics, 90*(5), 856-863.

Knoppers, B. M., Avard, D., Sénécal, K., & Zawati, M. n. H. (2014). Return of whole-genome sequencing results in paediatric research: a statement of the P3G international paediatrics platform. *Eur J Hum Genet, 22*(1), 3-5.

Kosugi, S., Hasebe, M., Tomita, M., & Yanagawa, H. J. P. o. t. N. A. o. S. (2009). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *106*(25), 10171-10176.

Krawitz, P. M., Murakami, Y., Rieß, A., Hietala, M., Krüger, U., Zhu, N., . . . Robinson, P. N. (2013). PGAP2 mutations, affecting the GPI-anchor-synthesis pathway, cause hyperphosphatasia with mental retardation syndrome. *The American Journal of Human Genetics, 92*(4), 584-589.

Krumm, N., Sudmant, P. H., Ko, A., O'Roak, B. J., Malig, M., Coe, B. P., . . . research, N. E. S. P. J. G. (2012). Copy number variation detection and genotyping from exome sequence data. *22*(8), 1525-1532.

Krumm, N., Turner, T. N., Baker, C., Vives, L., Mohajeri, K., Witherspoon, K., . . . He, Z.-X. (2015). Excess of rare, inherited truncating mutations in autism. *Nature genetics, 47*(6), 582-588.

Kutsche, K., Yntema, H., Brandt, A., Jantke, I., Nothwang, H. G., Orth, U., . . . Fryns, J.-P. (2000). Mutations in ARHGEF6, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation. *Nature genetics, 26*(2), 247-250.

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic acids research, 42*(D1), D980-D985.

Langen, M., Bos, D., Noordermeer, S. D., Nederveen, H., van Engeland, H., & Durston, S. J. B. p. (2014). Changes in the development of striatum are involved in repetitive behavior in autism. *76*(5), 405-411.

Law, R., Dixon-Salazar, T., Jerber, J., Cai, N., Abbasi, A. A., Zaki, M. S., . . . Khan, V. (2014). Biallelic truncating mutations in FMN2, encoding the actin-regulatory protein formin 2, cause nonsyndromic autosomal-recessive intellectual disability. *The American Journal of Human Genetics, 95*(6), 721-728.

Lee, H., Deignan, J. L., Dorrani, N., Strom, S. P., Kantarci, S., Quintero-Rivera, F., . . . Yourshaw, M. (2014). Clinical exome sequencing for genetic identification of rare Mendelian disorders. *Jama, 312*(18), 1880-1887.

Lee, J. H., Huynh, M., Silhavy, J. L., Kim, S., Dixon-Salazar, T., Heiberg, A., . . . Collazo, A. (2012). De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nature genetics, 44*(8), 941.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., . . . Cummings, B. B. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature, 536*(7616), 285-291.

Levy, D., Ronemus, M., Yamrom, B., Lee, Y.-h., Leotta, A., Kendall, J., . . . Ye, K. (2011). Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron, 70*(5), 886-897.

Li, C., Lan, Y., Schwartz-Orbach, L., Korol, E., Tahiliani, M., Evans, T., & Goll, M. G. (2015). Overlapping Requirements for Tet2 and Tet3 in Normal Development and Hematopoietic Stem Cell Emergence. *Cell reports, 12*(7), 1133-1143.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.

Li, J., Cai, T., Jiang, Y., Chen, H., He, X., Chen, C., . . . Li, Z. (2016). Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. *Molecular Psychiatry, 21*(12), 290-297.

Lim, E. T., Raychaudhuri, S., Sanders, S. J., Stevens, C., Sabo, A., MacArthur, D. G., . . . Fromer, M. (2013). Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron, 77*(2), 235-242.

Lim, K. K., & Chong, W. H. (2017). Moderating effect of child's autism spectrum disorder (ASD) diagnosis on benefit finding and negative affect of parents. *American Journal of Orthopsychiatry, 87*(3), 357.

Lord, C., DiLavore, P. C., & Gotham, K. (2012). *Autism diagnostic observation schedule*: Western Psychological Services Torrance, CA.

Lord, C., Rutter, M., Dilavore, P. C., & Risi, S. (2008). *ADOS: Autism diagnostic observation schedule*: Hogrefe Boston, MA.

Lord, C., Storoschuk, S., Rutter, M., & Pickles, A. J. I. M. H. J. (1993). Using the ADI□R to diagnose autism in preschool children. *14*(3), 234-252.

MacDonald, J. R., Ziman, R., Yuen, R. K., Feuk, L., & Scherer, S. W. (2013). The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic acids research*, gkt958.

Madsen, G. F., Bilenberg, N., Cantio, C., & Oranje, B. (2014). Increased prepulse inhibition and sensitization of the startle reflex in autistic children. *Autism Research, 7*(1), 94-103.

