Testing the network hypothesis for schizophrenia and autism spectrum disorder using whole exome sequencing data

Vanessa Fulginiti

Division of Human Genetics

McGill University

Montreal, Quebec, Canada

August, 2017

A thesis submitted to McGill University

in partial fulfillment of the requirements of the degree of

Master of Science in Human Genetics.

Abstract

Background Schizophrenia (SCZ) and autism spectrum disorder (ASD) are psychiatric diseases with complex inheritance. Genetic studies have not identified susceptibility genes to adequately explain the heritability and the etiology of these diseases is largely unknown.

Methods We identified susceptibility genes enriched for de novo mutations (DNMs) in at least two independent whole exome sequencing (WES) publications. Genes associated with hypertrophic cardiomyopathy (HC) were used as control genes. We selected for rare inherited and DNMs in the ASD network using a WES dataset (2392 ASD families) and in the SCZ network using three independent WES datasets (35 trios; 598 trios; 5090 case controls). We compared the mutation load in the 'disease network' between affected and unaffected individuals for each dataset. The analyses were repeated using the 'HC network'.

Results 14 SCZ genes and 143 ASD genes were identified. When using the 598 SCZ trios, probands were enriched in functional variants relative to the average mutation load of parents in the SCZ network ($p = 0.04$) but not in the HC genes ($p = 0.23$). All functional variants identified in the SCZ network were inherited. Similar results were obtained using the case control dataset (SCZ network: $p = 0.02$; HC network: $p = 0.09$). When analyzing ASD sibpairs, *unaffected* siblings were significantly enriched in functional variants in the ASD network ($p = 0.02$) but also throughout the exome based on a permutation analysis using all genes with functional variants. When controlling for sequencing depth through a conditional logistic regression and applying stricter filtering criteria, the difference was not statistically significant ($p = 0.1358$).

Conclusions We provide preliminary evidence that the accumulation of rare variants (mainly inherited) in the identified SCZ susceptibility genes is associated with SCZ. However, this was not the case for the ASD dataset that we had access to.

Résumé

Contexte La schizophrénie (SCZ) et le trouble du spectre autistique (ASD) sont des maladies psychiatriques avec des héritages complexes. Les études génétiques n'ont pas identifié des gènes de susceptibilité pour expliquer l'héritabilité et l'étiologie de ces maladies est inconnue.

Méthodes Nous avons identifié les gènes de susceptibilité de maladies enrichies pour les mutations de novo (DNM) signalés par au moins deux publications de séquençage complet (WES). Les gènes témoins sont associés à la cardiomyopathie hypertrophique (HC). Nous avons sélectionnées des variantes rares (héréditaires et de novo) des gènes ASD en utilisant 2392 familles atteints du ASD et des gènes SCZ en utilisant 35 trios, 598 trios et 5090 individus affectés/non affectés. Nous avons comparé le nombre des mutations dans le 'réseau de maladie' entre les individus affectés et non affectés. L'analyse a été répétée en utilisant les gènes de HC.

Résultats 14 gènes de SCZ et 143 gènes de ASD ont été identifié. Les proposants de les 598 trios ont été enrichis en variantes fonctionnelles du réseau de SCZ par rapport à la moyenne nombre de mutations de leurs parents ($p = 0,04$) mais pas dans les gènes de HC ($p = 0,23$). Les variantes dans les gènes de SCZ ont été héritées. Résultats similaires one été observé pour les 5090 individus affectés et non affectés (SCZ: $p = 0,02$; HC: $p = 0,09$). L'ordre de l'analyse de ASD, les frères et sœurs (FES) non affectés ont été enrichis en variantes fonctionnelles par rapport aux proposants dans les gènes de ASD ($p = 0.02$) et aussi tout au long de l'exome basé sur des permutations en utilisant tout les gènes avec les variantes fonctionnelles. Ce résultat n'est pas significative ($p = 0.1358$) lors de la comptabilisation de l'effet de séquençage par une régression logistique conditionnelle et en utilisant les critères de filtrage plus stricts.

Conclusions On observe que l'accumulation de variantes rares (principalement héréditaires) dans le réseau de 14 gènes est associée à SCZ. Cependant, ce n'était pas le cas pour l'ensemble de données ASD auquel nous avons eu accès.

Table of Contents

Acknowledgements

Thank you to my supervisor Dr. Yannis Trakadis for the opportunity to be a part of his team, for instilling in me the fundamental lessons of professionalism and for his guidance and suggestions over these past two years.

I would also like to thank my cherished committee members, Dr. Jean-Baptiste Rivière and Dr. Celia Greenwood, for the insightful scientific discussions and for sharing their expertise to help improve my project. I would like to acknowledge Dr. Celia Greenwood for providing the R code for the conditional logistic regression and statistical advice.

I'd like to extend my gratitude to Dr. Guy Rouleau for sharing the 35 local SCZ trios dataset. Thank you to Dr. Rouleau and Alexandre Dionne Laporte from his lab for the bioinformatics support. Thank you to Bill Qi and Anthony Chen for working alongside me and helping to develop the bioinformatic pipeline used for variant selection. I'd like to acknowledge Anthony Chen for his work on developing a script to automate the OMIM search.

Special thanks to Ross Mackay and Rimi Joshi for their administrative support. Thank you to the HGSS and all of the inspiring graduate students I've met, including Kellie McDonald for enriching my social experience during my time here.

Last but not least, my deep gratitude goes to the two most important women in my life and my source of strength, Patricia Lattuca and Genelle F Harrison, for their unwavering support, words of encouragement and boundless wisdom. This degree would not have been possible without them. Thank you.

Preface and Contribution of Authors

This thesis is presented as a traditional monograph format for a Master's of Science, as described in the Thesis Preparation Guidelines by the Department of Graduate and Postdoctoral Studies. The work described here was under the supervision of Dr. Yannis Trakadis.

Dr. Trakadis conceptualized the "network hypothesis" forming the basis of this thesis. Dr. Trakadis and the candidate applied for and obtained ethics approval from the McGill University Health Centre for this study. Dr. Trakadis obtained approved access to the autism dataset from the National Database of Autism Research and to the schizophrenia datasets from the Database of Genotypes and Phenotypes. Pertaining to the work presented in this thesis, the segregation analysis of the data was done by the candidate and Alexandre Dionne Laporte. The variant selection pipeline was developed with the help of Anthony Chen and Bill Qi. The candidate reviewed the literature summarized in the introduction section, performed a literature review of de novo mutation studies in autism and schizophrenia to identify the list of genes included in the susceptibility network and performed all analyses described in this thesis. Interpretation of results was performed in collaboration with Dr. Trakadis and in consultation with Dr. Rivière and Dr. Greenwood. Dr. Greenwood provided the code for the conditional logistic regressions.

A manuscript for the significant results supporting our hypothesis in schizophrenia is under review for publication.

List of Abbreviations

A2M	alpha-2-macroglobulin
ABCA12	ATP binding cassette subfamily A member 12
ABCA13	ATP binding cassette subfamily A member 13
ABCA3	ATP binding cassette subfamily A member 3
ABCD1	ATP binding cassette subfamily D member 1
ABI3BP	ABI family member 3 binding protein
ACADS	acyl-CoA dehydrogenase, C-2 to C-3 short chain
ACHE	acetylcholinesterase (Cartwright blood group)
ADAM22	ADAM metallopeptidase domain 22
ADH1C	alcohol dehydrogenase 1C (class I), gamma polypeptide
ADNP	activity dependent neuroprotector homeobox
AFF4	AF4/FMR2 family member 4
AHDC1	AT-hook DNA binding motif containing 1
AHNAK2	AHNAK nucleoprotein 2
AIP	aryl hydrocarbon receptor interacting protein
AKAP9	A-kinase anchoring protein 9
ALDH5A1	aldehyde dehydrogenase 5 family member A1
ANK2	ankyrin 2
ANK3	ankyrin 3
ANKRD11	ankyrin repeat domain 11
ANNOVAR	a software tool that annotates variants
AOC3	amine oxidase, copper containing 3
AP3D1	adaptor related protein complex 3 delta 1 subunit
APH1A	aph-1 homolog A, gamma-secretase subunit
APP	amyloid beta precursor protein
ARC	Activity-regulated cytoskeleton-associated protein
ARHGAP15	Rho GTPase activating protein 15
ARID1B	AT-rich interaction domain 1B
ARSA	arylsulfatase A
ASD	Autism spectrum disorder
ASH1L	ASH1 like histone lysine methyltransferase
ASS1	argininosuccinate synthase 1
ASXL3	additional sex combs like 3, transcriptional regulator
ATP13A2	ATPase 13A2
ATP1A3	ATPase Na+/K+ transporting subunit alpha 3
ATP2A2	ATPase sarcoplasmic/endoplasmic reticulum Ca2+ transporting 2
ATXN1	ataxin 1
ATXN2	ataxin 2
BCKDHA	branched chain keto acid dehydrogenase E1, alpha polypeptide
BCKDHB	branched chain keto acid dehydrogenase E1 subunit beta
BCKDK	branched chain ketoacid dehydrogenase kinase
BCL11A	B-cell CLL/lymphoma 11A

BICC1	BicC family RNA binding protein 1
BRD3	bromodomain containing 3
C10orf2	twinkle mtDNA helicase
C10orf90	chromosome 10 open reading frame 90
CACNA1A	calcium voltage-gated channel subunit alpha1 A
CACNA1E	calcium voltage-gated channel subunit alpha1 E
CACNA1G	calcium voltage-gated channel subunit alpha1 G
CACNB1	calcium voltage-gated channel auxiliary subunit beta 1
CADPS	calcium dependent secretion activator
CALR3	calreticulin 3
CAPN12	calpain 12
CDCA7L	cell division cycle associated 7 like
CDKL3	cyclin dependent kinase like 3
CDT	Critical disease threshold
CFAP157	cilia and flagella associated protein 157
CFH	complement factor H
CHD1	chromodomain helicase DNA binding protein 1
CHD2	chromodomain helicase DNA binding protein 2
CHD4	chromodomain helicase DNA binding protein 4
CHD8	chromodomain helicase DNA binding protein 8
CHEK2	checkpoint kinase 2
CHRNG	cholinergic receptor nicotinic gamma subunit
CHST5	carbohydrate sulfotransferase 5
CI	Confidence interval
CLCN2	chloride voltage-gated channel 2
CLN3	CLN3, battenin
CLN6	CLN6, transmembrane ER protein
CLSTN3	calsyntenin 3
CNV	Copy number variant
CP	ceruloplasmin
CPNE7	copine 7
CPOX	coproporphyrinogen oxidase
CREBBP	CREB binding protein
CSTB	cystatin B
CTLA4	cytotoxic T-lymphocyte associated protein 4
CTTNBP2	cortactin binding protein 2
CUL3	cullin 3
CYP27A1	cytochrome P450 family 27 subfamily A member 1
CYTH4	cytohesin 4
DAX1	nuclear receptor subfamily 0 group B member 1
dbGaP	Database of Genotypes and Phenotypes
DBT	dihydrolipoamide branched chain transacylase E2
DCAF17	DDB1 and CUL4 associated factor 17

DCAF4	DDB1 and CUL4 associated factor 4
DCTN5	dynactin subunit 5
DDX50	DExD-box helicase 50
DEPDC5	DEP domain containing 5
DICER1	dicer 1, ribonuclease III
DIP2A	disco interacting protein 2 homolog A
DIRC2	disrupted in renal carcinoma 2
DJ1	Parkinsonism associated deglycase
DMD	dystrophin
DMPK	DM1 protein kinase
DNAH12	dynein axonemal heavy chain 12
DNAH7	dynein axonemal heavy chain 7
DNAH9	dynein axonemal heavy chain 9
DNAJC16	DnaJ heat shock protein family (Hsp40) member C16
DNAJC5	DnaJ heat shock protein family (Hsp40) member C5
DNASE1	deoxyribonuclease 1
DNM	De novo mutation
DNMT1	DNA methyltransferase 1
DNMT3A	DNA methyltransferase 3 alpha
DSCAM	DS cell adhesion molecule
DSM-V	Diagnostic and statistical manual of mental disorders V
DSTYK	dual serine/threonine and tyrosine protein kinase
DTHD1	death domain containing 1
DYRK1A	dual specificity tyrosine phosphorylation regulated kinase 1A
ECM1	extracellular matrix protein 1
ECSIT	ECSIT signalling integrator
EEF1A2	eukaryotic translation elongation factor 1 alpha 2
EIF3G	eukaryotic translation initiation factor 3 subunit G
ELK1	ELK1, ETS transcription factor
ENO3	enolase 3
EPAS1	endothelial PAS domain protein 1
EPHB6	EPH receptor B6
EPM2A	EPM2A, laforin glucan phosphatase
ERBB2IP	erbb2 interacting protein
ETFB	electron transfer flavoprotein beta subunit
FAH	fumarylacetoacetate hydrolase
FAT1	FAT atypical cadherin 1
FCGR2A	Fc fragment of IgG receptor IIa
FCGR2B	Fc fragment of IgG receptor IIb
FGFR2	fibroblast growth factor receptor 2
FIG4	FIG4 phosphoinositide 5-phosphatase
FKBP5	FK506 binding protein 5
FLCN	folliculin

FLNC	filamin C
FN1	fibronectin 1
FOXP1	forkhead box P1
FTL	ferritin light chain
GABRB3	gamma-aminobutyric acid type A receptor beta3 subunit
GAL	galanin and GMAP prepropeptide
GALR1	galanin receptor 1
GALR2	galanin receptor 2
GATA6	GATA binding protein 6
GBA	glucosylceramidase beta
GCN1L1	GCN1, eIF2 alpha kinase activator homolog
GDF1	growth differentiation factor 1
GIGYF1	GRB10 interacting GYF protein 1
GLUD1	glutamate dehydrogenase 1
GLUD2	glutamate dehydrogenase 2
GNAS	GNAS complex locus
GNS	glucosamine (N-acetyl)-6-sulfatase
GPR139	G protein-coupled receptor 139
GPR151	G protein-coupled receptor 151
GPR153	G protein-coupled receptor 153
GRIN2B	glutamate ionotropic receptor NMDA type subunit 2B
GRN	granulin precursor
GSS	glutathione synthetase
GTF3C2	general transcription factor IIIC subunit 2
H2AFV	H2A histone family member V
H_A	Alternative hypothesis
HARS	histidyl-tRNA synthetase
HC	Hypertrophic cardiomyopathy
HCRT	hypocretin neuropeptide precursor
HERC1	HECT and RLD domain containing E3 ubiquitin protein ligase family member 1
HEXA	hexosaminidase subunit alpha
HEXB	hexosaminidase subunit beta
HFE	hemochromatosis
HGS	hepatocyte growth factor-regulated tyrosine kinase substrate
HIVEP3	human immunodeficiency virus type I enhancer binding protein 3
HLA-DQB1	major histocompatibility complex, class II, DQ beta 1
HMBS	hydroxymethylbilane synthase
HNF1A	HNF1 homeobox A
HNF1B	HNF1 homeobox B
HSD17B10	hydroxysteroid 17-beta dehydrogenase 10
HTR2A	5-hydroxytryptamine receptor 2A
IDUA	iduronidase, alpha-L-
IL1R2	interleukin 1 receptor type 2

