Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

7-23-2015 12:00 AM

Assessment of Genome-Wide Genetic and Epigenetic De Novo Variation in Families with Monozygotic Twins Discordant for Schizophrenia

Christina A. Castellani
The University of Western Ontario

Supervisor
Dr. Shiva M. Singh
The University of Western Ontario

Graduate Program in Biology
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Christina A. Castellani 2015

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Molecular Genetics Commons

Recommended Citation
Castellani, Christina A., "Assessment of Genome-Wide Genetic and Epigenetic De Novo Variation in Families with Monozygotic Twins Discordant for Schizophrenia" (2015). *Electronic Thesis and Dissertation Repository*. 2972.
https://ir.lib.uwo.ca/etd/2972

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Assessment of Genome-Wide Genetic and Epigenetic *De Novo* Variation in Families with Monozygotic Twins Discordant for Schizophrenia

(Thesis Format: Integrated Article)

by

Christina Angela Castellani

Graduate Program in Biology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada
July 2015

© Christina Angela Castellani 2015

Abstract

Schizophrenia (OMIM: 181500) is a common, debilitating and life-altering disorder. It affects 1% of the population worldwide and most often presents in early adulthood leading to devastating effects for patients, their families and society. Despite thousands of studies performed on the underlying mechanisms of schizophrenia, the causes of the disease remain unknown. However, what is known is that environmental, genetic and epigenetic factors contribute to the development of this complex disorder. Although a genetic role in schizophrenia is well established, the search for schizophrenia genes using traditional approaches has remained challenging. Interestingly, monozygotic twins show concordance for schizophrenia only 50% of the time and therefore provide a unique scenario for genomic analysis.

This Doctoral thesis examines the genetic and epigenetic contributions to schizophrenia discordance in monozygotic twins. In this thesis, I have identified and characterized genome-wide changes through the use of the Affymetrix SNP 6.0 Microarray, Complete Genomics whole genome sequencing and the Nimblegen Methylation 720k Microarray. Specifically, I have identified genetic and epigenetic differences between monozygotic twins discordant for schizophrenia.

The results show multiple genetic and epigenetic changes between monozygotic twins with discordance for schizophrenia. Some of these differences are patient-specific and others are shared between affected twins in the study. In addition, some of these differences affected genes and others did not. Many of the genes and genomic regions have been previously implicated in schizophrenia and neurodevelopmental disorders. The findings reinforce the concept that individual genomes harbor extensive variability, some inherited and some acquired. Even monozygotic twins are not identical and each individual may be a mosaic; carrying different sequence variations in different cells. The results also suggest that discordance for schizophrenia in monozygotic twins may result from the accumulation of genetic and epigenetic mutations that lead to the disease

threshold being met in one twin only. The results argue for the involvement of *de novo* mutations in genetic individuality and complex disease. Improved understanding of the genomic contributions to schizophrenia is critical for movement towards earlier and more accurate diagnosis, better treatment and further understanding of this complex mental health disorder.

Keywords: Monozygotic Twins, Discordance, Schizophrenia, Genomics, DNA Microarray, Complex Disease, Psychosis, Epigenomics, Complete Genomics Sequencing, meDIP, *de novo* change, DNA Variation, Copy Number Variation, Structural Variation, DNA Methylation, Differentially Methylated Regions, CNV Calling Methods

Co-Authorship

Chapter 2 of this thesis contains material from two published manuscripts. The first, “Biological relevance of CNV calling methods using familial relatedness including monozygotic twins” published in *BMC Bioinformatics* on April 21st 2014 and co-authored by Dr. Melkaye G Melka, Andrea E Wishart, Marjorie E Locke, Zain Awamleh, Dr. Richard L O’Reilly and Dr. Shiva M Singh. The second, “Copy number variation distribution in six monozygotic twin pairs discordant for schizophrenia” published in *Twin Research and Human Genetics* on February 20th 2014 and co-authored by Zain Awamleh, Dr. Melkaye G Melka, Dr. Richard L O’Reilly and Dr. Shiva M Singh; ©Cambridge University Press (Reprinted with permission).

Chapter 3 of this thesis contains material from the manuscript entitled, “Genomes of monozygotic twins discordant for schizophrenia delineate *de novo* features of twin differences and disease threshold” submitted to *Nature Scientific Reports* on May 14th 2015 and co-authored by Dr. Melkaye G Melka, Jane L Gui, Anthony J Gallo, Dr. Richard L O’Reilly and Dr. Shiva M Singh. In addition, Chapter 3 includes two figures that were created by Marjorie Locke (3.3 and 3.4).

Chapter 4 of this thesis contains material from the manuscript entitled, “DNA methylation differences in monozygotic twin pairs discordant for schizophrenia identifies psychosis related genes and networks” published in *BMC Medical Genomics* on May 6th 2015 and co-authored by Benjamin I Laufer, Dr. Melkaye G Melka, Eric J Diehl, Dr. Richard L O’Reilly and Dr. Shiva M Singh.

Chapter 5 of this thesis contains material from the manuscript entitled, “Integration of DNA Sequence and DNA Methylation Changes in Two Monozygotic Twin Pairs Discordant for Schizophrenia” submitted to *Schizophrenia Research* on May 29th 2015 and co-authored by Dr. Melkaye G Melka, Jane L Gui, Dr. Richard L O’Reilly and Dr. Shiva M Singh.

In each case, I was listed as first author on the manuscript and performed the experiments, analyzed the data, and wrote the manuscript alongside Dr. Shiva M Singh.

Dedication

This thesis is dedicated to my great-grandmother, Angelina Dipalma-Marti, who always believed in me without reservation.

It is equally dedicated to Santino, Alessandro and Isabella Castellani - may they always know how loved they are.

Acknowledgements

This thesis represents many years of effort that would not have been possible without the support of a list of individuals too large to count. I am incredibly thankful to my co-workers, family, friends, administrators and mentors.