Maenner, M. J., Shaw, K. A., & Baio, J. (2020). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. *MMWR Surveillance Summaries, 69*(4), 1.

Maki, T., Liang, A. C., Miyamoto, N., Lo, E. H., & Arai, K. (2013). Mechanisms of oligodendrocyte regeneration from ventricular-subventricular zone-derived progenitor cells in white matter diseases. *Frontiers in cellular neuroscience, 7*, 275.

Mancarci, B. O., Toker, L., Tripathy, S. J., Li, B., Rocco, B., Sibille, E., & Pavlidis, P. (2017). Cross-laboratory analysis of brain cell type transcriptomes with applications to interpretation of bulk tissue data. *Eneuro, 4*(6).

Marangi, G., Leuzzi, V., Manti, F., Lattante, S., Orteschi, D., Pecile, V., . . . Zollino, M. (2013). TRAPPC9-related autosomal recessive intellectual disability: report of a new mutation and clinical phenotype. *European Journal of Human Genetics, 21*(2), 229-232.

Marchuk, D. S., Crooks, K., Strande, N., Kaiser-Rogers, K., Milko, L. V., Brandt, A., . . . Vora, N. L. J. P. o. (2018). Increasing the diagnostic yield of exome sequencing by copy number variant analysis. *13*(12), e0209185.

Marin-Valencia, I., Gerondopoulos, A., Zaki, M. S., Ben-Omran, T., Almureikhi, M., Demir, E., . . . Appelhof, B. (2017). Homozygous mutations in TBC1D23 lead to a non-degenerative form of pontocerebellar hypoplasia. *The American Journal of Human Genetics, 101*(3), 441-450.

Marin-Valencia, I., Novarino, G., Johansen, A., Rosti, B., Issa, M. Y., Musaev, D., . . . Stanley, V. (2018). A homozygous founder mutation in TRAPPC6B associates with a neurodevelopmental disorder characterised by microcephaly, epilepsy and autistic features. *Journal of medical genetics, 55*(1), 48-54.

Marshall, C. R., Noor, A., Vincent, J. B., Lionel, A. C., Feuk, L., Skaug, J., . . . Ren, Y. (2008). Structural variation of chromosomes in autism spectrum disorder. *The American Journal of Human Genetics, 82*(2), 477-488.

Martin, C. L., Wain, K. E., Oetjens, M. T., Tolwinski, K., Palen, E., Hare-Harris, A., . . . Walsh, L. K. (2020). Identification of Neuropsychiatric Copy Number Variants in a Health Care System Population. *JAMA psychiatry*.

Martin, H. C., Gardner, E. J., Samocha, K. E., Kaplanis, J., Akawi, N., Sifrim, A., . . . Niemi, M. E. J. m. (2020). The contribution of X-linked coding variation to severe developmental disorders.

Martin, H. C., Jones, W. D., McIntyre, R., Sanchez-Andrade, G., Sanderson, M., Stephenson, J. D., . . . Bruntraeger, M. (2018). Quantifying the contribution of recessive coding variation to developmental disorders. *Science, 362*(6419), 1161-1164.

Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T., & Saxena, S. (2011). Prevalence of intellectual disability: a meta-analysis of population-based studies. *Research in developmental disabilities, 32*(2), 419-436.

McCarthy, S. E., Gillis, J., Kramer, M., Lihm, J., Yoon, S., Berstein, Y., . . . Ghiban, E. (2014). De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Molecular psychiatry, 19*(6), 652.

McLaren, J., & Bryson, S. E. (1987). Review of recent epidemiological studies of mental retardation: prevalence, associated disorders, and etiology. *American Journal on Mental Retardation*.

Mefford, H. C., Batshaw, M. L., & Hoffman, E. P. (2012). Genomics, intellectual disability, and autism. *New England Journal of Medicine, 366*(8), 733-743.

Meloni, I., Bruttini, M., Longo, I., Mari, F., Rizzolio, F., D'Adamo, P., . . . Renieri, A. (2000). A mutation in the Rett syndrome gene, MECP2, causes X-linked mental retardation and progressive spasticity in males. *The American Journal of Human Genetics, 67*(4), 982-985.

Merritt, J. K., Collins, B. E., Erickson, K. R., Dong, H., & Neul, J. L. J. H. M. G. (2020). Pharmacological readthrough of R294X Mecp2 in a novel mouse model of Rett Syndrome.

Metsu, S., Rooms, L., Rainger, J., Taylor, M. S., Bengani, H., Wilson, D. I., . . . Reyniers, E. (2014). FRA2A is a CGG repeat expansion associated with silencing of AFF3. *PLoS Genet, 10*(4), e1004242.

Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., . . . Epstein, C. J. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *The American Journal of Human Genetics, 86*(5), 749-764.