ILF2	interleukin enhancer binding factor 2
indel	Insertion and deletion
INTS6	integrator complex subunit 6
IRF2BPL	interferon regulatory factor 2 binding protein like
ITM2B	integral membrane protein 2B
ITPR1	inositol 1,4,5-trisphosphate receptor type 1
JAK2	Janus kinase 2
JMJD1C	jumonji domain containing 1C
JPH2	junctophilin 2
JPH3	junctophilin 3
KAT2B	lysine acetyltransferase 2B
KATNAL2	katanin catalytic subunit A1 like 2
KCNN4	potassium calcium-activated channel subfamily N member 4
KCNT1	potassium sodium-activated channel subfamily T member 1
KDM5B	lysine demethylase 5B
KDM6B	lysine demethylase 6B
KIAA0226	RUN and cysteine rich domain containing beclin 1 interacting protein
KIAA1644	KIAA1644
KIAA1967	cell cycle and apoptosis regulator 2
KIRREL3	kirre like nephrin family adhesion molecule 3
KMT2C	lysine methyltransferase 2C
KMT2E	lysine methyltransferase 2E
KRTAP9-3	keratin associated protein 9-3
LGI1	leucine rich glioma inactivated 1
LPHN2	adhesion G protein-coupled receptor L2
LRP1	LDL receptor related protein 1
LRP1B	LDL receptor related protein 1B
LRRC18	leucine rich repeat containing 18
MACC1	MACC1, MET transcriptional regulator
MACF1	microtubule-actin crosslinking factor 1
MAN2B1	mannosidase alpha class 2B member 1
MANSC1	MANSC domain containing 1
MAOA	monoamine oxidase A
MAP3K8	mitogen-activated protein kinase kinase kinase 8
MAPT	microtubule associated protein tau
MBD5	methyl-CpG binding domain protein 5
MCM2	minichromosome maintenance complex component 2
MCPH1	microcephalin 1
MDD1	major depressive disorder 1
MDD2	major depressive disorder 2
MECP2	methyl-CpG binding protein 2
MED12	mediator complex subunit 12
MED13L	mediator complex subunit 13 like

METTL2B	methyltransferase like 2B
MFRP	membrane frizzled-related protein
MIB1	mindbomb E3 ubiquitin protein ligase 1
MKL2	MKL1/myocardin like 2
MKLN1	muskelin 1
MMACHC	methylmalonic aciduria (cobalamin deficiency) cblC type, with homocystinuria
MOV10	Mov10 RISC complex RNA helicase
MPDZ	multiple PDZ domain crumbs cell polarity complex component
MPO	myeloperoxidase
MTHFR	methylenetetrahydrofolate reductase
MUC1	mucin 1, cell surface associated
MUC4	mucin 4, cell surface associated
MYCBP2	MYC binding protein 2, E3 ubiquitin protein ligase
MYH10	myosin heavy chain 10
MYH6	myosin heavy chain 6
MYH9	myosin heavy chain 9
MYL2	myosin light chain 2
MYL3	myosin light chain 3
MYO1A	myosin IA
MYO5B	myosin VB
MYO7A	myosin VIIA
MYO9B	myosin IXB
MYOF	myoferlin
MYOM2	myomesin 2
MYPN	myopalladin
MYT1L	myelin transcription factor 1 like
NAA15	N(alpha)-acetyltransferase 15, NatA auxiliary subunit
NAGS	N-acetylglutamate synthase
NCKAP1	NCK associated protein 1
NCKAP5	NCK associated protein 5
NDAR	National Database of Autism Research
NDP	NDP, norrin cystine knot growth factor
NEB	nebulin
NEXN	nexilin F-actin binding protein
NFASC	neurofascin
NHLRC1	NHL repeat containing E3 ubiquitin protein ligase 1
NINL	ninein like
NIMH	National Institute of Mental Health
NISCH	nischarin
NKX2-5	NK2 homeobox 5
NKX2-6	NK2 homeobox 6
NLGN3	neuroligin 3
NMDAR	N-Methyl-D-aspartic acid receptor

NOS3	nitric oxide synthase 3
NOTCH3	notch 3
NOV	nephroblastoma overexpressed
NPC1	NPC intracellular cholesterol transporter 1
NPC2	NPC intracellular cholesterol transporter 2
NRXN1	neurexin 1
NT5DC3	5'-nucleotidase domain containing 3
OBSCN	obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF
OBSL1	obscurin like 1
OGG1	8-oxoguanine DNA glycosylase
OMIM	Online Mendelian Inheritance in Man
OPLAH	5-oxoprolinase, ATP-hydrolysing
OR	Odds ratio
OR52M1	olfactory receptor family 52 subfamily M member 1
P2RX5	purinergic receptor P2X 5
PAF1	PAF1 homolog, Paf1/RNA polymerase II complex component
PAH	phenylalanine hydroxylase
PAK3	p21 (RAC1) activated kinase 3
PANX2	pannexin 2
PCDH15	protocadherin related 15
PCDH19	protocadherin 19
PCDHB16	protocadherin beta 16
PCNX	pecanex homolog 1
PDE11A	phosphodiesterase 11A
PDGFB	platelet derived growth factor subunit B
PDGFRB	platelet derived growth factor receptor beta
PHF2	PHD finger protein 2
PIAS1	protein inhibitor of activated STAT 1
PINK1	PTEN induced putative kinase 1
PITPNM1	phosphatidylinositol transfer protein membrane associated 1
PIWIL4	piwi like RNA-mediated gene silencing 4
PLAU	plasminogen activator, urokinase
PLCD1	phospholipase C delta 1
PLCD4	phospholipase C delta 4
PLD3	phospholipase D family member 3
PLEKHA4	pleckstrin homology domain containing A4
PLPP3	phospholipid phosphatase 3
PLXNB1	plexin B1
POGZ	pogo transposable element derived with ZNF domain
POLR1E	RNA polymerase I subunit E
POLR2A	RNA polymerase II subunit A
POM121C	POM121 transmembrane nucleoporin C
PPOX	protoporphyrinogen oxidase

PPP1R12B	protein phosphatase 1 regulatory subunit 12B
PPT1	palmitoyl-protein thioesterase 1
PRDM8	PR/SET domain 8
PRKAR1A	protein kinase cAMP-dependent type I regulatory subunit alpha
PRNP	prion protein
PRODH	proline dehydrogenase 1
PRPF6	pre-mRNA processing factor 6
PSD-95	Postsynaptic density protein 95
PSEN1	presenilin 1
PSEN2	presenilin 2
PTEN	phosphatase and tensin homolog
PTK7	protein tyrosine kinase 7 (inactive)
PTPA	protein phosphatase 2 phosphatase activator
PTPN22	protein tyrosine phosphatase, non-receptor type 22
PXK	PX domain containing serine/threonine kinase like
RANBP17	RAN binding protein 17
RBM10	RNA binding motif protein 10
RBMS3	RNA binding motif single stranded interacting protein 3
RELN	reelin
RFX3	regulatory factor X3
RGS12	regulator of G protein signaling 12
RHAG	Rh-associated glycoprotein
RNF139	ring finger protein 139
ROGDI	rogdi homolog
RPL10	ribosomal protein L10
RPS6KA3	ribosomal protein S6 kinase A3
RTN4RL1	reticulon 4 receptor like 1
RYR2	ryanodine receptor 2
SBF1	SET binding factor 1
SCN2A	sodium voltage-gated channel alpha subunit 2
SCNA	alpha-synuclein
SCRIB	scribbled planar cell polarity protein
SCZ	Schizophrenia
SETD1A	SET domain containing 1A
SETD2	SET domain containing 2
SETD5	SET domain containing 5
SETD7	SET domain containing lysine methyltransferase 7
SGCE	sarcoglycan epsilon
SHANK2	SH3 and multiple ankyrin repeat domains 2
SHANK3	SH3 and multiple ankyrin repeat domains 3
SLC12A6	solute carrier family 12 member 6
SLC1A1	solute carrier family 1 member 1
SLC20A2	solute carrier family 20 member 2

SLC25A13	solute carrier family 25 member 13
SLC30A5	solute carrier family 30 member 5
SLC39A5	solute carrier family 39 member 5
SLC6A1	solute carrier family 6 member 1
SLC6A19	solute carrier family 6 member 19
SLC6A3	solute carrier family 6 member 3
SLC6A8	solute carrier family 6 member 8
SLC7A7	solute carrier family 7 member 7
SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4
SMARCC2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 2
SNAPC5	small nuclear RNA activating complex polypeptide 5
SNCA	synuclein alpha
SNCB	synuclein beta
SNTG1	syntrophin gamma 1
SNV	Single nucleotide variant
SOBP	sine oculis binding protein homolog
SPAST	spastin
SPG20	spartin
STAC2	SH3 and cysteine rich domain 2
STAT2	signal transducer and activator of transcription 2
SUV420H1	lysine methyltransferase 5B
SVIL	supervillin
SYNE1	spectrin repeat containing nuclear envelope protein 1
SYNE2	spectrin repeat containing nuclear envelope protein 2
SYNGAP1	synaptic Ras GTPase activating protein 1
TADA	Transmission and de novo association
TBC1D7	TBC1 domain family member 7
TBL1XR1	transducin beta like 1 X-linked receptor 1
TBP	TATA-box binding protein
TBR1	T-box, brain 1
TBX1	T-box 1
TCAP	titin-cap
TCF3	transcription factor 3
TCF7L2	transcription factor 7 like 2
TDRD5	tudor domain containing 5
TECTA	tectorin alpha
TGM3	transglutaminase 3
TLK2	tousled like kinase 2
TMEM39B	transmembrane protein 39B
TNFSF4	TNF superfamily member 4
TNNC1	troponin C1, slow skeletal and cardiac type

TNNI3	troponin I3, cardiac type
TNNT2	troponin T2, cardiac type
TNRC6B	trinucleotide repeat containing 6B
TOM1L1	target of myb1 like 1 membrane trafficking protein
TP53	tumor protein p53
TPH2	tryptophan hydroxylase 2
TPM1	tropomyosin 1
TREX1	three prime repair exonuclease 1
TRIO	trio Rho guanine nucleotide exchange factor
TRIP12	thyroid hormone receptor interactor 12
TTC19	tetratricopeptide repeat domain 19
TTN	titin
TTR	transthyretin
TUBA1A	tubulin alpha 1a
TWIST1	twist family bHLH transcription factor 1
UBR4	ubiquitin protein ligase E3 component n-recognin 4
UGT2B10	UDP glucuronosyltransferase family 2 member B10
UNC79	unc-79 homolog, NALCN channel complex subunit
USH2A	usherin
USP45	ubiquitin specific peptidase 45
VAV3	vav guanine nucleotide exchange factor 3
VCL	vinculin
VHL	von Hippel-Lindau tumor suppressor
VPS13A	vacuolar protein sorting 13 homolog A
VPS35	VPS35, retromer complex component
WAC	WW domain containing adaptor with coiled-coil
WDFY3	WD repeat and FYVE domain containing 3
WDR20	WD repeat domain 20
WDR66	WD repeat domain 66
WES	Whole exome sequencing
WFS1	wolframin ER transmembrane glycoprotein
XIRP1	xin actin binding repeat containing 1
XPR1	xenotropic and polytropic retrovirus receptor 1
XRCC5	X-ray repair cross complementing 5
XRN2	5'-3' exoribonuclease 2
ZBTB20	zinc finger and BTB domain containing 20
ZDHHC9	zinc finger DHHC-type containing 9
ZFC3H1	zinc finger C3H1-type containing
ZFHX3	zinc finger homeobox 3
ZFPM2	zinc finger protein, FOG family member 2
ZFYVE26	zinc finger FYVE-type containing 26
ZNF155	zinc finger protein 155
ZNF559	zinc finger protein 559

List of Tables

Chapter 2:

Chapter 3:

List of Figures

Chapter 3:

Chapter 4:

Chapter 1: Introduction

Schizophrenia Genetics

Schizophrenia is characterized by different features including hallucinations, delusions, incoherent speech, catatonia, or negative symptoms such as cognitive deficits. Symptoms of SCZ can be heterogeneous in nature and as such the timing and duration of symptoms can lead to other diagnoses in the schizophrenic spectrum including but not limited to schizoaffective disorder, schizotypal personality disorder and schizophreniform disorder. The first sign of psychosis typically occurs in late adolescence to early adulthood (age of onset 18 – 35 years old) for both males and females. However, there are sex-specific differences in the clinical signs of SCZ and the age of onset. For example, the incidence of SCZ is lower in females, they tend to be diagnosed relatively later in life and tend to be less severely affected with relatively better social functioning and less frequent negative symptoms (e.g. apathy). For diagnostic criteria regarding SCZ, please refer to the diagnostic and statistical manual of mental disorders V, DSM V (1). No laboratory or psychometric tests are currently available for SCZ.

Genetic diagnostic tools could aid in an earlier and more accurate diagnosis of SCZ, given its a disorder of complex inheritance with a high estimated heritability of ~80% based on family, twin, and adoption studies (2, 3). In the 1990's and early 2000's whole genome linkage analysis, traditionally successful for Mendelian disorders, was used to narrow down chromosomal segments associated with SCZ by analysing inherited polymorphic genetic markers (specifically microsatellites) that segregate with affected family members (4-13). Linkage studies supported by the NIMH Genetics Initiative Millenium Schizophrenia Consortium analyzing families recruited by Washington University, Columbia University and Harvard University did not find any significant loci in 43 European-American pedigrees (7) or 30 African-American pedigrees

(8). Nonparametric linkage analysis showed chromosomes 13q32 (6) and 7q22 (13) to have genome wide significance as per Lander and Kruglyak's reporting guidelines (14). However, to date, definitive causal variants, pathways, or regions have yet to be identified. Genome-wide association studies (GWAS) investigated the association of common variants to SCZ in large populations (15, 16). A multi-stage GWAS of approximately 21000 cases and 40000 controls estimated that independent common variants contributing to SCZ risk account for 32% of variance in liability (16). A GWAS of 20,000 case-controls of European descent identified significant common variants in the microRNA-137 (15), an important regulator of genes in neuronal proliferation and synaptic maturation (17). Large GWAS studies using data obtained through meta-analyses have implicated SCZ to be associated with common single nucleotide variants on chromosome 6 in the major histocompatibility complex (18, 19). Rare copy number variants (CNVs), defined as a deletion or duplication of a chromosomal segment, have been implicated in SCZ risk (20-23). Rare CNV deletions of 1q21.1, 15q13.3, 22q11.21 and duplications of 16p11.2 have been implicated in SCZ (24, 25). Walsh et al. examined the whole genome of SCZ case controls for CNVs greater than 100 kilobases and found three times as many CNVs deleting or duplicating at least one gene in affected individuals (20). Genes within the CNVs of affected individuals were implicated in signalling during neurodevelopment. A whole genome scan found the burden of rare CNVs to be 1.15 times greater in affected individuals relative to controls (21).