First and foremost, I am thankful to my supervisor, Dr. Shiva Singh, for giving me the opportunity to find my true passion through his guidance, leadership and patience. I feel very fortunate to have been continuously encouraged by Dr. Singh throughout my training and to always be able to rely on his support. He has carefully ignited a spark for research, curiosity, teaching and discovery that I hope will only continue to flourish in the coming years.

I have also received invaluable mentorship, support and training from our collaborator, Dr. Richard O'Reilly. Dr. O'Reilly has never hesitated to assist me and has been a very important link to the clinical aspects of this research. Without Dr. O'Reilly's hard work and persistence, we would not have been able to collect the unique samples included in this report. I am also thankful to the individuals and families who participated in these studies.

I wish to thank the Department of Biology at Western University, particularly the faculty, administrators and staff for the opportunity to pursue this Degree and the various forms of support I have received from them.

I am grateful to my advisors Dr. Kathleen Hill and Dr. Robert Cumming, for their support throughout every step of my degree. I especially wish to extend thanks to Dr. Cumming for the critical review of my thesis, and to Dr. Hill for her ongoing mentorship, advice and encouragement.

In addition, I wish to thank the faculty who have assisted me, and/or provided me with invaluable teaching opportunities, particularly Dr. Brenda Murphy, Dr. Mark Daley, Dr. Greg Kelly, Dr. Graham Thompson and Dr. Susanne Kohalmi.

To my co-workers who have provided the utmost of support along the way. I extend special thanks to Dr. Melkaye Melka, Benjamin Laufer, Bonnie

Alberry, Eric Diehl, Aniruddho Hoque, Morgan Kleiber, Raihan Uddin, Randa Stringer and Aditi Thakur.

I am thankful to David Carter from the London Regional Genomics Centre at Robarts Research Institute for his assistance over the years and to Marjorie E Locke for her computational assistance.

I wish to thank the countless volunteers and undergraduate students who contributed to this research with special thanks to Ali Pensamiento, Zain Awamleh, Emma Reble, Anthony Gallo, Courtney Gendron and Jane Gui.

To my parents, who will forever be my number one supporters. I have always felt fortunate to have parents who encourage me to go out into the world and seek opportunities whilst knowing that I can always count on them to catch me if I fall. Words cannot express the appreciation I have for the constant love and encouragement that you have given me all my life. To my brother Carlo, for being the person I can call when all else fails, and to his wonderful wife Kelly who has become a sister to me. Not only for their love and support but also for giving me the joy of being an Aunt to Santino, Alessandro and Isabella, a constant source of joy. Having a network of family members behind me has made all the difference, special thanks to my Nonna Pina, Nonna Gina, Aunt Dina, Uncle Vince, Aunt Liz, Uncle John, Cousins Daniel, Marcus and Monica and so many more. In addition, the unending support and many important lessons of Nonno Mario, Nonno Italo and Nonna Angelina stay with me always.

I have been fortunate enough in my life to be blessed not only with an amazing network of family support but also to have friends who treat me like their family. Although it would be impossible to name everyone, many thanks goes to Ashley Hooke, Mark Salvati, Julie Lawlor-Salvati, Christopher Fera, Mike Marynowski, Aaron Coutu, Courtney Coutu, Simon Coutu, and Mark Kaszowski. Special thanks to Mark Froncisz for being someone I have always been able to rely on without hesitation. I further extend appreciation to the additional mentors and cheerleaders who have assisted me along the way including Sharon McKillop, Frank Stilson, and the Gioiosa Family.

Funding Acknowledgements

This work was supported by grants from the Canadian Institutes of Health Research (CIHR), The Schizophrenia Society of Ontario (SSO) and the Ontario Mental Health Foundation (OMHF).

I am also thankful to the Canadian Institutes of Health Research (CIHR) for their support through a Canada Graduate Scholarship (CGS) Doctoral Research Award (DRA) and to the Government of Ontario for their support through the Ontario Graduate Scholarship (OGS) program.

Table of Contents

Abstract ii
Keywords iii
Co-Authorship iv
Acknowledgements vi
Funding Acknowledgements viii
Table of Contents ix
List of Tables xiii
List of Abbreviations xvi
Chapter 1 – Introduction 1
1.0 Overview 1
1.1 Schizophrenia 1
1.2 Genetics of Schizophrenia 3
1.3 Epigenetics of Schizophrenia 7
1.4 Genetic Differences Between Monozygotic Twins 9
1.5 Epigenetic Differences Between Monozygotic Twins 10
1.6 Evidence for Somatic Mutation 11
1.7 Advancing Technology in the Era of Genomics 12
1.7.1 Affymetrix® Human SNP 6.0 Microarray 13
1.7.2 Complete Genomics Inc. Whole Genome Sequencing 13
1.7.3 NimbleGen Human DNA Methylation 3x720k CpG Island Plus RefSeq Promoter Microarray 14
1.8 Hypothesis and Thesis Objectives 14
1.8.1 Objectives 14
1.9 Overview of Participants in Thesis Experiments 15
1.10 References 19
Chapter 2 - Assessing Copy Number Variation (CNV) in Monozygotic Twins Discordant for Schizophrenia Using the Affymetrix Human SNP 6.0 Microarray 29
2.0 Overview of Chapter 2 29
2A Biological relevance of CNV calling methods using familial relatedness including monozygotic twins 30
2A.0 Overview of Chapter 2A 30
2A.1 Introduction 31
2A.2 Methods 33
2A.2.1 Subjects 33
2A.2.2 Microarray Hybridization and Analysis 34
2A.2.3 CNV Call Merging 35
2A.2.4 Comparison of CNV Calls Between Software Programs 36
2A.2.5 Comparison of CNV Calls Between Subjects 36
2A.3 Results 37