Miller, J. A., Ding, S.-L., Sunkin, S. M., Smith, K. A., Ng, L., Szafer, A., . . . Aiona, K. (2014). Transcriptional landscape of the prenatal human brain. *Nature, 508*(7495), 199-206.

Mir, A., Kaufman, L., Noor, A., Motazacker, M. M., Jamil, T., Azam, M., . . . Nasr, T. (2009). Identification of mutations in TRAPPC9, which encodes the NIK-and IKK-β-binding protein, in nonsyndromic autosomal-recessive mental retardation. *The American Journal of Human Genetics, 85*(6), 909-915.

Mir, A., Sritharan, K., Mittal, K., Vasli, N., Araujo, C., Jamil, T., . . . Rauf, S. (2014). Truncation of the E3 ubiquitin ligase component FBXO31 causes non-syndromic autosomal recessive intellectual disability in a Pakistani family. *Human genetics, 133*(8), 975-984.

Mirzaa, G. M., Chong, J. X., Piton, A., Popp, B., Foss, K., Guo, H., . . . Aldinger, K. A. (2020). De novo and inherited variants in ZNF292 underlie a neurodevelopmental disorder with features of autism spectrum disorder. *Genetics in Medicine, 22*(3), 538-546.

Mittal, K., Rafiq, M. A., Rafiullah, R., Harripaul, R., Ali, H., Ayaz, M., . . . So, J. (2016). Mutations in the genes for thyroglobulin and thyroid peroxidase cause thyroid dyshormonogenesis and autosomal-recessive intellectual disability. *Journal of Human Genetics*,. 61(10), 867-872.

Mochida, G. H., Mahajnah, M., Hill, A. D., Basel-Vanagaite, L., Gleason, D., Hill, R. S., . . . Walsh, C. A. (2009). A truncating mutation of TRAPPC9 is associated with autosomal-recessive intellectual disability and postnatal microcephaly. *The American Journal of Human Genetics, 85*(6), 897-902.

Modell, B., & Darr, A. (2002). Genetic counselling and customary consanguineous marriage. *Nature Reviews Genetics, 3*(3), 225-229.

Morrow, E. M., Yoo, S.-Y., Flavell, S. W., Kim, T.-K., Lin, Y., Hill, R. S., . . . Hashmi, A. (2008). Identifying autism loci and genes by tracing recent shared ancestry. *Science, 321*(5886), 218-223.

Musante, L., & Ropers, H. H. (2014). Genetics of recessive cognitive disorders. *Trends in Genetics, 30*(1), 32-39.

Myakhar, O., Unichenko, P., & Kirischuk, S. (2011). GABAergic projections from the subplate to Cajal–Retzius cells in the neocortex. *Neuroreport, 22*(11), 525-529.

Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S. S., Chen, W., . . . Jamali, P. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature, 478*(7367), 57-63.

Najmabadi, H., Motazacker, M. M., Garshasbi, M., Kahrizi, K., Tzschach, A., Chen, W., . . . Abedini, S. S. (2007). Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci. *Human genetics, 121*(1), 43-48.

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research, 31*(13), 3812-3814.

Niemi, M. E., Martin, H. C., Rice, D. L., Gallone, G., Gordon, S., Kelemen, M., . . . Yu, S. J. N. (2018). Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *562*(7726), 268-271.

Noor, A., Whibley, A., Marshall, C. R., Gianakopoulos, P. J., Piton, A., Carson, A. R., . . . Pinto, D. (2010). Disruption at the PTCHD1 Locus on Xp22. 11 in Autism spectrum disorder and intellectual disability. *Science translational medicine, 2*(49), 49ra68-49ra68.

Noor, A., Windpassinger, C., Patel, M., Stachowiak, B., Mikhailov, A., Azam, M., . . . Paterson, A. D. (2008). CC2D2A, encoding a coiled-coil and C2 domain protein, causes autosomal-recessive mental retardation with retinitis pigmentosa. *The American Journal of Human Genetics, 82*(4), 1011-1018.

O'Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., . . . Smith, J. D. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature, 485*(7397), 246-250.

Ofner, M., Coles, A., Decou, M., Do, M., Bienek, A., Snider, J., & Ugnat, A. (2018). *Autism spectrum disorder among children and youth in Canada 2018*: Public Health Agency of Canada Ottawa.

Omoyinmi, E., Melo Gomes, S., Standing, A., Rowczenio, D. M., Eleftheriou, D., Klein, N., . . . Brogan, P. A. (2014). Brief Report: Whole□Exome Sequencing Revealing Somatic NLRP3 Mosaicism in a Patient With Chronic Infantile Neurologic, Cutaneous, Articular Syndrome. *Arthritis & Rheumatology, 66*(1), 197-202.