Despite the observation of these variants in affected individuals, much of the heritability of SCZ remains to be explained. A portion of the missing heritability could be attributed to de novo mutations (DNMs), which would be undetected by linkage and GWAS studies. The World Health Organisation surveyed clinical centres from 12 countries and determined that the

incidence of SCZ is high (approximately 1%) and relatively constant despite variable environmental conditions (26). The incidence of SCZ remains constant despite negative selection pressures such as reduced reproductive fitness of affected individuals (27). Moreover, SCZ is associated with higher paternal age (28) likely due to DNMs occurring with spermatozoa cell divisions over time. In support of this, Awadalla et al. resequenced 401 synaptic genes and found a significant enrichment of pathogenic DNMs (primarily nonsynonymous mutations) in affected individuals from 143 SCZ trios and 142 ASD trios (29). To corroborate their findings, the rate of neutral point mutations that they directly calculated (1.36×10^{-8} mutations per site per generation) was not statistically different from previous indirect estimates. Whole exome sequencing (WES) has enabled the investigation of small variants, namely insertions and deletions (indels) and single nucleotide variants (SNVs). A WES study reported that the nonsynonmous:synonymous ratio of de novo SNVs in 231 affected families (4.84) was significantly greater than the corresponding ratio in control samples from the Simons Simplex Collection (2.23) (30). The number of de novo loss of function variants in probands from SCZ trios was 3 times higher than offspring from unaffected trios (30). Individuals that were more severely affected (as measured with school grades below the median) had an increased ratio of loss of function:missense variants compared to unaffected individuals and SCZ cases with higher grades (31). McCarthy et al. selected de novo SNVs with a minimum coverage of 10 reads that were known in the Exome Variant Server 6500 and 1000 Genomes and reported three times as many nonsense DNMs in SCZ trios relative to unaffected trios (32). Li et al. selected rare SNVs and indels with a minor allele frequency $< 0.1\%$ and determined that individuals affected with SCZ have a higher prevalence of exonic DNMs in schizophrenic probands compared to sibling controls (33). Similar to the findings pertaining to adult onset SCZ, an increased number of

exonic DNMs was reported in individuals affected with child onset SCZ diagnosed before 13 years of age (34). Moreover, the number of disruptive DNMs is higher compared to inherited variants in childhood onset SCZ. On the contrary, genes reported to be enriched for DNMs were also enriched for rare nonsynonymous alleles inherited from the parents of 623 trios (31). Several recent publications have reported an enrichment of DNMs in gene networks related to the glutamatergic postsynaptic proteins ARC and NMDAR (31, 35), FMRP (31, 36), chromatin remodeling (32), and neuronal calcium signaling (16). Other pathways implicated in SCZ based on an increased DNM rate in SCZ patients include voltage gated calcium channels, the NMDAR network, activity-regulated cytoskeleton, PSD-95 (37), synaptic genes (29), and a network of genes co-expressed in the prefrontal cortex of SCZ patients (38).

Autism Genetics

Autism spectrum disorder (ASD) includes autistic disorder, Asperger's disorder or pervasive developmental disorder and is mainly characterized by repetitive behaviours and deficits in social interaction and communication (1). The pathophysiology of ASD is unknown, but neuronal dysfunction particularly in the synapse through altered neuronal migration, abnormal synaptic pruning, or impaired synaptic function is believed to play a role (39-41). Other pathways, shown to be important for ASD include chromatin modifiers (42, 43), genes expressed in the embryo and genes associated with the fragile X protein (44, 45). ASD is a common disorder diagnosed in approximately 1 of 68 children (46) and is 4 times more prevalent in males relative to females (47). The heritability of ASD is estimated to be 70-90% (48). The disorder is genetically heterogeneous and specific etiologic factors in ASD can be identified in 30% - 40% of cases (the majority of which are mono-factorial etiologies) (48).

Similar to studies in SCZ, linkage studies have not successfully identified regions of major genetic contributions (49, 50). Findings from linkage studies are often not replicated at the level of genome wide significance with some exceptions supported by meta-analyses such as the chromosomal region 7q11-35 (51, 52). CNVs identified by karyotyping or array-based comparative genomic hybridisation, that have been reported to be implicated in ASD on more than one occasion include 1q21.1 deletions and duplications (53, 54), 15q11.2 deletion between break point 1 and break point 2 (55, 56), 16p11.2 deletion (57, 58) and 22q11.2 deletion (54). De novo CNVs also contribute to the risk of ASD (43, 59-61).

Current genetic studies have reported large numbers of candidate genes ranging from at least 100 to 1000 (41, 44, 45). Despite our uncertainty of the genetic architecture of ASD, recent WES studies have found that ASD patients have a significant enrichment of pathogenic nonsynonymous DNMs in 401 synaptic genes when analysing 142 ASD trios (29). Using a binomial test, they determined that the number of missense to nonsense DNMs was significantly higher than the neutral expectation (19.7:1) derived from the Human Gene Mutation Database (62). O'Roak et al. sequenced 20 ASD trios and selected DNMs that were known in dbSNP and 1000 Genomes (63). These variants were Sanger sequenced and verified in the Integrative genome browser. Protein altering mutations were significantly enriched at genomic regions that were highly conserved evolutionarily (as determined by Genomic evolutionary rate profiling scores (64) and Grantham matrix scores(65)). O'Roak et al. report DNMs found in probands from 200 ASD trios and list genes with recurrent protein-altering DNMs. Probands were found to have an increased nonsynonymous DNM rate relative to unaffected siblings (66). De novo gene disruptive mutations (nonsense, splice site, and frameshift) are twice as frequent in affected children compared to the unaffected siblings (44).

Neale et al., determined the rate of missense and nonsense DNMs in 175 ASD trios (0.92 mutations per exome) and found that it was not significantly different than the expected rate (0.871 mutations per exome). A statistically significant number of genes with DNMs encoded for proteins that had an enrichment of protein-protein interactions with a list of known ASD genes (67). Modelling of variants in these ASD families indicated that the DNMs attributed to the disease increase risk 5 to 20 fold but none were found to be completely penetrant. A WES study on 238 ASD quads determined that the burden of nonsense and splice site DNMs in brain-expressed genes was significantly increased in probands relative to unaffected siblings. The odds ratio (OR) of nonsense and splice site:silent SNVs considering sibpairs is significiant (OR = 5.65, p = 0.01) (68).

An observed correlation of higher paternal age and the number of observed DNMs per affected child (30, 44, 66, 69, 70) suggests a paternal bias in the origin of DNMs (44, 66, 67, 70, 71). Although DNMs are implicated in ASD, there is evidence to suggest the importance of *inherited* variants on increasing autism risk, with the emphasis of a maternal transmission bias to sons. Private inherited truncating SNVs in conserved genes (defined as genes with a residual variation intolerance score < 50) are in transmission disequilibrium with a significant maternal transmission bias to affected sons. The transmission disequilibrium was tested using Fisher's exact and Mann-Whitney U tests and by logistic regression (where the dependent variable was the presence of a variant found in a proband or sibling) (72). Genes with at least two de novo likely gene disruptive mutations were considered potential ASD risk genes (73). Likely gene disruptive variants in these genes are transmitted more frequently from mothers to affected children relative to unaffected siblings. Toma et al. identified genes with inherited rare truncating heterozygous variants present in both affected siblings of multiplex families to increase the

likelihood of these variants being implicated in ASD (74). In the identified genes, the number of inherited truncating mutations were two-fold higher than the number of de novo variants. Probands had a significant increase in the burden of rare inherited CNVs (72, 75, 76) relative to unaffected individuals.

Hypothesis and Objectives

The genetic architecture for ASD and SCZ is largely unknown. Based on genetic studies, it is suggested that polygeneticity is at work. The disease model to understand how these disorders are transmitted considers that every individual has some liability to develop the disease however the disease only occurs when an individual accumulates enough genetic and enviromental predispositions that increases the individual's liabiltiy to reach a critical disease threshold (CDT) (77). The advancement of WES technology and the ease of data sharing eliminated limitations that existed at the time the model was proposed in the 1960's. Griswold et al. resequenced genes previously identified by GWAS and analysed rare variants (41). They identified a significant increase in the number of genes with at least two loss of function variants in the same ASD individual relative to controls, which is suggestive of a multi-hit model. Based on these observations, we hypothesize that ASD and SCZ occurs when an individual reaches the CDT by accumulating rare variants (inherited and de novo) in a subset of genes from the susceptibilty 'network' that cooperatively increases susceptiblity to disease. Although not considered here, environmental threats occuring in utero or postnatally would be expected to also interact with genetic predispositions in the network to cause an individual to reach the CDT.

One of the objectives of this work is to identify SCZ susceptibility genes that are reported to be enriched in DNMs based on a Pubmed literature review. We will test if there is a significant enrichment of functional variants (including inherited and DNMs) of the "SCZ

network" (defined in this thesis as a set of susceptibility genes for SCZ) in affected individuals compared to unaffected individuals. We also aim to identify additional susceptibility genes based on genes responsible for single-gene conditions characterized by schizophrenia-like signs or symptoms or through an online tool, called Genemania, that recommends related genes based on functional data. Similar to SCZ, we will test if individuals affected with ASD are enriched in functional variants of the ASD susceptibility genes (i.e. ASD network) relative to their unaffected family members.

Chapter 2: Methods

Part 2.1: Candidate and control gene selection

Autism genes selection

To create a list of candidate genes to test for DNMs, we systematically searched Pubmed to identify genes with de novo single nucleotide variants (SNVs) in patients affected with ASD who had undergone WES. The Pubmed search used the following search terms: "exome*[Title/Abstract]" and "de novo[Title/Abstract]" and ("autism spectrum disorder" [Mesh] or "autistic disorder"[Mesh] or "autism[Title/Abstract]"). A total of 87 papers originally met these search criteria. Candidate genes were selected only if they were reported in at least 2 independent publications meeting our inclusion and exclusion criteria below.

Inclusion Criteria: Papers were included if they involved an original study using WES data (novel or obtained from a database) and the Pubmed search terms were used in the right context in the abstract (i.e. the paper was looking at DNMs and trying to understand the genes that are associated with autism). In the case of manuscripts describing autistic symptoms/features in patients with another condition (i.e. where autism was a comorbidity), the study was only included if **all** patients of the cohort had autistic symptoms/features. Case studies were included only if all individuals had undergone WES and all had ASD (as a primary condition or comorbidity).

Exclusion criteria: Papers that met any of the following conditions were removed: (1) the research question was not focused on autism (i.e. some of the patients included in the cohort did not have ASD but other psychiatric or neurodevelopmental disorders), (2) the paper was a review article, (3) the study was investigating inherited variants only, (4) the study was evaluating a specific number of genes (using targeted sequencing or deep re-sequencing), (5) the study was

evaluating protein expression (using RNA sequencing), (6) the study was using whole genome sequencing (i.e. the observed variants were not limited to protein coding regions of the genome).

We included papers that used WES, as opposed to whole genome sequencing, because the datasets available to us also used this technology and we were only interested in variants occurring in the protein coding regions of the genome.

Amongst papers that met the above inclusion and exclusion criteria, to the extent that it was possible based on the information available in each paper, the ones with original independent datasets were identified. A gene was selected only if it was reported in at least 2 publications using independent (i.e. not recurrent) datasets.

Schizophrenia genes selection

A similar search was conducted for SCZ using the key terms: "exome*[Title/Abstract]" and "de novo [Title/Abstract]" and ("schizophrenia"[Mesh] or "schizophrenia[Title/Abstract]" or "schizophrenia and other psychotic disorders"[Mesh]). A total of 33 papers originally met the search criteria. The same inclusion and exclusion criteria were applied, as described above for ASD, and candidate genes were selected if they were reported in at least 2 independent publications.

Control genes selection

A search was conducted in the Online Mendelian in Man (OMIM) database to identify genes associated with single-gene subtypes of hypertrophic cardiomyopathy (HC). The top 14 HC genes identified in OMIM (equal in number to the SCZ candidate genes) with no known

association with a neuropsychiatric phenotype were selected and served as control genes in our analyses (Supplementary Table 2.1).

Part 2.2: WES Datasets

The research ethics board of McGill University Health Centre approved this study. Approval to access the large SCZ and ASD WES datasets was obtained from the Database of Genotypes and Phenotypes (dbGaP) and the National Database of Autism Research (NDAR), respectively.

Schizophrenia WES data

First WES dataset, local trios: 35 local SCZ trios consisting of probands and unaffected parents from Dr. Rouleau's laboratory in affiliation with McGill University and Université de Montréal were obtained and analysed. These SCZ trios comprised individuals from previously published DNM studies using WES (34, 78). The dataset contained 20 early onset schizophrenic probands in which the age of onset was between the ages of 6-12 ($\bar{x}$ = 9.8 years), and 15 late onset schizophrenic probands ($\bar{x}_{male}$ = 18 years, $\bar{x}_{female}$ = 25 years) in which the age of onset was 12 years and above. Overall, 23 male probands (13 early onset and 10 late onset) and 12 female probands (7 early onset and 5 late onset) were studied.

Larger independent dataset, dbGaP trios: This dataset includes 623 SCZ families and is available in the dbGaP (study phs000687.v1.p1). The samples in this dataset were collected from the University Hospital Alexander in Sofia, Bulgaria and individuals with intellectual disability were excluded. We selected for unrelated families with parents who did not have schizophrenia. Since there were twelve families with two affected probands, we randomly selected 1 proband for each family. Overall, 598 trios were included in the analysis.

Larger independent dataset, dbGaP case-controls: 5090 individuals (2545 cases with SCZ and 2545 unrelated controls) were also analysed. This WES dataset was downloaded from the dbGaP (study phs000473.v1.p1), which included vcf files and additional data files.

Autism WES data

NDAR trios WES dataset: The data for 2,392 families (1,800 quads and 592 trios) with ASD was obtained from the NDAR (doi: *10.15154/1169318*; doi: *10.15154/1169195*). Raw data including VCF files and additional data file were submitted to NDAR by Krumm et al., 2015 (72).

Part 2.3: Variant annotation and variant filtering criteria

For each dataset, SNVs and small insertions/deletions (indels) were annotated by ANNOVAR according to the reference genome hg19/GRCh37 followed by segregation analysis. We selected rare (minor allele frequency ≤ 0.01) functionally important variants (herein defined as missense, frameshift, stop gain, stop loss, intronic and exonic splicing variants) with a genotype quality ≥ 90 in the candidate and control genes. If a variant did not meet any of these criteria it was considered wildtype.

Part 2.4 Statistical methods

Testing the de novo schizophrenia network

The mutation load (defined as the number of genetic variants meeting our filtering criteria) of the SCZ de novo 'network' and control 'network' was calculated for each individual in our three independent SCZ datasets (35 local SCZ trios, 598 dbGaP trios and 5090 dbGaP

case controls). To analyze the 35 local SCZ trios, a 1-tailed paired t-test was used to compare the mutation load of SCZ candidate genes in affected individuals versus the average mutation load in parents (see Supplementary Table 2.2 for details). The paired 1-tailed t-test analysis was repeated for the mutation load of the control genes using Statsplus (http://www.analystsoft.com/en/). Similar paired 1-tailed t-tests for the SCZ candidate genes and HC control genes respectively, were conducted using the larger dataset of 598 dbGaP trios. A homoscedastic 1-tailed t-test was used for the 5,090 case-control dataset, comparing the mutation load of SCZ candidate genes in affected individuals and external controls and this was repeated for the mutation load of the HC control genes. We applied reactome, an online tool that determines statistically likely pathways of a list of input genes (79), to the full SCZ candidate gene list to explore the functional importance of the candidate network.

Identifying additional SCZ genes using an OMIM search and Genemania

The SCZ susceptibility genes that were recurrently supported by the DNM literature, were inputted into Genemania (80), a bioinformatics tool that suggests genes that are related to an inputted list by assessing functional association data (such as co-expression and protein interaction), using the default settings. Genemania predicted genes to be functionally related to the inputted genes. The odds ratio (OR) for each of the Genemania suggested genes and the 95% confidence interval (CI) was calculated using rare variants meeting our filtering criteria in the dbGaP case control dataset. Only positions with a clear genotype were included in the calculation. For example, if the genotype was unknown (as per the genotype caller) due to a lack of information it was excluded from the calculation. The OR is defined as ad/bc and the confidence interval is calculated according to $ln(OR) \pm 1.96(1/a + 1/b + 1/c + 1/d)$; a = number of patients with mutation in the gene, b = number of controls with mutation in the gene, c =

number of patients without mutations in the gene, d = number of controls without mutations in the gene. Genemania suggested genes which had significant ORs (defined as OR > 1, lower threshold of the 95% CI > 1) were selected as additional SCZ susceptibility genes in the expanded network.

OMIM was searched for single-gene conditions characterized by SCZ or SCZ-like signs or symptoms using the following key terms: "schizo*" or "psychot*" or "hallucin*" or "psychosis". 163 OMIM entries originally met the search criteria. OMIM disorders were selected for analysis if there was a known phenotype description and/or molecular basis, if it was caused by point mutations in nuclear genes and if there were fewer than 10 associated genes. The OR for each of the above OMIM genes meeting our inclusion criteria was calculated using the dbGaP case control dataset, and those with significant ORs were inputted into Genemania. For each Genemania suggested gene, the OR and CI was calculated in order to select genes with significant ORs to be included in the expanded network.