2A.4 Discussion **46**
2A.5 Conclusion **47**
2B Assessment of Copy Number Variation in Monozygotic Twins Discordant for Schizophrenia **49**
2B.0 Overview of 2B **49**
2B.1 Introduction **49**
2B.2 Methods **53**
2B.2.1 Ethics Statement and Clinical Background 53
2B.2.2 DNA Preparation, Hybridization and CEL File Analysis 53
2B.2.3 Calling and Merging of CNVs for Individual Genomes 55
2B.2.4 Identification of Overlapping CNV Calls for Individual Genomes 55
2B.2.5 Comparison of CNV Calls Within Monozygotic Twin Pairs 56
2B.2.6 Confirmation of Unique CNVs 56
2B.3 Results **57**
2B.3.1 Twin Pair 1 63
2B.3.2 Twin Pair 2 63
2B.3.3 Twin Pair 3 64
2B.3.4 Twin Pair 4 65
2B.3.5 Twin Pair 5 65
2B.3.6 Twin Pair 6 66
2B.4 Discussion **67**
2B.4.1 Selected Regions/Genes 68
2B.4.2 *De novo* Mutation and Discordance 72
2B.5 Conclusion **73**
2.6 References **75**

Chapter 3 - Whole Genome Sequencing of Monozygotic Twins Discordant for Schizophrenia Defines *de novo* Features of Twin Differences **85**
3.0 Overview **85**
3.1 Introduction **86**
3.2 Methods **88**
3.2.1 Subjects 88
3.2.2 Complete Genomics Sequencing and Calling of Variants 92
3.2.3 Confirmation of Findings 94
3.2.4 Pathway Analysis 95
3.3 Results **96**
3.3.1 Defining genetic variation including *de novo* events 96
3.3.1.1 Sequence characteristics of six complete genomes 96
3.3.1.2 Mutational characteristics in twins affected with Schizophrenia 111
(i) Small Sequence Changes (SSCs) 111
(ii) Copy Number Variants (CNVs) 115
(iii) Structural Variants (SVs) 117
3.3.1.5 Independent confirmation of NGS results 120
3.3.2 Genetic differences between MZ twins and their discordance for schizophrenia 125

3.4 Discussion **136**
3.4.1 Individual human genomes contain extensive variability 136
3.4.2 MZ twins are genetically different 137
3.5 Conclusion **142**
3.6 References **142**

Chapter 4 - DNA Methylation Analysis of Families with Monozygotic Twins Discordant for Schizophrenia Identifies Psychosis Related Genes and Networks **151**
4.0 Overview **151**
4.1 Introduction **151**
4.2 Methods **153**
4.2.1 Subjects 153
4.2.2 DNA Collection and Array Processing 154
4.2.3 Microarray Analysis 157
4.3 Results **157**
4.3.1 MZ twins show differences in DNA Methylation profiles 158
4.3.2 MZ twins discordant for schizophrenia share genomic regions of differential methylation 161
4.3.3 MZ twins discordant for schizophrenia have differentially methylated networks and pathways; some (i) Pair-specific and others (ii) Shared 167
4.4 Discussion **176**
4.4.1 Conclusions 180
4.5 References **181**

Chapter 5 - Integration of Genome-Wide DNA Sequence and DNA Methylation Changes in Two Unrelated Monozygotic Twin Pairs Discordant for Schizophrenia **187**
5.0 Overview **187**
5.1 Introduction **188**
5.2 Methods **190**
5.2.4 Annotation and candidate gene analysis 192
5.3 Results **193**
5.3.1 Twin Pair 1 193
5.3.2 Twin Pair 2 202
5.3.3 Twin Pair 1 and Twin Pair 2 205
5.4 Discussion **209**
5.5 Conclusion **212**
5.6 References **213**

Chapter 6 - Discussion **218**
6.0 Overview **218**
6.1 Threshold Model **220**
6.1.1 Role of *de novo* Genetic and Epigenetic Mutations in the Threshold Model of Schizophrenia 222
6.1.2 Further Support for The Threshold Model of Schizophrenia 223

6.2 Context of Study Findings **224**
6.3 Methodological Considerations **226**
6.3.1 Study Strengths 226
6.3.2 Study Caveats 227
6.4 Implications **230**
6.5 Future Directions **230**
6.6 Final Conclusions **232**
6.7 References **233**

Appendices **238**

Curriculum Vitae **304**

List of Tables

Table 1.1. Summary of participants included in the thesis experiments. 16
Table 2.1. Raw (pre-merge) Copy Number Variant Calls by Program. 38
Table 2.2. Shared and unshared status of CNVs called by 3+ software programs in identical twins. 59
Table 2.3. CNVs identified by Affymetrix Genotyping Console, Partek Genomics Suite and PennCNV (A/P/p) and only present in the affected co-twin. 60
Table 2.4. CNVs identified by Affymetrix Genotyping Console, Partek Genomics Suite and PennCNV (A/P/p) and only present in the unaffected co-twin. 61
Table 2.5. TaqMan Real Time PCR Results. 70
Table 3.1. Identity of Sequenced Genomes. 97
Table 3.2. Shared and *de novo* Small Sequence Changes identified in blood (SNVs, Substitutions, Indels) as compared to NCBI Build 37. 104
Table 3.3. Exonic *de novo* variants in Affected Twins. 105
Table 3.4. Shared and unshared Copy Number Variants (CNVs) identified in blood as compared to NCBI Build 37. 110
Table 3.5. Copy Number Variants Unique to Affected Twins. 116
Table 3.6. Shared and unshared Structural Variation (SV) identified in blood as compared to NCBI Build 37. 118
Table 3.7. Regions analyzed using PCR Amplification of Specific Alleles. 121
Table 3.8. Regions predicted by NGS to be unshared between twins and selected for assessment by Real Time Quantitative PCR Analysis. 124
Table 3.9. Top 20 Canonical Pathways identified by Ingenuity Pathway Analysis (IPA) in the affected twin of each family. 128
Table 3.10. Genes affected in two pathways (Glutamate Receptor Signaling and Dopamine Feedback in cAMP Signaling) unique to the two unrelated schizophrenia patients representing family 1 and family 2. 135
Table 4.1. Differentially methylated regions identified in two affected MZD twins belonging to two unrelated families. 162
Table 4.2. Ingenuity Pathway Analysis (IPA) results A) Family 1 B) Family 2 C) Subset of 27 genes found in both affected twins. 168
Table 5.1. Genes (58) affected by *de novo* methylation and *de novo* sequence changes in Family 1. 196
Table 5.2. Ingenuity Pathway Analysis results showing the most significant pathways and networks for co-localized sequence variations and promoter methylation differences in Family 1. 201
Table 5.3. *De novo* sequence variations in the affected twin in Family 1 which have been previously associated with schizophrenia. 203
Table 5.4. Genes (13) affected by *de novo* methylation and *de novo* sequence changes in Family 2. 206
Table 5.5. *De novo* sequence variations in the affected twin in Family 2 which have been previously associated with schizophrenia. 208