Ouyang, Q., Kavanaugh, B. C., Joesch-Cohen, L., Dubois, B., Wu, Q., Schmidt, M., . . . Mishra, S. (2019). GPT2 mutations in autosomal recessive developmental disability: extending the clinical phenotype and population prevalence estimates. *Human genetics, 138*(10), 1183-1200.

Packer, J. S., Maxwell, E. K., O'Dushlaine, C., Lopez, A. E., Dewey, F. E., Chernomorsky, R., . . . Reid, J. G. (2016). CLAMMS: a scalable algorithm for calling common and rare copy number variants from exome sequencing data. *Bioinformatics, 32*(1), 133-135.

Page, M. M., Bell, D. A., & Watts, G. F. J. C. G. (2020). Widening the spectrum of genetic testing in familial hypercholesterolaemia: Will it translate into better patient and population outcomes? *, 97*(4), 543-555.

Pagnamenta, A. T., Bacchelli, E., de Jonge, M. V., Mirza, G., Scerri, T. S., Minopoli, F., . . . Paracchini, S. (2010). Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia. *Biological psychiatry, 68*(4), 320-328.

Pang, A. W. C., MacDonald, J. R., Yuen, R. K., Hayes, V. M., & Scherer, S. W. (2014). Performance of high-throughput sequencing for the discovery of genetic variation across the complete size spectrum. *G3: Genes| Genomes| Genetics, 4*(1), 63-65.

Patel, S., Howard, D., Man, A., Schwartz, D., Jee, J., Felsky, D., . . . French, L. (2019). Transcriptomic analysis of brain hypometabolism links the cholinergic and gingipain hypotheses of Alzheimer's disease pathogenesis. *bioRxiv*.

Patwardhan, A., Harris, J., Leng, N., Bartha, G., Church, D. M., Luo, S., . . . Salit, M. (2015). Achieving high-sensitivity for clinical applications using augmented exome sequencing. *Genome medicine, 7*(1), 1.

Peters, C., Nicholas, A. K., Schoenmakers, E., Lyons, G., Langham, S., Serra, E. G., . . . Schoenmakers, N. (2019). DUOX2/DUOXA2 Mutations Frequently Cause Congenital Hypothyroidism that Evades Detection on Newborn Screening in the United Kingdom. *Thyroid, 29*(6), 790-801.

Philippe, O., Rio, M., Carioux, A., Plaza, J.-M., Guigue, P., Molinari, F., . . . Smahi, A. (2009). Combination of linkage mapping and microarray-expression analysis identifies NF-κB signaling defect as a cause of autosomal-recessive mental retardation. *The American Journal of Human Genetics, 85*(6), 903-908.

Philips, A. K., Pinelli, M., de Bie, C. I., Mustonen, A., Määttä, T., Arts, H. H., . . . Raza, S. (2016). Identification of C12orf4 as a gene for autosomal recessive intellectual disability. *Clinical genetics,* 91(1), 100-105.

Popp, B., Ekici, A. B., Thiel, C. T., Hoyer, J., Wiesener, A., Kraus, C., . . . Zweier, C. (2017). Exome Pool-Seq in neurodevelopmental disorders. *Eur J Hum Genet, 25*(12), 1364-1376. doi:10.1038/s41431-017-0022-1

Pott, A., Rottbauer, W., & Just, S. J. E. O. o. D. D. (2020). Streamlining drug discovery assays for cardiovascular disease using zebrafish. *15*(1), 27-37.

Poultney, C. S., Goldberg, A. P., Drapeau, E., Kou, Y., Harony-Nicolas, H., Kajiwara, Y., . . . Rehnström, K. (2013). Identification of small exonic CNV from whole-exome sequence data and application to autism spectrum disorder. *The American Journal of Human Genetics, 93*(4), 607-619.

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research, 35*(suppl 1), D61-D65.

Puffenberger, E. G., Jinks, R. N., Sougnez, C., Cibulskis, K., Willert, R. A., Achilly, N. P., . . . Lawrence, J. J. (2012). Genetic mapping and exome sequencing identify variants associated with five novel diseases. *PLoS One, 7*(1), e28936.

Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., . . . Kähler, A. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature, 506*(7487), 185-190.

Qiao, Y., Mondal, K., Trapani, V., Wen, J., Carpenter, G., Wildin, R., . . . Jiang, R. J. H. m. (2014). Variant ATRX Syndrome with Dysfunction of ATRX and MAGT 1 Genes. *35*(1), 58-62.

Radnikow, G., Feldmeyer, D., & Lübke, J. (2002). Axonal projection, input and output synapses, and synaptic physiology of Cajal–Retzius cells in the developing rat neocortex. *Journal of Neuroscience, 22*(16), 6908-6919.