To summarize, the expanded network included (1) the SCZ susceptibility genes with recurrent evidence for an enrichment of DNMs in affected individuals, (2) genes with significant ORs in our SCZ dataset and known to be responsible for single-gene conditions (identified through an OMIM search), and (3) Genemania suggested genes that were related to the aforementioned genes and had significant ORs. We performed a 1-tailed paired t-test using the dbGaP trios (an independent dataset relative to the one used to calculate the ORs when selecting additional genes) to compare the mutation load of the expanded set of SCZ susceptibility genes in probands versus the average mutation load in parents.

Testing the de novo autism network

Similar to SCZ, the mutational load of the selected 'ASD network' (as per the Pubmed search) and the 'HC control network' was calculated for each individual from NDAR. We also tested 65 ASD genes previously identified by Sanders et al. 2015 (81). To identify the genes, this group used a gene-based likelihood model called the transmission and de novo association test (TADA) (82), that incorporates de novo mutations, inherited variants and variants identified within cases and controls from WES data to identify risk genes. They analyzed three previously available ASD WES datasets by including SNVs, indels and evidence of association from genes within small de novo deletion CNVs in the TADA model (See Supplementary Table 2.3 for the list of TADA genes). A 1-tailed paired t-test was used to compare the mutation load of the 'ASD network' in affected individuals versus the average mutation load in parents. To analyze 1800 sibpairs, a paired 1-tailed t-test was conducted to compare the mutation load of ASD genes in probands versus the mutation load in unaffected siblings. This analysis was repeated for the HC control genes and the TADA genes.

Investigating a bias in the NDAR dataset

After reviewing the results of the analysis above using the NDAR dataset, we were interested to investigate if *unaffected* siblings were enriched in functional variants relative to probands across the whole exome or if this was specific to the preselected networks (i.e. de novo ASD network, TADA network and HC control network). Permutation analysis was performed using genes of the exome with at least 1 functional variant. The permutation analysis randomly selected 1000 sets of 143 genes, equal in number to the ASD susceptibility genes. Each iteration was a 1-tailed paired t-test comparing the mutation load in the permutated set of genes of unaffected siblings versus probands. The alternative hypothesis was that the difference in

mutation load between siblings and probands (sib-pro) was greater than 0. We were interested to explore whether an increased mutation load in unaffected siblings in the ASD susceptibility genes and throughout the exome could be due to differences in sequencing depth of functional variants between sibpairs. To investigate this further, we calculated the mean depth of sequencing considering all positions with functional variants in the ASD candidate genes for each proband with a sibling and represented the distribution in a histogram. This calculation was repeated for the siblings.

In an effort to control for any biases in the mutation load due to sequencing depth, we intended to further restrict our analysis to reduce false positives due to mapping errors by implementing a maximum threshold for sequencing depth and to include only functional variants that were covered sufficiently well by implementing a minimum threshold for sequencing depth. To decide the maximum threshold for sequencing depth of our restricted filtering criteria, we determined the mean depth of sequencing for each functional variant in the ASD genes considering the 1800 sibpairs and represented the distribution in a histogram. For each functional variant in the ASD genes, the mean sequencing depth was determined considering only probands with an unaffected sibling. In particular, for each functional variant the individual depth of sequencing in each proband was summed and the resulting value was divided by the number of individuals. The process was repeated independently for unaffected siblings. The mean depth per functional variant was plotted against the count of functional variants for each corresponding depth. The threshold for the maximum sequencing depth was determined as the 95^{th} percentile from the histogram described above. Functional variants with a minimum of 4 variant reads, total depth between 10 and 184 reads, inclusively, and a minimum of 20% variant reads out of the total were selected for analysis and the permutation was repeated.

Although we restricted our filtering criteria to include variants with a specified range of sequencing depth (and related parameters), we modeled the relationship between the mutation load of the ASD susceptibility genes and the odds to be affected controlling for the effects of sequencing depth amongst sibpairs using a conditional logistic regression. The first analysis considered functional variants meeting our original filtering criteria and the analysis was repeated using the restricted filtering criteria.

Supplementary Tables

Supplementary Table 2.1 Hypertrophic cardiomyopathy control genes

Gene	OMIM #
CALR3	611414
FLNC	102565
JPH2	605267
MYH6	160710
MYL2	160781
MYL3	160790
MYPN	608517
NEXN	613121
TCAP	607487
TNNC1	613243
TNNI3	191044
TNNT2	191045
TPM1	191010
VCL	613255

Supplementary Table 2.1 Fourteen control genes responsible for HC were selected using OMIM. These genes have no known association with neuropsychiatric phenotypes. The top HC genes identified in OMIM (equal in number to the SCZ candidate genes) were selected.

Abbreviations: HC: hypertrophic cardiomyopathy, OMIM: Online Mendelian Inheritance in Man, SCZ: schizophrenia

Supplementary Table 2.2 Analysis for family design

Family ID	Proband (A)	Mother (B)	Father (C)	Parental average (D)
0001	2	0	2	1
0002	1	0	1	0.5
0003	2	1	1	1
0004	1	0	0	0
0005	1	0	1	0.5
0006	3	1	1	1

Supplementary Table 2.2 The first column (Family ID) indicates the identification number for each of the families included in our study. The number of mutations in the disease network is listed in columns A, B and C for probands, mothers and fathers respectively. Column D indicates the average number of mutations present in the proposed network of both parents of any given proband. A paired 1-tailed t-test was performed comparing the mutation load of probands (column A) to the average mutation load in parents (column D).

Supplementary Table 2.3 ASD susceptibility genes previously identified using the transmission and de novo association test on WES data

Genes		
ACHE	*GRIN2B*	*POGZ*
ADNP	*ILF2*	*PTEN*
AKAP9	*INTS6*	*PTK7*
ANK2	*IRF2BPL*	*RANBP17*
APH1A	*KAT2B*	*SCN2A*
ARID1B	*KATNAL2*	*SETD5*
ASH1L	*KDM5B*	*SHANK2*
BCL11A	*KDM6B*	*SHANK3*
CAPN12	*KMT2C*	*SLC6A1*
CHD2	*KMT2E*	*SPAST*
CHD8	*MIB1*	*SUV420H1*
CTTNBP2	*MBD5*	*SYNGAP1*
CUL3	*MFRP*	*TBR1*
DIP2A	*MYT1L*	*TCF7L2*
DNMT3A	*NAA15*	*TNRC6B*
DSCAM	*NCKAP1*	*TRIO*
DYRK1A	*NINL*	*TRIP12*
ERBB2IP	*NLGN3*	*USP45*
ETFB	*NRXN1*	*WAC*
FOXP1	*OR52M1*	*WDFY3*
GABRB3	*P2RX5*	*ZNF559*
GIGYF1	*PHF2*	

Chapter 3: Results

Disease candidate gene sets

The Pubmed search we conducted identified 33 SCZ papers, 10 of which met our inclusion/exclusion criteria. In total, 14 SCZ candidate genes were selected; see Table 3.1 for the list of candidate genes and the referenced articles providing supporting evidence. Similarly, 87 ASD papers were generated by the Pubmed search, 26 of these met our inclusion/exclusion criteria. 143 ASD candidate genes were supported by 2 or more independent de novo studies using WES (Table 3.2).

De novo schizophrenia susceptibility network

There was no difference in mutational load of functional variants (SNVs and indels) in probands relative to parents in the SCZ candidate genes selected (Table 3.3, column A: $p = 0.43$) using the 35 local SCZ trios. However, in a larger independent dataset of 598 trios from dbGaP, there was a significant enrichment of functional variants (SNVs and indels) in the SCZ candidate genes of affected individuals (Table 3.3, column B: $p = 0.04$) but not in the HC control genes (Table 3.3, column B: $p = 0.23$). All of variants in the 'SCZ network' of dbGaP probands were inherited. Similarly, using the SCZ case-control dataset (n=5090) the mutation load in the SCZ susceptibility genes was significantly increased in affected individuals relative to external controls (Table 3.3, column C: $p = 0.02$) but not in the HC genes (Table 3.3, column C: $p = 0.09$). To better characterize the network functionally, we inputted our candidate genes in reactome, a tool that statistically determines the common pathways of genes. No pathways were statistically significant.

Expanded schizophrenia network

Affected individuals were significantly enriched in functional variants in the SCZ candidate genes (supported by at least 2 independent DNM publications) in the 598 dbGaP trios dataset and the 5090 dbGaP case control dataset. Genemania suggested 20 genes that were functionally related to the above network (Supplementary Table 3.4); 3 of which had significant ORs (*CPNE7*: OR = 1.14 [1.10 - 1.17], *NT5DC3*: OR = 1.11 [1.06 – 1.16], *PPP1R12B*: OR = 1.56 [1.39 – 1.74]) and were included in the expanded network.

The OMIM search for genes responsible for single-gene conditions characterized by SCZ-like signs or symptoms generated 163 entries. 121 OMIM entries (and 151 genes) met our inclusion criteria (Supplementary Table 3.5). 29 OMIM genes had significant ORs (OR > 1, lower threshold of the 95% CI > 1) in our SCZ dataset. 20 genemania suggested genes were related to the significant OMIM network above (Supplementary Table 3.6); 1 of which had a significant OR (*CP*: OR = 1.05 [1.03 – 1.07]). The expanded SCZ network includes 47 susceptibility genes with ORs ranging from 1.05 to 2.81 (Table 3.4). There was no difference in the mutation load of the expanded SCZ network between probands and the average mutation load in parents using the dbGaP trios ($p = 0.19$).

Increased mutation load in unaffected siblings relative to probands in the autism dataset

There was no difference in the mutational load of the ASD selected genes when comparing probands to the average mutation load of the parents using a 1-tailed paired t-test (Table 3.5, column A: $p = 0.08$). *Unaffected* siblings were significantly enriched in functional variants relative to probands in the selected ASD network (Table 3.5, column B: $p = 0.02$) and the TADA network (Table 3.5, column B: $p = 0.008$) when analyzing the 1800 sibpairs from

NDAR. Similar results were found for the HC control genes (Table 3.5, column B: $p = 7.36x10^{-6}$). Through the permutation analysis, we found a significant bias of an increased number of functional variants in unaffected siblings relative to probands throughout the exome (Figure 3.2A). On average a proband had a total depth of sequencing of 96 reads which was slightly greater than the average total depth of sequencing per sibling of 94 reads (Figure 3.3A and 3.3B). Even when we applied the stricter filtering criteria for depth of sequencing and repeated the permutation analysis, unaffected siblings were enriched for functional variants relative to probands throughout the exome (Figure 3.2B). A positive significant relationship between depth of sequencing and the odds to be affected (OR = 1.0049 [1.0013, 1.0086], p = 0.008) and a negative significant relationship between mutation load in the ASD network and the odds to be affected (OR = 0.9893 [0.9798, 0.998], p = 0.046) was indicated by our conditional logistic regression considering variants in the ASD network that met our original criteria in sibpairs (Table 3.6A). The mean depth of sequencing for each functional variant in the ASD genes considering the 1800 sibpairs is represented in Figure 3.1 (on average each functional variant was covered with 75 reads in probands and siblings). The x-axis indicates the mean depth for each functional variant and the y-axis indicates the count of functional variants that were covered with the corresponding mean depth of sequencing on the x-axis. The threshold for the maximum sequencing depth was determined as the 95th percentile of the distribution in Figure 3.1, which was equal to 184 reads. When we repeated the conditional logistic regression using the stricter filtering criteria selecting variants within a range of sequencing depth, we find that there is no significant relationship between the mutation load in the ASD network and the odds to be affected when controlling for depth of sequencing (p = 0.1358, see Table 3.6B).

Tables

Table 3.1 SCZ susceptibility genes reported in two or more DNM studies using WES

Genes	References	PMID	Number of independent references
AHDC1 *CHRNG* *SETD1A*	Xu *et al*., 2012 Guipponi *et al*., 2014 Takata *et al*., 2016	23042115 25420024 26938441	2
ANKRD11 *DNAH12*	Xu *et al*., 2012 Wang *et al*., 2015 Takata *et al*., 2016	23042115 26666178 26938441	2
CHD4	Girard *et al*., 2011 Li *et al*., 2016	21743468 25849321	2
DICER1	Wang *et al*., 2015 Li *et al*., 2016	26666178 25849321	2
GPR153	Xu *et al*., 2011 Xu *et al*., 2012 Ambalavanan *et al*., 2016	21822266 23042115 26508570	2
LRP1	Girard *et al*., 2011 Wang *et al*., 2015 Li *et al*., 2016	21743468 26666178 25849321	3
MACF1	Xu *et al*., 2012 Wang *et al*., 2015	23042115 26666178	2
PITPNM1	Xu *et al*., 2011 Xu *et al*., 2012 McCarthy *et al*., 2014	21822266 23042115 24776741	2
RGS12	Xu *et al*., 2011 Xu *et al*., 2012 Guipponi *et al*., 2014	21822266 23042115 25420024	2
RYR2	Xu *et al*., 2012 Takata *et al*., 2016 Ambalavanan *et al*., 2016	23042115 26938441 26508570	2
STAC2	Xu *et al*., 2012 Ambalavanan *et al*., 2016	23042115 26508570	2

Table 3.1 Fourteen candidate genes preselected for the SCZ susceptibility network includes genes reported to have increased DNMs in at least two or more WES studies. Dependent studies denoted in grey shared samples and were not considered indpendent: namely Xu *et al*., 2011, Xu *et al*., 2012, and Takata *et al*., 2016.