List of Figures

Figure 2.1. Distribution of raw copy number variant calls (3957) by size generated by each software package. 40
Figure 2.2. Chromosomal distribution of raw copy number variants in autosomes (3957) generated by each software package. 41
Figure 2.3. Overlap of copy number calls across programs. 43
Figure 2.4. Copy number call differences made by each program and categorized by relatedness. 45
Figure 2.5. Flowchart of Chapter 2B experimental design 54
Figure 2.6. Copy Number Variation Loss at 7q11.21 in Twin 3A. 71
Figure 3.1. Members of the two discordant monozygotic twin pairs and one set of parents included in the study. 89
Figure 3.2. Flowchart of Methods Employed. 90
Figure 3.3. Call Quality of raw variant calls in 6 samples. 98
Figure 3.4. Read Depth Density of raw variant calls in 6 samples. 99
Figure 3.5. Circos plots of high confidence variant calls. 101
Figure 3.6. Chromosomal distribution of small sequence changes (SNVs, Indels, Block Substitutions) across 6 samples. 113
Figure 3.7. Identify of *de novo* (Family 1) and Presumed *de novo* (Family 2) Variants Identified in the Affected Twin of Family 1 (I-2-1) and the Affected Twin of Family 2 (II-2-1). 114
Figure 3.8. Presence of Mosaicism in Sanger Sequencing Results (PASA). 122
Figure 3.9. Approach for analysis of Ingenuity Pathway Results. 127
Figure 3.10. Venn diagram of the top 20 canonical pathways found to be overrepresented by variants in the sequencing dataset of affected and unaffected twins. 130
Figure 3.11. Ingenuity Pathways Analysis Identified Pathways in Affected Twins Only. Legend for symbols available in Appendix L. 131
Figure 4.1. Pedigrees of families included in the study. 156
Figure 4.2. Differential methylation in two twin pairs visualized in a Circos plot covering all chromosomes. 159
Figure 4.3. Overlap of differentially methylated regions between two families. 160
Figure 4.4. Manhattan plot representing methylation in the 15q11.2-15q12 region (Chr 15:17527952-25617776). 165
Figure 4.5. Manhattan plot representing methylation in the 1q21.1-1q21.3 region (Chr 1:143518837-152631162). 166
Figure 4.6. Common networks identified in both families. Legend for symbols can be found in Appendix L. 172
Figure 5.1. Differences between the MZD pair in Family 1 in sequence changes (1896) and changes in DNA methylation (330). 195
Figure 5.2. Differences between the MZD pair in Family 2 in sequence changes (1682) and changes in DNA methylation (138). 204
Figure 6.1. Visual representation of the proposed threshold model. 221

List of Appendices

Appendix A - Availability of Supporting Data 238

Appendix B - Ethics Approval 239

Appendix C - Structural Variants unique to affected member of Family 1......... 240

Appendix D - Structural Variants unique to affected member of Family 2......... 246

Appendix E - Selected Regions for Sanger Sequencing 252

Appendix F - Promoters differentially methylated in Family 1. 254

Appendix G - Promoters differentially methylated in Family 2. 271

Appendix H - *De novo* sequencing variations in the affected twin of Family 1 that show co-localized differential methylation and sequence variation. 278

Appendix I - Observed DNA methylation changes in 58 genes that showed co-localized *de novo* sequence variations in Family 1. 290

Appendix J - *De novo* sequencing variations in the affected twin of Family 2 that show co-localized differential methylation and sequence variation. 296

Appendix K - Observed DNA methylation changes in 13 genes that showed co-localized *de novo* sequence variations in Family 2. 299

Appendix L - Ingenuity Pathway Analysis Legend 301

Appendix M - Journal Copyright Approval 303

List of Abbreviations

A/P/p	Called by Affymetrix Genotyping Console, Partek Genomics Suite and PennCNV
ABI	Applied Biosystems
ACG (A)	Affymetrix Genotyping Console 4.1.1
bp	Base pairs
cAMP	Cyclic Adenosine Monophosphate
CEL file	CEL Intensity File: A single intensity value for each probe
CG	Complete Genomics Inc.
Chr	Chromosome
CN	Copy Number
CNAM	Copy Number Analysis Method
CNP	Copy Number Polymorphism
CNV	Copy Number Variant, minimum size 1 kb
Ct value	The cycle threshold is defined as the number of cycles required for the fluorescent signal to cross the threshold
dbSNP	The Single Nucleotide Polymorphism Database
de novo	Arose anew; assessed as absent in parents and co-twin
DGV	Database of Genomic Variants
DMR	Differentially Methylated Region
DNA	Deoxyribonucleic Acid
DNM	*De Novo* Mutations
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders IV
EWAS	Epigenome-Wide Association Study
GC	Genotyping Console
GH	Golden Helix SVS Suite 7.0
gpd	Genetic predisposing contributing to disease
gp	Genetic predisposing
GWAS	Genome-wide Association Study
HMM	Hidden Markov Model