Rafiq, M. A., Kuss, A. W., Puettmann, L., Noor, A., Ramiah, A., Ali, G., . . . Garshasbi, M. (2011). Mutations in the alpha 1, 2-mannosidase gene, MAN1B1, cause autosomal-recessive intellectual disability. *The American Journal of Human Genetics, 89*(1), 176-182.

Rafiq, M. A., Leblond, C. S., Saqib, M. A., Vincent, A. K., Ambalavanan, A., Khan, F. S., . . . Ali, G. (2015). Novel VPS13B Mutations in Three Large Pakistani Cohen Syndrome Families Suggests a Baloch Variant with Autistic-Like Features. *BMC medical genetics, 16*(1), 41.

Rajab, A., Schuelke, M., Gill, E., Zwirner, A., Seifert, F., Gonzalez, S. M., & Knierim, E. (2015). Recessive DEAF1 mutation associates with autism, intellectual disability, basal ganglia dysfunction and epilepsy. *Journal of medical genetics, 52*(9), 607-611.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. J. N. p. (2013). Genome engineering using the CRISPR-Cas9 system. *8*(11), 2281-2308.

Rauch, A., Hoyer, J., Guth, S., Zweier, C., Kraus, C., Becker, C., . . . Rüschendorf, F. (2006). Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *American Journal of Medical Genetics Part A, 140*(19), 2063-2074.

Rehm, H. L., Bale, S. J., Bayrak-Toydemir, P., Berg, J. S., Brown, K. K., Deignan, J. L., . . . Lyon, E. (2013). ACMG clinical laboratory standards for next-generation sequencing. *Genetics in Medicine, 15*(9), 733-747.

Reiff, R. E., Ali, B. R., Baron, B., Timothy, W. Y., Ben-Salem, S., Coulter, M. E., . . . Al-Younes, B. (2014). METTL23, a transcriptional partner of GABPA, is essential for human cognition. *Human molecular genetics*, ddu054.

Reig, R., & Silberberg, G. J. N. (2014). Multisensory integration in the mouse striatum. *83*(5), 1200-1212.

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research, 47*(D1), D886-D894.

Retterer, K., Juusola, J., Cho, M. T., Vitazka, P., Millan, F., Gibellini, F., . . . Bale, S. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genet Med, 18*(7), 696-704. doi:10.1038/gim.2015.148

Reuter, M. S., Tawamie, H., Buchert, R., Gebril, O. H., Froukh, T., Thiel, C., . . . Zweier, C. (2017). Diagnostic yield and novel candidate genes by exome sequencing in 152 consanguineous families with neurodevelopmental disorders. *JAMA psychiatry, 74*(3), 293-299.

Reva, B., Antipin, Y., & Sander, C. (2010). Mutation Assessor. *Cancer*.

Riazuddin, S., Hussain, M., Razzaq, A., Iqbal, Z., Shahzad, M., Polla, D., . . . Tomas-Roca, L. (2016). Exome sequencing of Pakistani consanguineous families identifies 30 novel candidate genes for recessive intellectual disability. *Molecular Psychiatry*.

Riazuddin, S., Hussain, M., Razzaq, A., Iqbal, Z., Shahzad, M., Polla, D., . . . Tomas-Roca, L. (2017). Exome sequencing of Pakistani consanguineous families identifies 30 novel candidate genes for recessive intellectual disability. *Molecular Psychiatry, 22*(11), 1604-1614.

Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S. J., Zhong, Q., Sahni, N., . . . Mosca, R. (2014). A proteome-scale map of the human interactome network. *Cell, 159*(5), 1212-1226.

Ropers, H.-H., & Hamel, B. C. (2005). X-linked mental retardation. *Nature Reviews Genetics, 6*(1), 46-57.

Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., . . . Ayivi-Guedehoussou, N. (2005). Towards a proteome-scale map of the human protein–protein interaction network. *Nature, 437*(7062), 1173-1178.

Saadat, M. (2007). Consanguinity associated with child and adult mortality in 24 Asian and African countries, an ecological study. *Iranian Journal of Public Health, 36*, 35-39.

Saadat, M., Ansari-Lari, M., & Farhud, D. J. A. o. h. b. (2004). Short report consanguineous marriage in Iran. *31*(2), 263-269.

Sadedin, S. P., Ellis, J. A., Masters, S. L., & Oshlack, A. (2018). Ximmer: a system for improving accuracy and consistency of CNV calling from exome data. *GigaScience, 7*(10), giy112.

Safran, M., Dalah, I., Alexander, J., Rosen, N., Stein, T. I., Shmoish, M., . . . Krug, H. (2010). GeneCards Version 3: the human gene integrator. *Database, 2010*, baq020.

Samarakoon, P. S., Sorte, H. S., Stray-Pedersen, A., Rødningen, O. K., Rognes, T., & Lyle, R. (2016). cnvScan: a CNV screening and annotation tool to improve the clinical utility of computational CNV prediction from exome sequencing data. *BMC genomics, 17*(1), 51.