Abbreviations: SCZ: schizophrenia, DNM: de novo mutation, WES: whole exome sequencing

Table 3.2 ASD susceptibility genes reported in two or more DNM studies using WES

Gene	Reference	PMID	Number of independent references
ABCA12 *ABCA13* *ADAM22* *AHNAK2* *AOC3* *C10orf90* *CADPS* *CDCA7L* *CHD1* *CYTH4* *DCAF4* *DNAH7* *ELK1* *FAT1* *FN1* *JMJD1C* *KIAA1967* *KIRREL3* *KRTAP9-3* *MACC1* *MANSC1* *MCM2* *MCPH1* *MKL2* *MPDZ* *MUC4* *MYCBP2* *MYO5B* *MYOM2* *NFASC* *NISCH* *PCDHB16* *PIAS1* *PIWIL4* *PLCD4* *PLXNB1* *POLR2A* *RTN4RL1* *SBF1* *SCRIB* *SLC39A5* *SLC6A3*	Neale *et al*., 2012 Iossifov *et al*., 2014	22495311 25363768	2

SNTG1 *SPAST* *STAT2* *SVIL* *SYNE2* *TCF3* *TDRD5* *UGT2B10* *USH2A* *VAV3* *WDR66* *ZFC3H1* *ZNF155*			
ASXL3	Dinwiddie *et al*., 2013 De Rubeis *et al*., 2014 Hori *et al*., 2016	24044690 25363760 27075689	3
ABCA3 *DDX50* *DNAJC16* *ECSIT* *ENO3* *GCN1L1* *HERC1* *LRP1B* *OBSCN* *PCNX* *PLEKHA4* *POLR1E* *WDR20* *ZFHX3* *ZFPM2*	Iossifov *et al*., 2014 Hashimoto *et al*., 2016	25363768 26582266	2
ABI3BP *CDKL3* *EIF3G* *EPAS1* *GTF3C2* *JAK2* *KIAA0226* *METTL2B* *PCDH15* *SETD7* *SNAPC5* *UBR4* *UNC79* *XRCC5*	An *et al*., 2014 Iossifov *et al*., 2014	24893065 25363768	2
XRN2	An *et al*., 2014	24893065	2

	Iossifov *et al*., 2014	25363768	
AFF4 *ARHGAP15* *CHST5* *DCTN5* *EPHB6* *GPR139* *IL1R2* *MYO1A* *PLCD1* *RBMS3* *SLC30A5* *SYNE1* *TGM3* *TTN* *XIRP1*	O'Roak *et al*., 2011 Iossifov *et al*., 2014	21572417 25363768	2
ANK3	Bi *et al*., 2012 Iossifov *et al*., 2014	22865819 25363768	2
ASH1L *SYNGAP1*	De Rubeis *et al*., 2014 Dong *et al*., 2014 Iossifov *et al*., 2014	25363760 25284784 25363768	2
BICC1	Iossifov *et al*., 2014 Lee *et al*., 2014	25363768 24501278	2
BRD3 *DICER1* *H2AFV* *ITPR1* *LPHN2* *LRP1* *MOV10* *MYH9* *NEB* *RFX3* *SLC6A1* *TECTA*	Iossifov *et al*., 2014 Li *et al*., 2016	25363768 25849321	2
CACNA1E	Neale *et al*., 2012 O'Roak *et al*., 2012 Iossifov *et al*., 2014	22495311 22495309 25363768	2
CHD2	Neale *et al*., 2012 Dong *et al*., 2014 Iossifov *et al*., 2014 Pinto *et al*., 2016	22495311 25284784 25363768 26754451	3
CREBBP	Iossifov *et al*., 2014 Yoo *et al*., 2015	25363768 25768348	2
	O'Roak *et al*., 2012 De Rubeis *et al*., 2014	22495309 25363760	

CUL3	Iossifov *et al*., 2014 Codina-Solà *et al*., 2015 Li *et al*., 2016	25363768 25969726 25849321	2
EEF1A2	Iossifov *et al*., 2014 Nakajima *et al*., 2014	25363768 24697219	2
FOXP1	O'Roak *et al*., 2011 Iossifov *et al*., 2014 Lozano *et al*., 2015	21572417 25363768 25853299	3
GABRB3	De Rubeis *et al*., 2014 Iossifov *et al*., 2014 Li *et al*., 2016	25363760 25363768 25849321	3
ETFB *MYO9B* *MYT1L* *TRIO*	De Rubeis *et al*., 2014 Iossifov *et al*., 2014	25363760 25363768	2
HIVEP3 *MYH10*	Dong *et al*., 2014 Iossifov *et al*., 2014 Li *et al*., 2016	25284784 25363768 25849321	2
MED13L	Iossifov *et al*., 2012 Iossifov *et al*., 2014 Codina-Solà *et al*., 2015	22542183 25363768 25969726	2
MYOF *SMARCC2* *TUBA1A*	Neale *et al*., 2012 Iossifov *et al*., 2014 Li *et al*., 2016	22495311 25363768 25849321	2
POGZ	Iossifov *et al*., 2012 Neale *et al*., 2012 De Rubeis *et al*., 2014 Iossifov *et al*., 2014 Fukai *et al*., 2015 Hashimoto *et al*., 2016 Li *et al*., 2016	22542183 22495311 25363760 25363768 25694107 26582266 25849321	4
PTEN	O'Roak *et al*., 2012 De Rubeis *et al*., 2014 Iossifov *et al*., 2014 Codina-Solà *et al*., 2015 Li *et al*., 2016	22495309 25363760 25363768 25969726 25849321	2
RELN	Neale *et al*., 2012 De Rubeis *et al*., 2014 Iossifov *et al*., 2014 Li *et al*., 2016	22495311 25363760 25363768 25849321	2
SCN2A	Sanders *et al*., 2012 An *et al*., 2014 De Rubeis *et al*., 2014 Iossifov *et al*., 2014 Tavassoli *et al*., 2014 Codina-Solà *et al*., 2015	22495306 24893065 25363760 25363768 24650168 25969726	3

	Li *et al*., 2016	25849321	
SETD2	O'Roak *et al*., 2012 Iossifov *et al*., 2014 Lumish *et al*., 2015	22495309 25363768 26084711	2
SETD5	Neale *et al*., 2012 De Rubeis *et al*., 2014 Iossifov *et al*., 2014 Li *et al*., 2016	22495311 25363760 25363768 25849321	2
TBL1XR1	O'Roak *et al*., 2012 Iossifov *et al*., 2014 Saitsu *et al*., 2014	22495309 25363768 25102098	2
TBR1	Neale *et al*., 2012 O'Roak *et al*., 2012 De Rubeis *et al*., 2014 Dong *et al*., 2014 Iossifov *et al*., 2014 Li *et al*., 2016	22495311 22495309 25363760 25284784 25363768 25849321	2
TLK2	O'Roak *et al*., 2011 Iossifov *et al*., 2014 Li *et al*., 2016	21572417 25363768 25849321	2

Table 3.2 143 candidate genes preselected for the ASD susceptibility network includes genes reported to have increased DNMs in at least two or more WES studies. Dependent studies denoted in grey shared samples and were not considered independent. Genes in bold are supported by 3 or more independent publications.

Abbreviations: ASD: autism spectrum disorder, DNM: de novo mutation, WES: whole exome sequencing.

Table 3.3 Comparisons of the mutation load in individuals affected with SCZ versus unaffected individuals (familial and external controls)

		A	B		C	
i	**SCZ dataset**	**35 local trios**	**dbGaP trios**		**dbGaP case control**	
ii	**Sample size**	35 probands 35 mothers 35 fathers	598 probands 598 mothers 598 fathers		2545 cases 2545 controls	
iii	**Type of test**	1-tailed paired t-test	1-tailed paired t-test		1-tailed homoscedastic t-test	
iv	**Comparison group**	Probands vs. parental average	Probands vs. parental average		Cases vs. controls	
v	**Gene network**	**SCZ**	**SCZ**	**HC**	**SCZ**	**HC**
vi	**P-value**	0.43	0.04	0.23	0.02	0.09

Table 3.3 T-test comparing the mutation load in a set of preselected susceptibility genes (albeit the SCZ network or the HC control network) between affected and unaffected individuals using A) the 35 local trios B) 598 dbGaP trios C) 5090 case controls. There was no difference in the mutation load of probands relative to the average mutation load in parents (A). Affected individuals were significantly enriched for functional variants in the SCZ genes but not in the HC control genes relative to unaffected individuals (B and C).

Abbreviations: SCZ: schizophrenia, HC: hypertrophic cardiomyopathy, dbGaP: database of Genotypes and Phenotypes

Table 3.4 Additional SCZ susceptibility genes

Gene	OR	OR confidence interval	Total muts in cases per gene	Total muts in controls per gene
TP53	2.81	1.65-4.79	14	5
CPOX	1.84	1.43-2.37	22	12
A2M	1.69	1.22-2.34	71	67
MED12	1.66	1.33-2.09	32	18
BCKDHA	1.64	1.23-2.19	18	11
PRDM8	1.63	1.13-2.33	14	9
FGFR2	1.59	1.21-2.08	19	12
PPP1R12B	1.56	1.39-1.74	96	75
MMACHC	1.53	1.27-1.86	26	17
ROGDI	1.38	1.29-1.48	71	51
NAGS	1.38	1.01-1.88	15	12
TPH2	1.37	1-1.86	15	11
PAK3	1.30	1.06-1.6	23	18
KCNN4	1.29	1.09-1.52	27	21
GLUD2	1.28	1.05-1.56	31	24
PINK1	1.28	1.2-1.36	87	64
CACNA1A	1.26	1.22-1.29	158	137
PTPN22	1.23	1.2-1.27	164	136
PLAU	1.21	1.17-1.26	113	96
MAN2B1	1.20	1.15-1.25	138	118
ATP2A2	1.16	1.05-1.28	46	39
PPT1	1.15	1.03-1.28	39	34
DMPK	1.14	1.05-1.24	51	45
GAL	1.14	1.05-1.25	49	43
ZBTB20	1.14	1.08-1.21	78	66
CPNE7	1.14	1.1-1.17	144	129
MPO	1.12	1.09-1.14	216	199
NT5DC3	1.11	1.06-1.16	95	84
ZFYVE26	1.11	1.09-1.12	392	347
PDGFB	1.10	1.03-1.17	67	62
ECM1	1.09	1.07-1.12	204	187
NPC1	1.08	1.04-1.12	116	105
CP	1.05	1.03-1.07	236	229

Table 3.4 Thirty three genes with significant ORs were selected in the expanded SCZ susceptbility network. Twenty nine genes responsible for single-gene conditions characterized by schizophrenia-like signs or symptoms and four genes predicted to be functionally related to the

previously mentioined genes by Genemania. The expanded SCZ network also includes genes reported 2 or more DNMs studies using WES (Table 1).
Abbreviations: OR: odds ratio, SCZ: schizophrenia, muts: mutations

Table 3.5 Comparisons of the mutation load in probands with ASD versus unaffected family members (parents and unaffected siblings)

		A	B		
i	**ASD dataset**	**NDAR trios**	**NDAR sibpairs**		
ii	**Sample size**	2392 probands 2392 mothers 2392 fathers	1800 probands 1800 siblings		
iii	**Type of test**	1-tailed paired t-test	1-tailed paired t-test		
iv	**Comparison group**	Probands vs. parental average	Probands vs. siblings		
v	**Gene network**	**ASD**	**ASD**	**TADA**	**HC**
vi	**P-value**	0.08	0.02*	0.008*	7.36×10^{-6}*

Table 3.5 There was no difference in the mutation load of ASD susceptibility genes when comparing probands to the average mutation load in parents (A). * Unaffected siblings were significantly enriched for functional variants relative to probands in the ASD network, the TADA network and the HC control network, respectively.

Abbreviations: ASD: autism spectrum disorder, HC: hypertrophic cardiomyopathy, NDAR: National Database of Autism Research, TADA: transmission and de novo association

Table 3.6A Association of sequencing depth and mutation load with disease outcome using standard filtering criteria for functional variants

Model	**OR [95% confidence interval]**	**P-value**
Disease ~ Depth	1.005 [1.001, 1.009]	0.007
Disease ~ Number of muts	0.989 [0.978, 0.999]	0.039
Disease ~ Depth + Number of muts	OR_{depth} = 1.005 [1.001, 1.009] $OR_{mutations}$ = 0.989 [0.979, 0.998]	p_{depth} = 0.008 $p_{mutations}$ = 0.046

Table 3.6B Association of sequencing depth and mutation load with disease outcome using stricter filtering criteria for functional variants

Model	**OR [95% Confidence interval]**	**P-value**
Disease ~ Depth	1.004 [1, 1.009]	0.045
Disease ~ Number of muts	0.9892 [0.9755, 1.003]	0.13
Disease ~ Depth + Number of muts	OR_{depth} = 1.005 [1.0002, 1.009] $OR_{mutations}$ = 0.989 [0.9755, 1.003]	p_{depth} = 0.039 $p_{mutations}$ = 0.136

Table 3.6A Mutation load significantly decreases the odds to be affected when controlling for the effect of depth of sequencing in the conditional logistic regressions using ASD sibparis. Unaffected siblings have more mutations than probands. **Table 3.6B** There is no relationship between mutation load and the odds to be affected when applying the stricter filtering criteria and statistically controlling for the effect of depth of sequencing.

Abbreviations: ASD: autism spectrum disorder, OR: odds ratio

Figures

Figure 3.1: Mean depth of sequencing per functional variant in sibpairs.

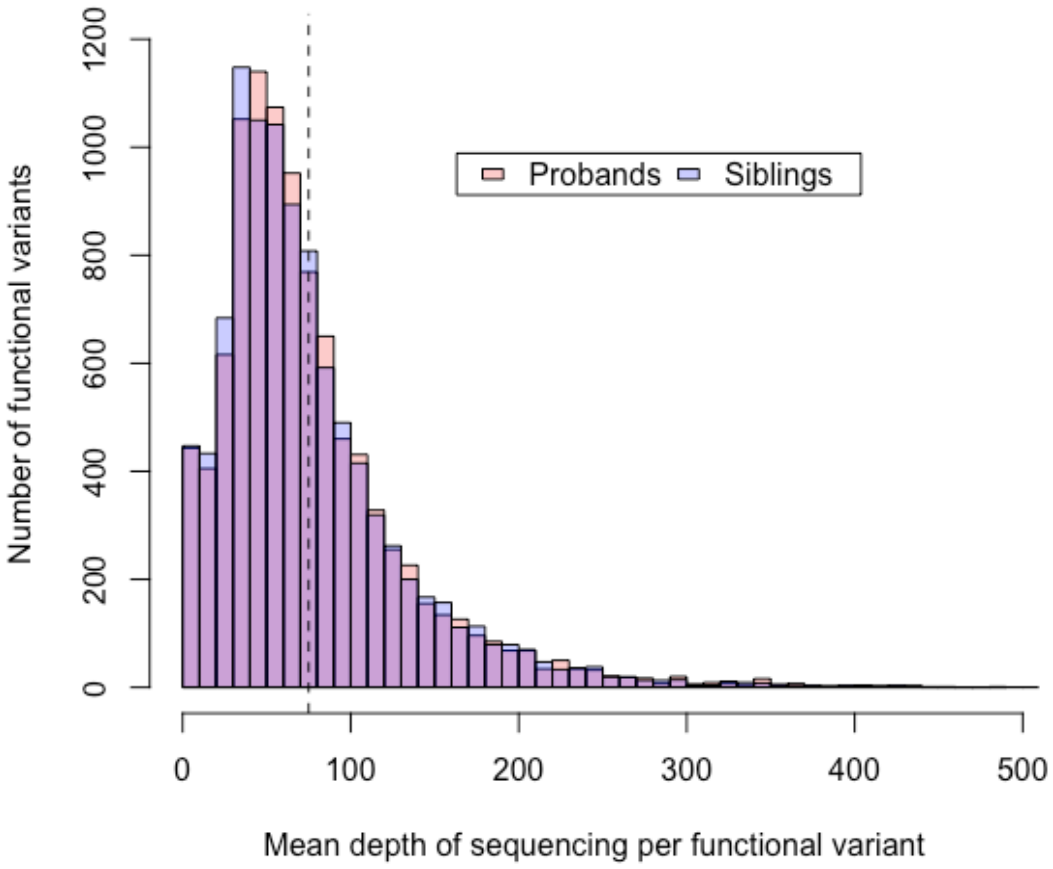


Figure 3.1 Several functional variants found in the ASD network have a very low or very high mean depth of sequencing in sibpairs, which is indicative of potential false positives. The 95th percentile of the distribution is 184 reads.