Indel	Short insertions and deletions up to 50 bp in size
IPA	Ingenuity Pathway Analysis
IVA	Ingenuity Variant Analysis
kb	Kilobases
MeDIP	Methylated DNA Immunoprecipitation
methQTL	Methylation Quantitative Trait Loci
Mb	Megabases
MHC	Major Histocompatibility Complex
Mosaicism	The presence of two or more populations of cells with different genotypes arising from a single fertilized egg
MZ	Monozygotic
MZC	Concordant Monozygotic
MZD	Discordant Monozygotic
NGS	Next Generation Sequencing
OMIM	Online Mendelian Inheritance in Man
PC (p)	PennCNV
PCR	Polymerase Chain Reaction
PGC	Psychiatric Genomics Consortium
PGS (P)	Partek Genomics Suite
PolyPhen	Polymorphism Phenotyping, http://genetics.bwh.harvard.edu/pph2/
Provisional *de novo*	Not found in co-twin, parental information not available
QC	Quality Control
Rare Variant	Allele frequency less than 0.5%
RefSeq	Reference Sequence
RO	Reciprocal Overlap
sdRNA	snoRNA-Derived RNA
SIFT	Predicts if an amino acid substitution affects protein function, http://sift.jcvi.org/
snoRNA	Small Nucleolar RNA Species

SSC	Small Sequence Change
SNV	Single Nucleotide Variant
SNP	Single Nucleotide Polymorphism (>1% of population)
SV	Structural Variant, minimum size 500 bp
Variant	A difference from expectation (reference sequence)

Chapter 1 – Introduction

1.0 Overview

In 1865, Francis Galton wrote that twins offer a ‘‘means of distinguishing between the effects of tendencies received at birth, and of those that were imposed by the circumstances of their after lives; in other words, between the effects of nature and nurture’’. Monozygotic twins arise from a single fertilized egg and thus begin life as identical entities. Interestingly, monozygotic twins can be discordant for a number of traits, including disease. Monozygotic twins thus offer the ideal condition to study genomic change and specifically the level of dynamic processes occurring in our genomes during ontogeny and the effect of this dynamic nature on individual physiology. The next decade will see increasing efforts to understand two related issues in genetics: the basis of human individuality and the genetic underpinnings of complex diseases. Identical twins have the potential to assist in unraveling the answer to both of these challenging questions.

Few studies have looked at multiple types of *de novo* genetic changes with single base pair resolution in the same twins. To better understand genome changes within and across generations as well as the implications of these mutations to disease, this thesis investigates multiple types of *de novo* genomic changes in families with monozygotic (MZ) twins discordant for schizophrenia. Indeed, a number of *de novo,* patient-specific variations may play a large role in the development of schizophrenia and an experimental design that includes families with MZ twins discordant for the disorder allows a unique approach to the identification of these variants.

1.1 Schizophrenia

Schizophrenia (OMIM 181500) is a damaging disease with 1% prevalence worldwide (Huxley *et al*., 1964; Regier *et al*., 1993). It has a devastating effect on patients, their families and society (Jablensky *et al*., 1992). Schizophrenia is a

highly polygenic disorder with a complex array of risk loci (Ripke *et al.*, 2014). Attempts to identify a set of genes that are definitively contributing to schizophrenia etiology have been unsuccessful. At present, the single greatest risk-factor for developing schizophrenia is a family history of the disease (Mortensen *et al.*, 2010; Perala *et al.*, 2007). Schizophrenia patients have long been known to have reduced fecundity which should cause negative selection pressure for schizophrenia-causing genes (Bassett *et al.*, 1996; Laursen and Munk-Olsen, 2010). However, the rate of schizophrenia in society remains constant at around 1%, which strongly suggests that there is likely to be a genetic mechanism working to balance the negative selection. This supports the notion that *de novo* insults may be playing a significant role in the etiology of schizophrenia. Further supporting the hypothesis that *de novo* mutation may be responsible is the observation that paternal age, and an associated increase in germ cell mutations, is associated with an increased risk of schizophrenia (Byrne *et al.*, 2003).

Schizophrenia onset is typically seen just prior to, or during, early adult life (Delisi, 2009). Schizophrenia has a high heritability (~80%) and is classified by both positive and negative symptoms including pronounced impairment in cognitive functioning (Sullivan *et al.*, 2003). The hallmark of this disorder is a distortion in the perception of reality (Regier *et al.*, 1993). Diagnosis of schizophrenia usually follows the Diagnostic and Statistical Manual of Mental Disorders (First *et al.*, 1996, 1997). The disease presents with significant heterogeneity across patients and variability in both presentation and classification of the disease is common. Monozygotic twins show a high discordance rate (~50%) and dizygotic twins are concordant for the disease only 17% of the time (McGuffin *et al.*, 1994). The symptoms of schizophrenia are generally divided into four main subtypes: positive, negative, cognitive and mood disturbances. Positive symptoms include delusions and hallucinations; negative symptoms include a flat affect and social/emotional withdrawal; cognitive symptoms include attention, memory, decision-making and executive functioning

deficits; and mood disturbances include dysphoria and depression (Cañas *et al.*, 2005).