Samuels, Y., Wang, Z., Bardelli, A., Silliman, N., Ptak, J., Szabo, S., . . . Riggins, G. J. J. S. (2004). High frequency of mutations of the PIK3CA gene in human cancers.

Sang, L., Miller, J. J., Corbit, K. C., Giles, R. H., Brauer, M. J., Otto, E. A., . . . Kwong, M. (2011). Mapping the NPHP-JBTS-MKS protein network reveals ciliopathy disease genes and pathways. *Cell, 145*(4), 513-528.

Satterstrom, F. K., Walters, R. K., Singh, T., Wigdor, E. M., Lescai, F., Demontis, D., . . . Bybjerg-Grauholm, J. J. N. n. (2019). Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. *22*(12), 1961-1965.

Scala, M., Bianchi, A., Bisulli, F., Coppola, A., Elia, M., Trivisano, M., . . . Lattanzi, S. J. E. R. o. N. (2020). Advances in genetic testing and optimization of clinical management in children and adults with epilepsy. *20*(3), 251-269.

Schluth-Bolard, C., Labalme, A., Cordier, M. P., Till, M., Nadeau, G., Tevissen, H., . . . Sanlaville, D. (2013). Breakpoint mapping by next generation sequencing reveals causative gene disruption in patients carrying apparently balanced chromosome rearrangements with intellectual deficiency and/or congenital malformations. *J Med Genet, 50*(3), 144-150. doi:10.1136/jmedgenet-2012-101351

Schmeisser, M. J., Ey, E., Wegener, S., Bockmann, J., Stempel, A. V., Kuebler, A., . . . Spilker, C. (2012). Autistic-like behaviours and hyperactivity in mice lacking ProSAP1/Shank2. *Nature, 486*(7402), 256-260.

Schwarz, J. M., Rödelsperger, C., Schuelke, M., & Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods, 7*(8), 575-576.

Scott, E. M., Halees, A., Itan, Y., Spencer, E. G., He, Y., Azab, M. A., . . . Abel, L. (2016). Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nature genetics,* 48(9), 1071-1076.

Seelow, D., Schuelke, M., Hildebrandt, F., & Nürnberg, P. (2009). HomozygosityMapper—an interactive approach to homozygosity mapping. *Nucleic acids research, 37*(suppl 2), W593-W599.

Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L., Edwards, K. J., . . . Gaunt, T. R. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human mutation, 34*(1), 57-65.

Short, P. J., McRae, J. F., Gallone, G., Sifrim, A., Won, H., Geschwind, D. H., . . . Barrett, J. C. J. N. (2018). De novo mutations in regulatory elements in neurodevelopmental disorders. *555*(7698), 611-616.

Singh, T., Kurki, M. I., Curtis, D., Purcell, S. M., Crooks, L., McRae, J., . . . Breen, G. (2016). Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nature neuroscience*.

Singh, T., Walters, J. T., Johnstone, M., Curtis, D., Suvisaari, J., Torniainen, M., . . . McIntosh, A. M. J. N. g. (2017). The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *49*(8), 1167-1173.

Sjøholt, G., Ebstein, R., Lie, R., Berle, J. Ø., Mallet, J., Deleuze, J., . . . Bannoura, I. (2004). Examination of IMPA1 and IMPA2 genes in manic-depressive patients: association between IMPA2 promoter polymorphisms and bipolar disorder. *Molecular Psychiatry, 9*(6), 621-629.

Sobreira, N., Schiettecatte, F., Valle, D., & Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Human mutation, 36*(10), 928-930.

Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shiel, J. A., Thomas, N. S., . . . Cooper, D. N. (2003). Human gene mutation database (HGMD®): 2003 update. *Human mutation, 21*(6), 577-581.

Stevens, S. J., van Ravenswaaij□Arts, C. M., Janssen, J. W., Klein Wassink□Ruiter, J. S., van Essen, A. J., Dijkhuizen, T., . . . Smeets, E. E. (2011). MYT1L is a candidate gene for intellectual disability in patients with 2p25. 3 (2pter) deletions. *American Journal of Medical Genetics Part A, 155*(11), 2739-2745.

Study, D. D. D., McRae, J. F., Clayton, S., Fitzgerald, T. W., Kaplanis, J., Prigmore, E., . . . Akawi, N. J. N. (2017). Prevalence and architecture of de novo mutations in developmental disorders. *542*(7642), 433-438.

Swaab, D. F. (2007). Sexual differentiation of the brain and behavior. *Best practice & research clinical endocrinology & metabolism, 21*(3), 431-444.

Tammimies, K., Marshall, C. R., Walker, S., Kaur, G., Thiruvahindrapuram, B., Lionel, A. C., . . . Weksberg, R. (2015). Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder. *Jama, 314*(9), 895-903.