Figure 3.2A P-value distribution of the permutation analysis using NDAR sibpairs

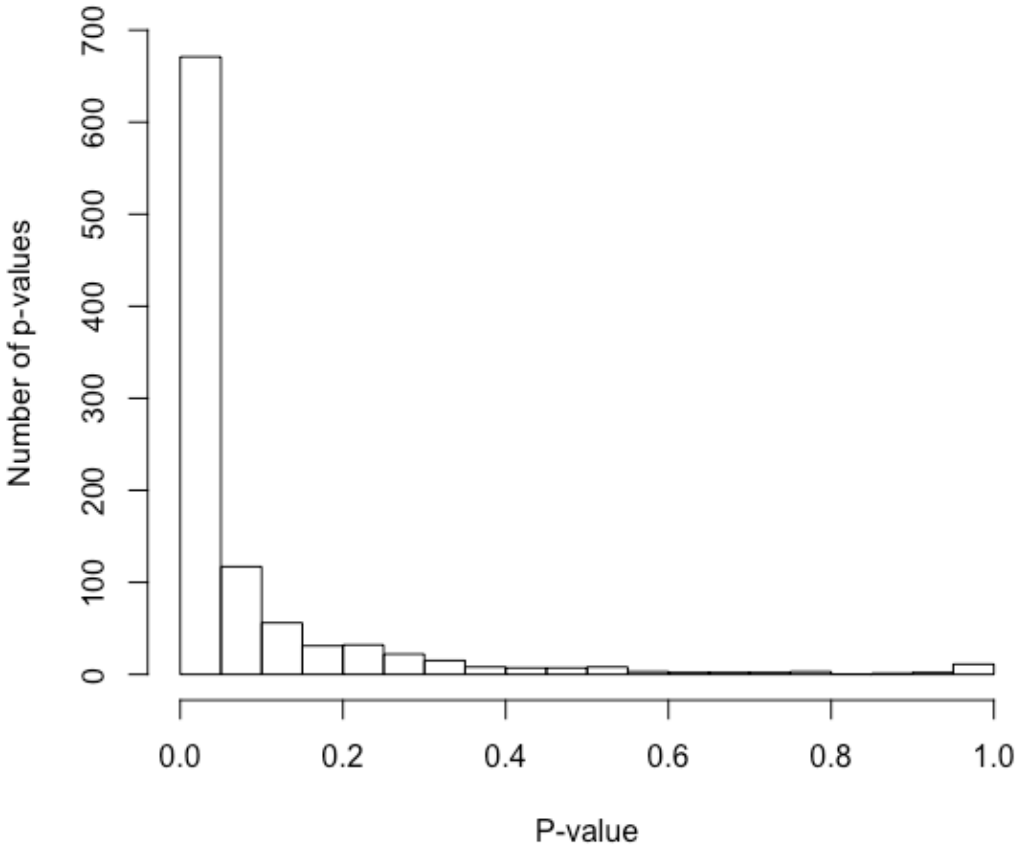


Figure 3.2A The permutation analysis randomly selected 1000 sets of 143 genes, equal in number to the ASD genes. Each iteration was a 1-tailed paired t-test comparing the mutation load in the permutated set of genes in unaffected siblings versus probands (H_A = mutation load in siblings > mutation load in probands). Unaffected siblings were enriched in functional variants throughout the exome relative to probands.

Abbreviations: H_A = Alternative hypothesis, NDAR = National Database of Autism Research

Figure 3.2B Repeated permutation analysis after implementing stricter filtering criteria for sequencing depth

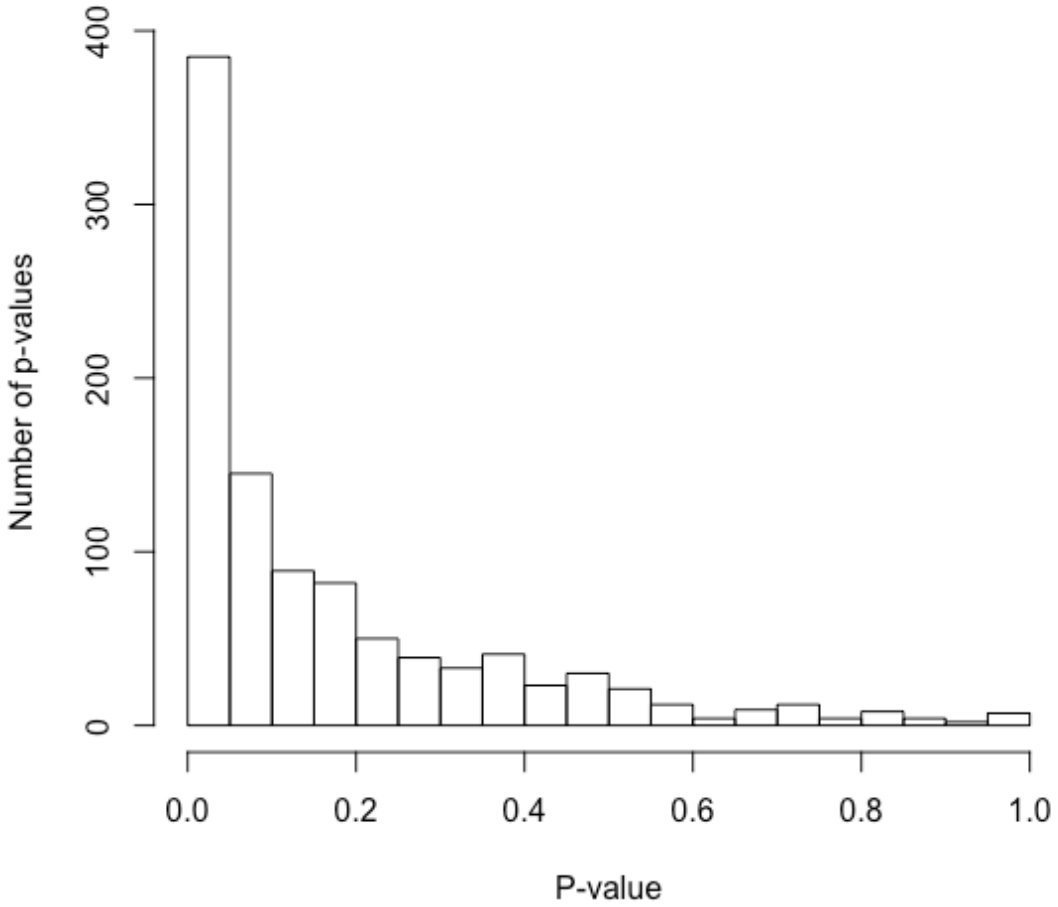


Figure 3.2B Stricter filtering criteria for depth of sequencing were applied to select functional variants and the permutation analysis was repeated. The distribution of p-values was less skewed but many p-values were still significant, indicating that unaffected siblings had more mutations than probands throughout the exome.

Figure 3.3A The mean depth of sequencing per functional variant in probands

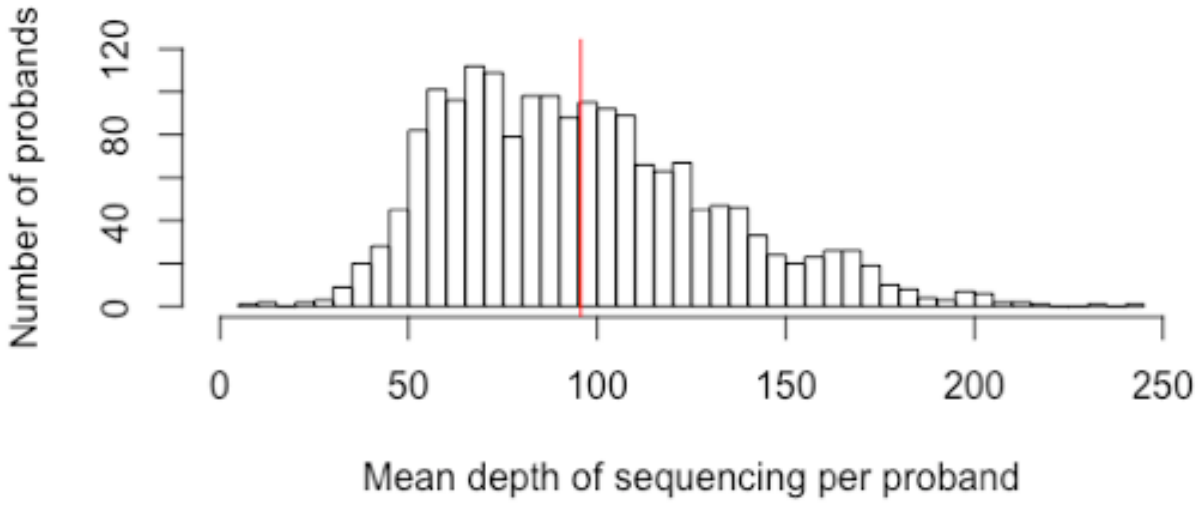


Figure 3.3B The mean depth of sequencing per functional variant in siblings

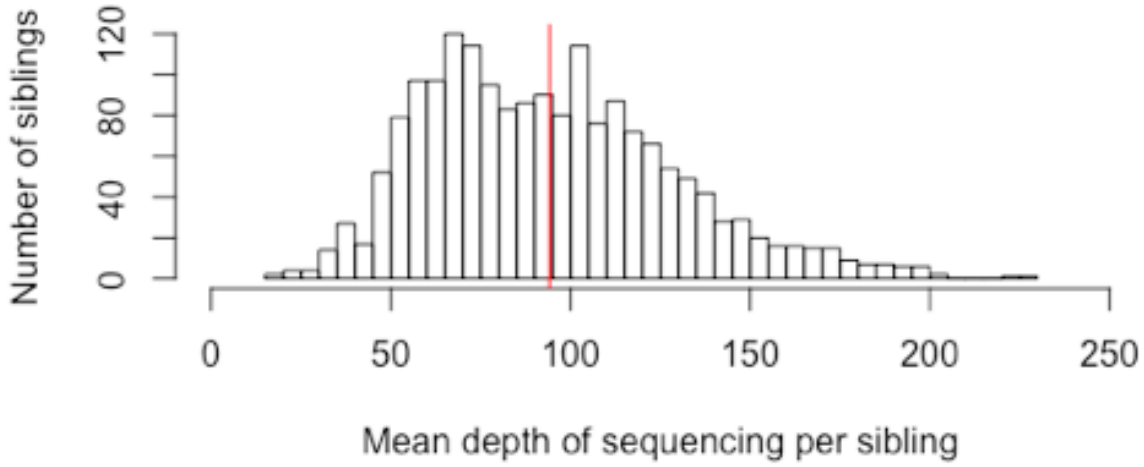


Figure 3.3A and 3.3B On average, a proband had a mean sequencing depth of 96 reads and a sibling had a mean sequencing depth of 94 reads considering all positions with functional variants (red lines).

Supplementary Tables

Supplementary Table 3.4 Genemania genes related to the de novo SCZ susceptibility network

Genes
AP3D1
CACNA1G
CFAP157
CLSTN3
CPNE7
DNAH9
DTHD1
HGS
KIAA1644
LRRC18
NCKAP5
NT5DC3
PAF1
PANX2
POM121C
PPP1R12B
PRPF6
PXK
RBM10
SMARCA4

Supplementary Table 3.4 Genemania suggested 20 genes that were related to the de novo SCZ network (see Table 1). The OR for each gene was calculated using the dbGaP case control dataset. Genes in bold had significant ORs and were included in the expanded susceptibility network.

Abbreviations: dbGaP: Database of Genotypes and Phenotypes, OR: odds ratio, SCZ: schizophrenia

Supplementary Table 3.5. Single-gene conditions characterized by SCZ-like signs or symptoms

Gene Name	Disease (OMIM #)
A2M	Alzheimer disease (OMIM #104300)
ABCD1	Adrenoleukodystrophy (OMIM #300100)
ACADS	Deficiency of acyl-coa dehydrogenase (OMIM #201470)
ADH1C	Late-onset parkinson disease (OMIM #168600)
AIP	Pituitary adenoma (OMIM #219090)
ALDH5A1	Succinic semialdehyde dehydrogenase deficiency (OMIM #271980)
APP	Alzheimer disease (OMIM #104300)
ARSA	Metachromatic leukodystrophy (OMIM #250100)
ASS1	Classic citrullinemia (OMIM #215700)
ATP13A2	Kufor-rakeb syndrome (OMIM #606693)
ATP1A3	Dystonia 12 (OMIM #128235)
ATP2A2	Darier-white disease (OMIM #124200)
ATXN1	Spinocerebellar ataxia 1 (OMIM #164400)
ATXN2	Late-onset parkinson disease (OMIM #168600)
BCKDHA	Maple syrup urine disease (OMIM #248600)
BCKDHB	Maple syrup urine disease (OMIM #248600)
BCKDK	Branched-chain keto acid dehydrogenase kinase deficiency (OMIM #614923)
C10orf2	Mitochondrial dna depletion syndrome 7 (OMIM #271245)
CACNA1A	Migraine, familial hemiplegic (OMIM #141500)
CFH	Complement factor h deficiency (OMIM #609814)
CHD8	Susceptibility to autism 18 (OMIM #615032)
CHEK2	Li-fraumeni syndrome 2 (OMIM #609265)
CLCN2	Leukoencephalopathy with ataxia (OMIM #615651)
CLN3	Ceroid lipofuscinosis (OMIM #204200)
CLN6	Ceroid lipofuscinosis (OMIM #204300)
CPOX	Coproporphyria (OMIM #121300)
CSTB	Myoclonic epilepsy of unverricht and lundborg (OMIM #254800)
CTLA4	Systemic lupus erythematosus (OMIM #152700)
CYP27A1	Cerebrotendinous xanthomatosis (OMIM #213700)
DAX1	Adrenal hypoplasia (OMIM #300200)
DBT	Maple syrup urine disease (OMIM #248600)
DCAF17	Woodhouse-sakati syndrome (OMIM #241080)

DEPDC5	Epilepsy (OMIM #604364)
DIRC2	Renal cell carcinoma (OMIM #144700)
DJ1	Parkinson disease 7 (OMIM #606324)
DMD	Muscular dystrophy (OMIM #300376)
DMPK	Myotonic dystrophy 1 (OMIM #160900)
DNAJC5	Ceroid lipofuscinosis (OMIM #162350)
DNASE1	Systemic lupus erythematosus (OMIM #152700)
DNMT1	Cerebellar ataxia (OMIM #604121)
DSTYK	Congenital anomalies of kidney and urinary tract 1(OMIM #610805)
ECM1	Lipoid proteinosis of urbach and wiethe (OMIM #247100)
EPM2A	Myoclonic epilepsy of lafora (OMIM #254780)
FCGR2A	Systemic lupus erythematosus (OMIM #152700)
FCGR2B	Systemic lupus erythematosus (OMIM #152700)
FGFR2	Saethre-chotzen syndrome (OMIM #101400)
FIG4	Polymicrogyria, bilateral temporooccipital (OMIM #612691)
FKBP5	Major depressive disorder (OMIM #608516)
FLCN	Renal cell carcinoma (OMIM #144700)
FTL	Neurodegeneration with brain iron accumulation 3 (OMIM #606159)
GAL	Epilepsy (OMIM #616461)
GATA6	Conotruncal heart malformations (OMIM #217095)
GBA	Dementia (OMIM #127750) Late-onset parkinson disease (OMIM #168600)
GDF1	Conotruncal heart malformations (OMIM #217095)
GLUD2	Late-onset parkinson disease (OMIM #168600)
GNAS	Pseudohypoparathyroidism, type Ia (OMIM #103580) Acth-independent macronodular adrenal hyperplasia (OMIM #219080)
GRN	Frontotemporal lobar degeneration with tdp43 inclusions (OMIM #607485)
GSS	Glutathione synthetase deficiency (OMIM #266130)
HARS	Usher syndrome type IIIb (OMIM #614504)
HCRT	Narcolepsy 1 (OMIM #161400)
HEXA	Tay-sachs disease (OMIM #272800)
HEXB	Sandhoff disease (OMIM #268800)
HFE	Alzheimer disease (OMIM #104300) Porphyria variegate (OMIM #176200)
HLA-DQB1	Creutzfeldt-jakob disease (OMIM #123400)