There is no cure for schizophrenia and treatment often includes a combination of antipsychotic medications, cognitive therapy, daily support and social skills training. Current treatments seek to target the positive symptoms of schizophrenia and have very little effect on the other categories of symptoms (Wright, 2014). In addition, the current drugs available for schizophrenia treatment are not without negative side effects. Due to the lack of effective treatment of schizophrenia, more than two-thirds of patients live their lives with detrimental symptoms and 5-13% eventually die by suicide (Pompili *et al.*, 2007; Saha *et al.*, 2005). Schizophrenia is therefore a large social, financial and health burden to patients, families and society (Knapp *et al.*, 2004). In fact, schizophrenia has been estimated to cost $2.02 billion per year in direct health care costs in Canada (Goeree *et al.*, 2005).

A summary of current findings regarding the genetics and epigenetics of schizophrenia is discussed below, it should be noted that a number of environmental factors (too many to list here) have been associated with this disease. Increased incidence in migrants, increased incidence in those living in urban settings, increased incidence in those exposed to maternal infection during pregnancy or suffering from childhood infection as well as substance abuse and stressful/traumatic events, have all been linked to the disease (Betts *et al.*, 2014; Khandaker *et al.*, 2013; Nielsen *et al.*, 2013; Selten *et al.*, 2013). However, the effect of each individual environmental factor appears to be quite small. Physiologically, studies have shown that there is a significant increase in ventricle size in the brains of affected individuals and that whole brain volume appears to be reduced up to 3% (Keshavan *et al.*, 2007).

1.2 Genetics of Schizophrenia

The genetic basis of schizophrenia is very complex and almost certainly polygenic in nature (Consortium *et al.*, 2009; Sullivan *et al.*, 2012). Twin and

family studies point to a heritability estimate in schizophrenia that is approximately 80% and 60%, respectively (Lichtenstein *et al*., 2009; Sullivan *et al*., 2003; Wray and Gottesman, 2012). A heritability estimate of less than 100% suggests that factors other than simple inheritance are playing a role. These additional factors could arise from environmental, epigenetic or *de novo* genetic insults. To date, a number of common variants with small effects have been identified by GWAS (Genome-Wide Association Studies) with increasing numbers of patients and controls. In addition, although the relative contributions of common and rare variants to this disease are not known, rare variants (mutations found at a low population frequency) have been estimated to account for about 20% of risk (Wright, 2014) and would be expected to have a relatively larger effect than a common variant. The morbid risk in families rises with increasing genetic relatedness to the index case, and different families, as well as different patients from the same family, may carry different susceptibility genes (Perala *et al*., 2007). However, to date, no one specific locus, rare or common, can be considered to be causal towards this heterogeneous and complex disease (McClellan and King, 2010).

Copy Number Variants (CNVs) are differences in the number of expected copies of a section of DNA and it has been suggested that *de novo* CNVs are up to eight times more frequent in schizophrenia cases versus controls (Xu *et al*., 2008). In fact, in general, schizophrenia patients appear to exhibit an increased burden of rare (<0.5%) variants when compared to controls (Kavanagh *et al*., 2014), and affected offspring seem to have increased levels of *de novo* genome-wide mutations as compared to their unaffected siblings (Xu *et al*., 2011). An elevation in *de novo* mutation rate in schizophrenia may explain the development of schizophrenia despite strong negative selection and significantly reduced fecundity of patients with schizophrenia (Nimgaonkar, 1998). Given this complexity, it is not surprising that a causal association or linkage to schizophrenia has been difficult to establish (Arribas-Ayllon *et al*., 2010).

In recent years, GWAS have shown the important contribution of genetic variants to complex human traits and disease susceptibility at the population level (Ripke *et al.*, 2014; Rudan, 2010; Stranger *et al.*, 2011). By comparing thousands of variants between cases and controls at a population level, GWAS identifies disease-associated Single Nucleotide Variants (SNVs) and regions of the genome that may elevate disease risk (Ott and Wang, 2011; Stranger *et al.*, 2011). However, the heterogeneity of genomic differences at the population level, often between entirely unrelated individuals, may be limiting (Ott and Wang, 2011; Rudan, 2010). In contrast, family studies reduce the heterogeneity and narrow the pool of genomic differences that may account for the phenotypic difference under study. The new wave of GWAS data has far exceeded the search for schizophrenia candidate genes by linkage and candidate gene studies, which produced results that were inconsistent and difficult to reproduce. However, most of the variants identified by GWAS have been common variants, that is, variants with allele frequencies greater than or equal to 1%. Some of the most interesting and consistent GWAS findings have highlighted genes involved in immunity including the major histocompatibility complex (MHC) on Chromosome 6 and the genes *TCF4*, *NRGN* and *DPYD/MIR317* (Chen *et al.*, 2015). These findings are not surprising given that the immune system has long been thought to be involved in schizophrenia pathology (Chen *et al.*, 2015). In addition to these loci, the Psychiatric Genomics Consortium (PGC) recently published the largest schizophrenia GWAS to date of 150,000 individuals including over 35,000 affected with schizophrenia, which also represents the largest genetic study on the disease that has currently been performed (Ripke *et al.*, 2014). This collaboration of over 80 research groups identified 108 common genetic loci overrepresented in schizophrenia patients, covering 348 genes, 83 of which had not been previously identified (Ripke *et al.*, 2014). These loci included a large list of potential genes, most notably *DRD2*, a common anti-psychotic target, as well as a number of glutamate receptors (*GRIA1*, *GRIN2A*, *GRM3*) (Ripke *et al.*, 2014). In addition, many CNVs have been associated with

schizophrenia risk, the most notable being deletions at 2p16.3 (overlapping the *NRXN1* gene) and duplications at 7p36.3 (overlapping the *VIPR2* gene) (Levinson *et al*., 2011; Rujescu *et al*., 2009). Other regions identified to harbour structural variations (SVs) contributing to risk include deletions and duplications at 1q21.1 and 3q29, as well as deletions at 15q13.3 and 22q11.2 and duplications at 16p13.1 (Chen *et al*., 2015; Levinson *et al*., 2011; Sullivan *et al*., 2012). However, it has been shown that a polygenic burden exists in schizophrenia patients arising from rare disruptive mutations primarily comprising ultra-rare nonsense mutations distributed across many genes (Purcell *et al*., 2014) and increased studies on rare variants are likely to uncover the missing heritability not yet identified.