Tan, C., Topper, S., Del Gaudio, D., Nelakuditi, V., Shchelochkov, O., Nowaczyk, M., . . . Meeks, N. (2015). Characterization of patients referred for non□specific intellectual disability testing: the importance of autosomal genes for diagnosis. *Clinical genetics*.

Tarpey, P. S., Smith, R., Pleasance, E., Whibley, A., Edkins, S., Hardy, C., . . . Menzies, A. (2009). A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nature genetics, 41*(5), 535-543.

Tatton-Brown, K., Loveday, C., Yost, S., Clarke, M., Ramsay, E., Zachariou, A., . . . Rittinger, O. J. T. A. J. o. H. G. (2017). Mutations in epigenetic regulation genes are a major cause of overgrowth with intellectual disability. *100*(5), 725-736.

Teng, B. L., Nikolova, V. D., Riddick, N. V., Agster, K. L., Crowley, J. J., Baker, L. K., . . . Moy, S. S. (2016). Reversal of social deficits by subchronic oxytocin in two autism mouse models. *Neuropharmacology, 105*, 61-71.

Thye, M. D., Bednarz, H. M., Herringshaw, A. J., Sartin, E. B., & Kana, R. K. (2018). The impact of atypical sensory processing on social impairments in autism spectrum disorder. *Developmental cognitive neuroscience, 29*, 151-167.

Tick, B., Bolton, P., Happé, F., Rutter, M., & Rijsdijk, F. (2016). Heritability of autism spectrum disorders: a meta□analysis of twin studies. *Journal of Child Psychology and Psychiatry, 57*(5), 585-595.

Trost, B., Engchuan, W., Nguyen, C. M., Thiruvahindrapuram, B., Dolzhenko, E., Backstrom, I., . . . Dov, A. J. N. (2020). Genome-wide detection of tandem DNA repeats that are expanded in autism. 1-7.

Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J. J., Moser, M.-B., & Moser, E. I. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature, 561*(7721), 57-62.

Turrini, M., Bourgain, C. J. S. S., & Medicine. (2020). Genomic susceptibility in practice: the regulatory trajectory of non-rare thrombophilia (NRT) genetic tests in the clinical management of venous thrombo-embolism (VTE). 112903.

Turro, E., Astle, W. J., Megy, K., Gräf, S., Greene, D., Shamardina, O., . . . Thys, C. J. N. (2020). Whole-genome sequencing of patients with rare diseases in a national health system. *583*(7814), 96-102.

Tzschach, A., Grasshoff, U., Beck-Woedl, S., Dufke, C., Bauer, C., Kehrer, M., . . . Di Donato, N. (2015). Next-generation sequencing in X-linked intellectual disability. *European Journal of Human Genetics*, *23*(11), 1513-1518

Ullah, M. A., Husseni, A. M., & Mahmood, S. U. J. I. J. C. M. P. H. (2017). Consanguineous marriages and their detrimental outcomes in Pakistan: an urgent need for appropriate measures. *5*(1), 1-3.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy□Moonshine, A., . . . Thibault, J. (2013). From FastQ data to high□confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics, 43*(1), 11.10. 11-11.10. 33.

van der Vaart, B., Franker, M. A., Kuijpers, M., Hua, S., Bouchet, B. P., Jiang, K., . . . Akhmanova, A. (2012). Microtubule plus-end tracking proteins SLAIN1/2 and ch-TOG promote axonal development. *The Journal of Neuroscience, 32*(42), 14722-14728a.

van Karnebeek, C. D., & Stockler, S. (2012). Treatable inborn errors of metabolism causing intellectual disability: a systematic literature review. *Molecular genetics and metabolism, 105*(3), 368-381.

Van Scherpenzeel, M., Timal, S., Rymen, D., Hoischen, A., Wuhrer, M., Hipgrave-Ederveen, A., . . . Edvardson, S. (2014). Diagnostic serum glycosylation profile in patients with intellectual disability as a result of MAN1B1 deficiency. *Brain, 137*(4), 1030-1038.

Vasli, N., Ahmed, I., Mittal, K., Ohadi, M., Mikhailov, A., Rafiq, M. A., . . . Ayub, M. (2015). Identification of a homozygous missense mutation in LRP2 and a hemizygous missense mutation in TSPYL2 in a family with mild intellectual disability. *Psychiatric Genetics*, *26*(2), 66-73

Vincent, J. B. (2013). Identification of the dcps gene on 11q24. 2, which encodes the human decapping enzyme scavenger, in non-syndromic autosomal recessive mental retardation, diagnostic probes thereof and methods of identifying subjects with same. In: Google Patents.