HMBS	Porphyria (OMIM #176000)
HNF1A	Renal cell carcinoma (OMIM #144700)
HNF1B	Renal cell carcinoma (OMIM #144700)
HSD17B10	Mental retardation, x-linked, syndromic 10 (OMIM #300220)
HTR2A	Major depressive disorder (OMIM #608516)
IDUA	Hurler-scheie syndrome (OMIM #607015)
ITM2B	Cerebral amyloid angiopathy (OMIM #117300)
JPH3	Huntington disease-like 2 (OMIM #606438)
KCNN4	Dehydrated hereditary stomatocytosis 2 (OMIM #616689)
KCNT1	Epilepsy (OMIM #615005)
LGI1	Epilepsy, familial temporal lobe (OMIM #600512)
MAN2B1	Mannosidosis (OMIM #248500)
MAOA	Brunner syndrome (OMIM #300615)
MAPT	Frontotemporal dementia (OMIM #600274) Late-onset parkinson disease (OMIM #168600)
MDD1	Major depressive disorder (OMIM #608516)
MDD2	Major depressive disorder (OMIM #608516)
MECP2	Mental retardation, x-linked, syndromic 13 (OMIM #300055) Mental retardation, x-linked, syndromic 13 (OMIM #300055)
MED12	Lujan-fryns syndrome (OMIM #309520)
MMACHC	Methylmalonic aciduria and homocystinuria (OMIM #277400)
MPO	Alzheimer disease (OMIM #104300)
MTHFR	Homocystinuria (OMIM #236250)
MUC1	Medullary cystic kidney disease 1 (OMIM #174000)
MYO7A	Usher syndrome, type I (OMIM #276900)
NAGS	N-acetylglutamate synthase deficiency (OMIM #237310)
NDP	Norrie disease (OMIM #310600)
NHLRC1	Myoclonic epilepsy of lafora (OMIM #254780)
NKX2-5	Conotruncal heart malformations (OMIM #217095)
NKX2-6	Conotruncal heart malformations (OMIM #217095)
NOS3	Alzheimer disease (OMIM #104300)
NOTCH3	Cerebral arteriopathy (OMIM #125310)
NPC1	Niemann-pick disease, type c1 (OMIM #257220)
NPC2	Niemann-pick disease, type c2 (OMIM #607625)
OGG1	Renal cell carcinoma (OMIM #144700)
PAH	Phenylketonuria (OMIM #261600)
PAK3	Mental retardation (OMIM #300558)
PCDH19	Epileptic encephalopathy (OMIM #300088)

PDE11A	Pigmented nodular adrenocortical disease (OMIM #610475)
PDGFB	Basal ganglia calcification (OMIM #615483)
PDGFRB	Kosaki overgrowth syndrome (OMIM #616592)
PINK1	Early onset parkinson disease 6 (OMIM #605909)
PLAU	Alzheimer disease (OMIM #104300)
POGZ	White-sutton syndrome (OMIM #616364)
PPOX	Porphyria variegate (OMIM #176200)
PPT1	Ceroid lipofuscinosis (OMIM #256730)
PRDM8	Epilepsy (OMIM #616640)
PRKAR1A	Pigmented nodular adrenocortical disease (OMIM #610489)
PRODH	Hyperprolinemia (OMIM #239500) Schizophrenia 4 (OMIM #600850)
PRNP	Fatal familial insomnia (OMIM #600072) Gerstmann-straussler disease (OMIM #137440) Creutzfeldt-jakob disease (OMIM #123400) Spongiform encephalopathy with neuropsychiatric features (OMIM #606688)
PSEN1	Alzheimer disease 3 (OMIM #607822) Frontotemporal dementia (OMIM #600274)
PSEN2	Alzheimer disease 4 (OMIM #606889)
PTPN22	Systemic lupus erythematosus (OMIM #152700)
RHAG	Overhydrated hereditary stomatocytosis (OMIM #185000)
RNF139	Renal cell carcinoma (OMIM #144700)
ROGDI	Kohlschutter-tonz syndrome (OMIM #226750)
RPL10	Susceptibility to x-linked autism 5 (OMIM #300847)
RPS6KA3	Coffin-lowry syndrome (OMIM #303600)
SCNA	Parkinson disease 4 (OMIM #605543)
SGCE	Dystonia 11 (OMIM #159900)
SHANK3	Phelan-mcdermid syndrome (OMIM #606232)
SLC12A6	Agenesis of the corpus callosum with peripheral neuropathy (OMIM #218000)
SLC1A1	Schizophrenia 18 (OMIM #615232)
SLC20A2	Basal ganglia calcification (OMIM #213600)
SLC25A13	Adult onset citrullinemia type II (OMIM #603471)
SLC6A19	Hartnup disorder (OMIM #234500)
SLC6A8	Cerebral creatine deficiency syndrome 1 (OMIM #300352)
SLC7A7	Lysinuric protein intolerance (OMIM #222700)
SNCA	Dementia (OMIM #127750) Parkinson disease 1 (OMIM #168601)
SNCB	Dementia (OMIM #127750)

SOBP	Mental retardation (OMIM #613671)
SPG20	Spastic paraplegia 20 (OMIM #275900)
TBC1D7	Macrocephaly/megalencephaly syndrome (OMIM #248000)
TBP	Late-onset parkinson disease (OMIM #168600)
TBX1	Conotruncal heart malformations (OMIM #217095)
TP53	Li-fraumeni syndrome 1 (OMIM #151623)
TPH2	Major depressive disorder (OMIM #608516)
TREX1	Systemic lupus erythematosus (OMIM #152700)
TTC19	Mitochondrial complex iii deficiency, nuclear type (OMIM #615157)
TTR	Amyloidosis (OMIM #105210)
TWIST1	Saethre-chotzen syndrome (OMIM #101400)
VHL	Renal cell carcinoma (OMIM #144700)
VPS13A	Choreoacanthocytosis (OMIM #200150)
VPS35	Parkinson disease 17 (OMIM #614203)
WFS1	Wolfram-like syndrome (OMIM #614296) Wolfram syndrome 1 (OMIM #222300)
XPR1	Basal ganglia calcification (OMIM #616413)
ZBTB20	Primrose syndrome (OMIM #259050)
ZDHHC9	Mental retardation, x-linked, syndromic, raymond type (OMIM #300799)
ZFYVE26	Spastic paraplegia 15 (OMIM #270700)

Supplementary Table 3.5 Single-gene conditions characterized by schizophrenia-like signs or symptoms identified through an OMIM search using the following keywords: "schizo*" or "psychot*" or "hallucin*" or "psychosis". The OR for each gene was calculated using the dbGaP case control dataset. Gene in bold has significant ORs and were included in the expanded SCZ susceptibility network.

Abbreviations: SCZ: schizophrenia, OMIM: Online Mendelian Inheritance in Man, OR: odds ratio

Supplementary Table 3.6 Genes predicted by genemania to be functionally related with genes having significant odds ratios in Supplementary Table 3.5

Gene
BCKDHB
CACNB1
CP
FAH
GALR1
GALR2
GLUD1
GNS
GPR151
MAP3K8
MKLN1
NOV
OBSL1
OPLAH
PLD3
PLPP3
PTPA
TMEM39B
TNFSF4
TOM1L1

Supplementary Table 3.6 Genes responsible for single-gene conditions characterized by SCZ or SCZ-like signs or symptoms , which had significant ORs using our SCZ case-control dataset, were inputted into Genemania. Genemania suggested 20 genes that functionally related to the inputted list. The OR for each gene was calculated using the dbGaP case control dataset. The gene in bold had a significant OR and was included in the expanded susceptibility network.

Abbreviations: dbGaP: database for Genotypes and Phenotypes, OR: Odds ratio

Chapter 4: Discussion

SCZ and ASD are common diseases with complex inheritance and a high estimated heritability. However, there is a general consensus that there are no genetic risk factors of major contribution based on linkage and GWAS studies. Recent WES studies support the importance of de novo SNVs (in multiple genes) to increase one's risk to develop ASD and SCZ (Table 3.1 and 3.2, respectively), especially in sporadic cases. DNMs are a plausible explanation for the high incidence of the diseases worldwide that remain constant despite different cultures, variability in environmental factors and reduced fitness (26). The network hypothesis states that the genetics of SCZ and ASD (and potentially other diseases of complex inheritance) can be explained when affected individuals accumulate *de novo and inherited mutations* within a susceptibility network (or a subset of it), which increases the individual's liability to reach the CDT and express the disease. In both familial and sporadic cases, we believe it is possible that parents are unaffected because the mutation load of ASD genes in each of them is not sufficient to reach the CDT *however* the child is affected after inheriting a high mutation load in the network from **both** parents and the occurrence of DNMs.

Network hypothesis is not supported in autism

Our results do not support the network hypothesis for ASD; we find that there is no difference in the mutation load in ASD susceptibility genes between probands and the average mutation load in parents. However, when we compared the mutation load in ASD genes in sibpairs, we observed that *unaffected* siblings were enriched in functional variants in the de novo ASD network ($p = 0.02$), the TADA network ($p = 0.0008$) *and* the randomly selected HC control genes ($p = 7.36 \times 10^{-6}$), respectively. We explored and confirmed that there is a systematic bias for

a significant enrichment of functional variants in unaffected siblings occurring throughout the exome (Figure 3.2A) and that this bias is not specific to the selected networks of ASD susceptibility genes and HC control genes.

We explored if there was a discrepancy in sequencing depth between affected and unaffected siblings since depth of sequencing is indicative of genotype quality and calling accuracy (83) and could influence the number of variants that are called. We report that on average, probands were sequenced with a similar depth to their unaffected siblings when considering positions with functional variants in the ASD susceptibility network. We applied stricter filtering criteria to remove variants with very low or very high depth of sequencing in order to limit false positives, and repeated the permutation analysis using functional variants in the whole exome (Figure 3.2B). Although the p-values were generally higher, the distribution remained skewed indicating that we did not fully correct the bias in mutation load of functional variants between sibpairs by applying our stricter filtering criteria.

However, when we controlled for the effect of depth of sequencing through conditional logistic regression using the original filtering criteria, we find that probands had significantly fewer variants in the ASD network relative to unaffected siblings (Table 3.6A). This counterintuitive significant difference in the mutation load of ASD genes between sibpairs no longer remained once we applied our stricter filtering criteria and controlled for depth of sequencing in a subsequent conditional logistic regression (Table 3.6B). We do not have evidence to support the network hypothesis for ASD at this time.

Evidence to support the network hypothesis in schizophrenia

We tested the network approach using 35 local trios and report nonsignificant results,

potentially due to a small sample size. In two larger independent SCZ datasets, we show that 14 genes reported to carry DNMs in schizophrenic patients, were enriched in functional variants relative to familial controls ($p = 0.04$) and external controls ($p = 0.02$), while randomly selected HC control genes were not ($p = 0.23$ and $p = 0.09$, respectively). Moreover, in the 598 trios we showed that none of the functional variants that the probands had in the SCZ susceptibility genes were DNMs. Our results support that DNM literature can be used to identify susceptibility genes but variants transmitted from parents are also important in disease risk and may be sufficient to cause disease. We argue that affected individuals already have a high mutation load of transmitted variants in the network and the DNM is a final hit that causes the individual to reach the CDT.

The network hypothesis is a plausible explanation for some of the genetic heterogeneity of SCZ but requires further inquiry to be consolidated. We believe that affected individuals accumulate inherited and de novo variants in the SCZ network and reach the CDT. However, it is possible that two affected individuals have different mutational profiles meaning that although they require a certain number of variants to reach the CDT, the specific susceptibility genes contributing to their condition may vary from one individual to the other (even among affected individuals from the same family). Since every affected individual can vary in the subset of susceptibility genes with mutations from the overall network, this may explain the relatively low yield of replicable susceptibility loci identified by SCZ linkage studies. In the example pedigree (Supplementary Figure 4.1) with no clear mode of inheritance, the children of the couple from the first generation (II-1 and II-3) are unaffected despite having an affected father. According to our hypothesis, it is possible that the children were unaffected because the mutation load in the SCZ susceptibility genes transmitted by both parents was not sufficient to reach the CDT. In

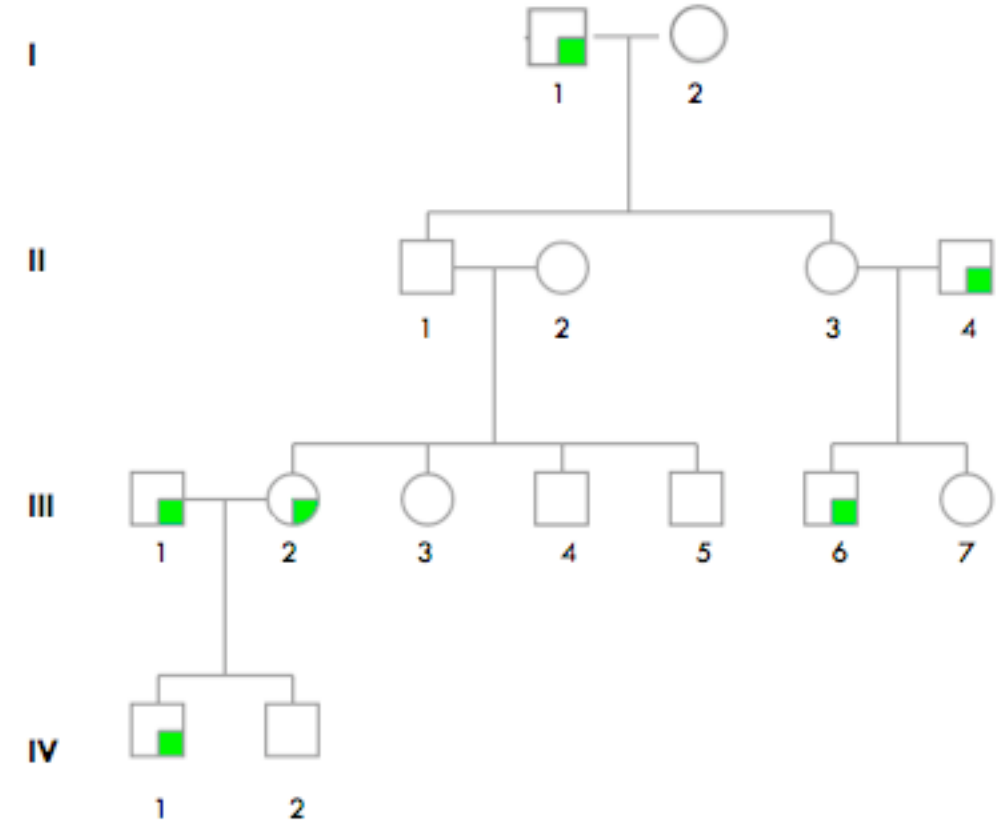
contrast, the unaffected parents from the second generation (II-1 and II-2) have an affected child. The affected child (III-2) inherits a high mutation load in their SCZ susceptibility genes (and possible DNMs occur), although the mutation load in each of the parents is not sufficient for them to be affected.

Although not considered here, other genetic variants such as CNVs, rare relatively deterministic genes, some chromosomal anomalies and environmental factors also contribute to the genetic heterogeneity of SCZ. Similarly, in utero and postnatal environmental threats are also interacting with genetics to increase one's predisposition to disease. A nested case-control study interested in determining the effect of maternal infection (in this case influenza) on SCZ risk investigated birth records in California between 1959 and 1966 and determined the diagnosis of children 40 years later. After confirming the presence of influenza anti-body in maternal serum, individuals exposed to influenza in the first trimester had a 7-fold increase risk for SCZ (84). Although viruses do not usually penetrate the placenta, it is believed that maternal immune response indirectly impacts the fetus through exposure to proinflammatory cytokines. The investigation of 17 adult SCZ cases that were exposed to increased level of interluken-8 in utero had changes to the structure of their brain recoded by magnetic resonance imaging that was consistent with SCZ (85). Bobetsis et al, identified 74 fetal genes that were epigenetically misregulated by bacterial infection in a mouse model (86). Epidemiological studies have found evidence that difficult social environments such as severe bullying could increase one's risk for psychosis 2-fold (87). Thus, it would be important to consider association of environment and genetic predispositions in the network hypothesis as the same number of genetic hits in susceptibility genes may have different impact in two individuals exposed to different environmental stressors.