It is noteworthy that the earliest definitive report of rare variation contributing to schizophrenia includes copy number variability at 22q11.2 which confers close to a 25-fold increase in risk for schizophrenia (Karayiorgou *et al*., 1995; Malhotra and Sebat, 2012). This suggest that structural variability may play a large role in the cause of this disease. To date, at least 11 rare risk alleles have been identified to be associated with this disorder (Rees *et al*., 2014).

Evidence is now emerging for the effect of *de novo* mutations in schizophrenia. Identification of genetic variants in schizophrenia patients has also given some insight into the biological processes that might be affected in schizophrenia. Neurodevelopmental genes appear to be preferentially represented in the gene sets identified by multiple groups; specifically, synaptic strength modulation at glutamatergic synapses appears to have been strongly implicated in schizophrenia (Kirov *et al*., 2012; Walsh *et al*., 2008). Together, these recent advances represent a new era in the study of schizophrenia genetics and there is no doubt that the advances in genetic technology such as high throughput sequencing are in part responsible for these successes (Kavanagh *et al*., 2014).

These associations of regions of interest with this complex disease are helping the field to inch closer to a clearer understanding of the genetics

underlying predisposition and/or causation in schizophrenia. In addition, these studies reinforce the highly polygenetic nature of schizophrenia. It has been estimated that the variants that have been previously identified account for >50% of the heritability of schizophrenia (Lee *et al*., 2012). However, there are still unidentified risk genes for schizophrenia and continuing to increase sample size may not uncover them. Alternative approaches should be considered, including the use of families with monozygotic twins discordant for schizophrenia. In addition, the accumulation of results on schizophrenia genetics suggests that the insults responsible for schizophrenia in individual patients may not overlap, which implies the need for patient specific inquiry and treatment.

1.3 Epigenetics of Schizophrenia

It is now apparent that the manifestation of the genetic code into psychiatric phenotypes including mental disorders is not determined solely by DNA sequence (Bell and Saffery, 2012; McCarthy and Hirschhorn, 2008). The causation of psychiatric disorders likely involves complex interactions involving chromatin, where epigenetic signals superimpose a regulatory role. In fact, it has been suggested that the missing heritability seen in some neuropsychiatric disorders could be due in part to the effect of epigenetic patterning (Bell and Saffery, 2012; McCarthy and Hirschhorn, 2008). The epigenome is in a dynamic state influenced by both deterministic as well as stochastic processes. This complexity also makes it difficult to tease apart the underlying factors that contribute to its state at any given time (Mill *et al*., 2008). It represents a major challenge for future studies. DNA methylation in mammals involves the modification of cytosine to methylated cytosine (or its equivalent) in the genome and is the most widely studied of epigenetic marks. This form of gene regulation is sequence-specific and is needed for the proper functioning of the genome. In addition, DNA methylation provides regulatory roles in cellular functioning via regulation of gene transcription (Razin and Riggs, 1980), genomic imprinting (Li *et al*., 1993), gene splicing (Shukla *et al*., 2011) and chromatin structure and

stability (Xu *et al*., 1999). Indeed, any alteration from normal patterns of methylation may cause abnormal cellular functioning including disease phenotypes (Zhao *et al*., 2014). DNA methylation profiles can potentially be altered by various factors including seasonal, social and additional environmental factors such as chemicals and prescription drugs (Alvarado *et al*., 2014; Singh *et al*., 2003). This property may help to further the understanding of disease processes including mechanisms of actions of drugs which are often used to treat this disease.

Studies on brains from patients with schizophrenia and matched controls have identified differences in DNA methylation (Wockner *et al*., 2014). The results on brain samples are comparable to similar studies on blood samples from schizophrenia patients (Wockner *et al*., 2014). The questions of both tissue specificity and the effect of antipsychotic drugs on epigenetics are critical, and present concerns in studies on methylation involving human brain disorders. Some non-brain tissues may serve as markers for abnormalities in the brain (Singh *et al*., 2002a). Further, the genes affected by enriched methylation changes in patients are related to a number of pathways particularly the glutamatergic and GABAergic neurotransmission pathways, which have been previously implicated in psychosis (Bonsch *et al*., 2012; Shimabukuro *et al*., 2006). These findings suggest that a common epigenetic regulation mechanism may be applicable both in the brain and in peripheral tissues of schizophrenia patients (Auta *et al*., 2013). However, it is not entirely clear if the observed changes in methylation in these studies represent epigenetic changes that promote the disorder or are reflective of changes induced by the drugs administered to treat the patients (Melka *et al*., 2014). The results of studies involving medication-free schizophrenia patients show that some of the altered DNA methylation seen in schizophrenia is likely to be directly involved in the pathophysiology of the disease. For example, a recent study on medication-free patients suggests that methylation is indeed a part of the complexity of manifestation of schizophrenia (Kinoshita *et al*., 2013).

Conclusions drawn from genome-wide association studies are compatible with methylation studies that have reported hypermethylation of the serotonin transporter gene promoter particularly in schizophrenia patients (Abdolmaleky *et al.*, 2014). Some pathways are shared by a number of neurological disorders. As an example, CDK5 and CREB signaling is often reported in schizophrenia (Allen *et al.*, 2008), in excessive anxiety induced by stress (Bignante *et al.*, 2008), as well as in depressive-like behavior (Zhu *et al.*, 2012). Interestingly, altered DNA methylation of genes involved in the CREB pathway has been reported in a recent study of DNA methylation in schizophrenia (Yu *et al.*, 2014). Further, *DRD2, DRD4* and *DRD5* promoters were significantly methylated in schizophrenia patients as compared with healthy controls (Kordi-Tamandani *et al.*, 2013), suggesting that the dopamine network is actively involved in an increased risk for psychosis. Thus, although only a small number of studies have been performed on the epigenetics of schizophrenia, alterations in DNA methylation appear to be important to the etiology of psychosis.