Vissers, L. E., Gilissen, C., & Veltman, J. A. (2015). Genetic studies in intellectual disability and related disorders. *Nature Reviews Genetics*, *17*(1), 9-18

Walker, R. L., Ramaswami, G., Hartl, C., Mancuso, N., Gandal, M. J., De La Torre-Ubieta, L., . . . Geschwind, D. H. J. C. (2019). Genetic control of expression and splicing in developing human brain informs disease mechanisms. *179*(3), 750-771. e722.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F., . . . Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome research, 17*(11), 1665-1674.

Wang, T., Guo, H., Xiong, B., Stessman, H. A., Wu, H., Coe, B. P., . . . Hoekzema, K. J. N. c. (2016). De novo genic mutations among a Chinese autism spectrum disorder cohort. *7*, 13316.

Ware, J. S., Samocha, K. E., Homsy, J., & Daly, M. J. J. C. p. i. h. g. (2015). Interpreting de novo variation in human disease using denovolyzeR. *87*(1), 7.25. 21-27.25. 15.

Wasser, C. R., & Herz, J. (2017). Reelin: neurodevelopmental architect and homeostatic regulator of excitatory synapses. *Journal of Biological Chemistry, 292*(4), 1330-1338.

Weiner, D. J., Wigdor, E. M., Ripke, S., Walters, R. K., Kosmicki, J. A., Grove, J., . . . Bybjerg-Grauholm, J. (2017). Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nature genetics, 49*(7), 978-985.

Weiss, L., Escayg, A., Kearney, J., Trudeau, M., MacDonald, B., Mori, M., . . . Meisler, M. (2003). Sodium channels SCN1A, SCN2A and SCN3A in familial autism. *Molecular Psychiatry, 8*(2), 186-194.

Welch, J. D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., & Macosko, E. Z. J. C. (2019). Single-cell multi-omic integration compares and contrasts features of brain cell identity. *177*(7), 1873-1887. e1817.

Wolf, E., Kim, P. S., & Berger, B. J. P. S. (1997). MultiCoil: a program for predicting two□and three□stranded coiled coils. *6*(6), 1179-1189.

Wright, A. (1997). Limbic system: amygdala. *Neuroscience Online. http://neuroscience. uth. tmc. edu/s4/chapter06. html*.

Wright, C. F., Fitzgerald, T. W., Jones, W. D., Clayton, S., McRae, J. F., van Kogelenberg, M., . . . Bayzetinova, T. (2015). Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *The Lancet, 385*(9975), 1305-1314.

Wynshaw□Boris, A. (2007). Lissencephaly and LIS1: insights into the molecular mechanisms of neuronal migration and development. *Clinical genetics, 72*(4), 296-304.

Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K., . . . Hughes, T. R. J. S. (2015). The human splicing code reveals new insights into the genetic determinants of disease. *347*(6218).

Yang, H., & Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature protocols, 10*(10), 1556-1566.

Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A., . . . Niu, Z. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *New England Journal of Medicine, 369*(16), 1502-1511.

Yang, Y., Muzny, D. M., Xia, F., Niu, Z., Person, R., Ding, Y., . . . Buhay, C. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. *Jama, 312*(18), 1870-1879.

Yu, H., Tardivo, L., Tam, S., Weiner, E., Gebreab, F., Fan, C., . . . Yang, X. (2011). Next-generation sequencing to generate interactome datasets. *Nature methods, 8*(6), 478-480.

Yuen, R. K., Merico, D., Bookman, M., Howe, J. L., Thiruvahindrapuram, B., Patel, R. V., . . . Wang, Z. J. N. n. (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *20*(4), 602-611.

Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., Van Der Zwan, J., . . . La Manno, G. J. C. (2018). Molecular architecture of the mouse nervous system. *174*(4), 999-1014. e1022.

Zhang, K., Lentini, J. M., Prevost, C. T., Hashem, M. O., Alkuraya, F. S., & Fu, D. (2020). An intellectual disability□associated missense variant in TRMT1 impairs tRNA modification and reconstitution of enzymatic activity. *Human mutation*.

Zhang, R.-R., Cui, Q.-Y., Murai, K., Lim, Y. C., Smith, Z. D., Jin, S., . . . Wu, H.-P. (2013). Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell, 13*(2), 237-245.

Zhubi, A., Chen, Y., Dong, E., Cook, E., Guidotti, A., & Grayson, D. (2014). Increased binding of MeCP2 to the GAD1 and RELN promoters may be mediated by an enrichment of 5-hmC in autism spectrum disorder (ASD) cerebellum. *Translational psychiatry, 4*(1), e349.

Zyla, J., Marczyk, M., Domaszewska, T., Kaufmann, S. H., Polanska, J., & Weiner 3rd, J. (2019). Gene set enrichment for reproducible science: comparison of CERNO and eight other algorithms. *Bioinformatics, 35*(24), 5146-5154.