Our results provide preliminary evidence to support the network hypothesis in SCZ using 14 susceptibility genes identified by WES studies reporting an enrichment of DNMs in affected individuals. It is likely that we have identified a small portion of the SCZ susceptibility network, given that we have selected genes based on the incidence of rare DNMs and current studies estimate 100 - 1000 susceptibility genes. In an effort to identify additional susceptibility genes, we identified the genes responsible for single-gene conditions with published evidence that (in some patients) they are characterized by SCZ-like signs or symptoms. These genes were inputted in Genemania to identify functionally related genes which could constitute potential susceptibility genes. Genes with significant ORs were prioritized and included in the expanded network. There was no difference in the mutation load of the expanded SCZ susceptibility network between SCZ probands and the average mutation load in parents ($p = 0.19$), thus we do not find any evidence to support the hypothesis in the expanded gene network. According to our hypothesis, variants of most low effect present in many genes interact cooperatively to cause disease. Thus, it is not unusual that many genes considered for the expanded SCZ network did not have exceedingly high ORs ranging from 1 to 2 (i.e. many more variants in affected individuals relative to unaffected individuals). This is a potential limitation in our ability to prioritize novel susceptibility genes.

Supplementary Figure

Supplementary Figure 4.1 Example schizophrenia pedigree



Chapter 5: Conclusions and Future Directions

We provide preliminary evidence for the hypothesis in SCZ and we report the importance of inherited variants from both parents within the network of susceptibility genes for an individual to reach the CDT. We believe that affected individuals have variants in different combinations of genes within the susceptibility network, which may explain the heterogeneity of the disease and the inability of linkage and association studies to find genetic risk factors of major effect. The network hypothesis is not supported for ASD since there was no difference in the mutation load of ASD susceptibility genes between probands and unaffected family members when controlling for the effect of sequencing depth. To rule out if these null results are specific to our dataset, further testing of the network hypothesis would be required in an independent family-based dataset. Although we did not have approved access to an ASD case control dataset, this design would emphasize the difference in the network mutation load between affected and unaffected individuals because they do not share a family history. We could then determine if affected individuals have a higher network mutation load as compared to the unrelated unaffected individuals.

The focus of future research should be to replicate our findings for SCZ and to test the hypothesis for ASD using independent datasets. Given the genetic complexities of these disorders, it would be interesting to identify additional susceptibility genes by repeating our approach (OMIM search and Genemania) in larger datasets or to explore novel approaches to identify risk genes such as TADA (82). In this project, we analyzed rare SNVs and indels, which were assumed to be of equal weight when calculating the mutation load. Evidence suggests that other types of variants such as CNVs and common variants contribute to ASD and SCZ. Krumm et al. determined that the burden of rare inherited CNVs, de novo CNVs and de novo likely gene

disruptive SNVs were independent risk factors in ASD using a logistic regression model (76). In the future, we should incorporate different classes of variants in the network hypothesis and to assign differential weights to each class of variant based on our understanding of the relative functional impact and their effect on risk when calculating the mutation load so that we can more accurately model an individual's ability to reach the CDT. Importantly, epidemiological studies have identified many environmental risk factors (albeit in utero or postnatal) implicated in SCZ and ASD that often interact with genetics to exasperate risk that have not been considered in this work. These environmental risk factors should be considered in future work.

References

1. American Psychiatric A. Diagnostic and Statistical Manual of Mental Disorders (DSM-5Â®): American Psychiatric Pub; 2013.
2. Cardno AG, Gottesman II. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. Am J Med Genet. 2000;97(1):12-7.
3. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry. 2003;60(12):1187-92.
4. Coon H, Jensen S, Holik J, Hoff M, Myles-Worsley M, Reimherr F, et al. Genomic scan for genes predisposing to schizophrenia. Am J Med Genet. 1994;54(1):59-71.
5. Moises HW, Yang L, Kristbjarnarson H, Wiese C, Byerley W, Macciardi F, et al. An international two-stage genome-wide search for schizophrenia susceptibility genes. Nat Genet. 1995;11(3):321-4.
6. Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G, et al. Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. Nat Genet. 1998;20(1):70-3.
7. Faraone SV, Matise T, Svrakic D, Pepple J, Malaspina D, Suarez B, et al. Genome scan of European-American schizophrenia pedigrees: results of the NIMH Genetics Initiative and Millennium Consortium. Am J Med Genet. 1998;81(4):290-5.
8. Kaufmann CA, Suarez B, Malaspina D, Pepple J, Svrakic D, Markel PD, et al. NIMH Genetics Initiative Millenium Schizophrenia Consortium: linkage analysis of African-American pedigrees. Am J Med Genet. 1998;81(4):282-9.
9. Levinson DF, Mahtani MM, Nancarrow DJ, Brown DM, Kruglyak L, Kirby A, et al. Genome scan of schizophrenia. Am J Psychiatry. 1998;155(6):741-50.
10. Shaw SH, Kelly M, Smith AB, Shields G, Hopkins PJ, Loftus J, et al. A genome-wide search for schizophrenia susceptibility genes. Am J Med Genet. 1998;81(5):364-76.
11. Hovatta I, Varilo T, Suvisaari J, Terwilliger JD, Ollikainen V, Arajärvi R, et al. A genomewide screen for schizophrenia genes in an isolated Finnish subpopulation, suggesting multiple susceptibility loci. Am J Hum Genet. 1999;65(4):1114-24.
12. Williams NM, Rees MI, Holmans P, Norton N, Cardno AG, Jones LA, et al. A two-stage genome scan for schizophrenia susceptibility genes in 196 affected sibling pairs. Hum Mol Genet. 1999;8(9):1729-39.
13. Ekelund J, Lichtermann D, Hovatta I, Ellonen P, Suvisaari J, Terwilliger JD, et al. Genome-wide scan for schizophrenia in the Finnish population: evidence for a locus on chromosome 7q22. Hum Mol Genet. 2000;9(7):1049-57.
14. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995;11(3):241-7.
15. Schizophrenia Psychiatric Genome-Wide Association Study C. Genome-wide association study identifies five new schizophrenia loci. Nature genetics. 2011;43(10):969-76.
16. Ripke S, O'Dushlaine C, Chambert K, Moran JL, KÃ¤hler AK, Akterin S, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nature genetics. 2013;45(10):1150-9.
17. Yin J, Lin J, Luo X, Chen Y, Li Z, Ma G, et al. miR-137: a new player in schizophrenia. Int J Mol Sci. 2014;15(2):3262-71.
18. Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature. 2009;460(7256):753-7.

19. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, et al. Common variants conferring risk of schizophrenia. Nature. 2009;460(7256):744-7.
20. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science. 2008;320(5875):539-43.
21. Consortium IS. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature. 2008;455(7210):237-41.
22. Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S, et al. Large recurrent microdeletions associated with schizophrenia. Nature. 2008;455(7210):232-6.
23. Vacic V, McCarthy S, Malhotra D, Murray F, Chou HH, Peoples A, et al. Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. Nature. 2011;471(7339):499-503.
24. Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J, et al. Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am J Psychiatry. 2011;168(3):302-16.
25. Karayiorgou M, Morris MA, Morrow B, Shprintzen RJ, Goldberg R, Borrow J, et al. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. Proc Natl Acad Sci U S A. 1995;92(17):7612-6.
26. Jablensky A, Sartorius N, Ernberg G, Anker M, Korten A, Cooper JE, et al. Schizophrenia: manifestations, incidence and course in different cultures A World Health Organization Ten-Country Study. Psychological medicineMonograph supplement. 1992;20:1-97.
27. Bassett AS, Bury A, Hodgkinson KA, Honer WG. Reproductive fitness in familial schizophrenia. Schizophrenia research. 1996;21(3):151-60.
28. Malaspina D, Brown A, Goetz D, Alia-Klein N, Harkavy-Friedman J, Harlap S, et al. Schizophrenia risk and paternal age: a potential role for de novo mutations in schizophrenia vulnerability genes. CNS spectrums. 2002;7(01):26-9.
29. Awadalla P, Gauthier J, Myers RA, Casals F, Hamdan FF, Griffing AR, et al. Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. American Journal of Human Genetics. 2010;87(3):316-24.
30. Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y, et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nature genetics. 2012;44(12):1365-9.
31. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, et al. De novo mutations in schizophrenia implicate synaptic networks. Nature. 2014;506(7487):179-84.
32. McCarthy SE, Gillis J, Kramer M, Lihm J, Yoon S, Berstein Y, et al. De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. Molecular psychiatry. 2014;19(6):652-8.
33. Li J, Cai T, Jiang Y, Chen H, He X, Chen C, et al. Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. Molecular psychiatry. 2016;21(2):290-7.
34. Ambalavanan A, Girard SL, Ahn K, Zhou S, Dionne-Laporte A, Spiegelman D, et al. De novo variants in sporadic cases of childhood onset schizophrenia. European Journal of Human Genetics. 2015.

35. Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Molecular psychiatry. 2012;17(2):142-53.
36. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature. 2014;506(7487):185-90.
37. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature. 2014;506(7487):185-90.
38. Gulsuner S, Walsh T, Watts AC, Lee MK, Thornton AM, Casadei S, et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell. 2013;154(3):518-29.
39. Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron. 2011;70(5):898-907.
40. Phillips M, Pozzo-Miller L. Dendritic spine dysgenesis in autism related disorders. Neurosci Lett. 2015;601:30-40.
41. Griswold AJ, Dueker ND, Van Booven D, Rantus JA, Jaworski JM, Slifer SH, et al. Targeted massively parallel sequencing of autism spectrum disorder-associated genes in a case control cohort reveals rare loss-of-function risk variants. Molecular autism. 2015;6:43-015-0034-z. eCollection 2015.
42. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014;515(7526):209-15.
43. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron. 2015;87(6):1215-33.
44. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. Neuron. 2012;74(2):285-99.
45. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014;515(7526):216-21.
46. McCarthy M. Autism diagnoses in the US rise by 30%, CDC reports. BMJ. 2014;348:g2520.
47. Fombonne E. Epidemiology of pervasive developmental disorders. Pediatr Res. 2009;65(6):591-8.
48. Schaefer GB. Clinical Genetic Aspects of ASD Spectrum Disorders. Int J Mol Sci. 2016;17(2).
49. Altmüller J, Palmer LJ, Fischer G, Scherb H, Wjst M. Genomewide scans of complex human diseases: true linkage is hard to find. Am J Hum Genet. 2001;69(5):936-50.
50. Risch NJ. Searching for genetic determinants in the new millennium. Nature. 2000;405(6788):847-56.
51. Cantor RM, Kono N, Duvall JA, Alvarez-Retuerto A, Stone JL, Alarcón M, et al. Replication of autism linkage: fine-mapping peak at 17q21. Am J Hum Genet. 2005;76(6):1050-6.
52. Badner JA, Gershon ES. Regional meta-analysis of published data supports linkage of autism with markers on chromosome 7. Mol Psychiatry. 2002;7(1):56-66.

53. Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K, et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. N Engl J Med. 2008;359(16):1685-99.
54. Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med. 2011;13(9):777-84.
55. Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, Huguet G, et al. Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. PLoS Genet. 2012;8(2):e1002521.
56. De Wolf V, Brison N, Devriendt K, Peeters H. Genetic counseling for susceptibility loci and neurodevelopmental disorders: the del15q11.2 as an example. Am J Med Genet A. 2013;161A(11):2846-54.
57. Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, Badner JA, et al. Recurrent 16p11.2 microdeletions in autism. Hum Mol Genet. 2008;17(4):628-38.
58. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature. 2009;459(7246):569-73.
59. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of de novo copy number mutations with autism. Science (New York, NY). 2007;316(5823):445-9.
60. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010;466(7304):368-72.
61. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron. 2011;70(5):863-85.
62. Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am J Hum Genet. 2007;80(4):727-39.
63. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nature genetics. 2011;43(6):585-9.
64. Cooper GM, Goode DL, Ng SB, Sidow A, Bamshad MJ, Shendure J, et al. Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. Nat Methods. 2010;7(4):250-1.
65. Grantham R. Amino acid difference formula to help explain protein evolution. Science. 1974;185(4154):862-4.
66. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012;485(7397):246-50.
67. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. 2012;485(7397):242-5.
68. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012;485(7397):237-41.

69. Hultman CM, Sandin S, Levine SZ, Lichtenstein P, Reichenberg A. Advancing paternal age and risk of autism: new evidence from a population-based study and a meta-analysis of epidemiological studies. Mol Psychiatry. 2011;16(12):1203-12.
70. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, et al. Rate of de novo mutations and the importance of father/'s age to disease risk. Nature. 2012;488(7412):471-5.
71. Dong S, Walker MF, Carriero NJ, DiCola M, Willsey AJ, Adam YY, et al. De novo insertions and deletions of predominantly paternal origin are associated with autism spectrum disorder. Cell reports. 2014;9(1):16-23.
72. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al. Excess of rare, inherited truncating mutations in autism. Nature genetics. 2015;47(6):582-8.
73. Iossifov I, Levy D, Allen J, Ye K, Ronemus M, Lee YH, et al. Low load for disruptive mutations in autism genes and their biased transmission. Proc Natl Acad Sci U S A. 2015;112(41):E5600-7.
74. Toma C, Torrico B, Hervas A, Valdes-Mas R, Tristan-Noguero A, Padillo V, et al. Exome sequencing in multiplex autism families suggests a major role for heterozygous truncating mutations. Molecular psychiatry. 2014;19(7):784-90.
75. Poultney CS, Goldberg AP, Drapeau E, Kou Y, Harony-Nicolas H, Kajiwara Y, et al. Identification of small exonic CNV from whole-exome sequence data and application to autism spectrum disorder. American Journal of Human Genetics. 2013;93(4):607-19.
76. Krumm N, O'Roak BJ, Karakoc E, Mohajeri K, Nelson B, Vives L, et al. Transmission disequilibrium of small CNVs in simplex autism. American Journal of Human Genetics. 2013;93(4):595-606.
77. Falconer DS. The inheritance of liability to certain diseases, estimated from the incidence among relatives. Annals of Human Genetics. 1965;29(1):51-76.
78. Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. Nature genetics. 2011;43(9):860-3.
79. Haw R, Hermjakob H, D'Eustachio P, Stein L. Reactome pathway analysis to enrich biological discovery in proteomics data sets. Proteomics. 2011;11(18):3598-613.
80. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38(Web Server issue):W214-20.
81. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron. 2015;87(6):1215-33.
82. He X, Sanders SJ, Liu L, De Rubeis S, Lim ET, Sutcliffe JS, et al. Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. PLoS Genet. 2013;9(8):e1003671.
83. Garner C. Confounded by sequencing depth in association studies of rare alleles. Genet Epidemiol. 2011;35(4):261-8.
84. Brown AS, Begg MD, Gravenstein S, Schaefer CA, Wyatt RJ, Bresnahan M, et al. Serologic evidence of prenatal influenza in the etiology of schizophrenia. Arch Gen Psychiatry. 2004;61(8):774-80.
85. Ellman LM, Deicken RF, Vinogradov S, Kremen WS, Poole JH, Kern DM, et al. Structural brain alterations in schizophrenia following fetal exposure to the inflammatory cytokine interleukin-8. Schizophr Res. 2010;121(1-3):46-54.

86. Bobetsis YA, Barros SP, Lin DM, Arce RM, Offenbacher S. Altered gene expression in murine placentas in an infection-induced intrauterine growth restriction model: a microarray analysis. J Reprod Immunol. 2010;85(2):140-8.
87. Schreier A, Wolke D, Thomas K, Horwood J, Hollis C, Gunnell D, et al. Prospective study of peer victimization in childhood and psychotic symptoms in a nonclinical population at age 12 years. Arch Gen Psychiatry. 2009;66(5):527-36.