1.4 Genetic Differences Between Monozygotic Twins

Monozygotic twins result from the division of the morula during development. In some cases, the morula divides before day 5 post-fertilization, in other cases, twinning occurs after the separation of the morula (between day 5 and 9), the latter being more common. In rare exceptions, twins are formed after day 9 and will end up sharing only one placenta. Genetic mutations that are seen in one twin but not in the other identical twin are presumed to have occurred after the separation of the morula (Weber-Lehmann *et al.*, 2014).

Although MZ twins originate from a single zygote, many reports have now identified genetic differences between them (Bruder *et al.*, 2008; Ehli *et al.*, 2012; Maiti *et al.*, 2011). For example, Bruder *et al* (2008) reported that all of 19 MZ twins studied differed in CNVs (Bruder *et al.*, 2008). Similarly, *de novo* CNVs can be identified by comparison of the genomes of MZ twins with their parents (Maiti *et al.*, 2011). Further, a duplication found only in the affected monozygotic twin in

a tissue specific manner has been implicated in the causation of Mayer-Rokitansky-Kuster-Hauser syndrome (Rall *et al*., 2015). In addition, a recent report identified somatic mutations at the base pair level in monozygotic twins and found two *de novo* somatic mutations that occurred early in embryonic development (Li *et al*., 2014). In fact, many post-twinning single nucleotide mutations have been reported (Kondo *et al*., 2002; Reumers *et al*., 2012; Sakuntabhai *et al*., 1999; Vadlamudi *et al*., 2010; Ye *et al*., 2013), however these are expected to be rarer than post-twinning CNVs.

Many studies have looked at identifying CNV differences between monozygotic twins. The largest study of this kind looked at CNVs in 159 monozygotic pairs and confirmed 10 post-twinning mutations (Forsberg *et al*., 2012). It is likely that older twin pairs have an increase in somatic mutations based solely on *de novo* mutation rates and accumulation of mutations over the lifespan (Ye *et al*., 2013). The rate is also likely to be tissue specific (Piotrowski *et al*., 2008). Lending further support, Dal *et al* found that early post-zygotic mutations exist in humans and arise from early mitotic events occurring during embryogenesis leading to genome mosaicism (multiple genotypes arising from a single fertilized egg) (Dal *et al*., 2014). This genomic mosaicism could be an important contributor to the development of disease.

1.5 Epigenetic Differences Between Monozygotic Twins

Differences in methylation between identical twins have been identified as early as in newborn twin pairs (Ollikainen *et al*., 2010). DNA methylation profiles are more similar within pairs of dichorionic twins than between monochorionic twins and this suggests that sharing a placenta may cause imbalanced *in utero* conditions leading to epigenetic differences (Castillo-Fernandez *et al*., 2014). Also, these differences change over time, supporting the potential for neurodevelopmental programming and reprogramming in the causation of disease (Dempster *et al*., 2011). It has been suggested that contributions to the discordance of monozygotic twins may involve epigenetic change (Kim *et al*.,

2014), strengthening the case for dynamic processes including DNA methylation in psychosis. These processes are likely directed by genetic as well as random and environmental contributors over the lifetime (Wong *et al*., 2010).

Results on DNA methylation analysis of blood DNA from monozygotic twins discordant for schizophrenia support the involvement of DNA methylation in psychosis (Dempster *et al*., 2011). Methylation of genomic DNA and promoter methylation of specific genes in blood samples of twins discordant for schizophrenia showed hypermethylation and hypomethylation of several genes (Bonsch *et al*., 2012). The results on twins to-date argue for the potential involvement of regulatory mechanisms, particularly DNA methylation, in the development of disease (Singh *et al*., 2002b). These mechanisms may underlie aberrations in neurodevelopment known to exist in a number of mental disorders (Rapoport *et al*., 2012; Singh *et al*., 2004; Weinberger, 1996).

1.6 Evidence for Somatic Mutation

Evidence has emerged in the literature to support the existence of somatic mutations in individual genomes. Generally, without involving identical twins, the estimated number of SNV substitution mutations per generation ranges from 1×10^{-8} - 3×10^{-8} per human single base pair which is equal to approximately 10–40 expected SNVs per generation (Krawczak *et al*., 2012; Xue *et al*., 2009). Although little is known about the exact mechanism by which somatic mutations may arise, it is suspected that they occur randomly and as a by-product of replication. This would make the occurrence of the same somatic mutation, in the same tissue, in different individuals, a rare occurrence. If post-zygotic mutations occur late in development, they would be expected to result in mosaics. Given the trillions of cell divisions occurring, post-zygotic replication errors are likely to be present in all individuals. The impact of these errors will most often be benign, however, the likelihood for contribution to pathology will depend on the tissue affected. Somatic mutations may reshape the genetic circuitry that underpins normal and abnormal neurobiological processes (Baillie *et al*., 2011). As we

know that not all mutations are deleterious, some may represent normal aspects of genetic diversity and development in some organs (Muotri and Gage, 2006) and cell types (Muotri *et al*., 2010). It follows then, that rare and unique somatic mutations should exist between the genomes of identical twins.

Accordingly, somatic point mutations have been shown to occur during early development with a frequency of 1.2×10^{-7} per base pair per twin pair (Li *et al*., 2014). This somatic point mutation frequency does not take into account the frequency of lesions (deletions, insertions, indels) which are expected to occur at a ratio of approximately one in three relative to SNVs (Krawczak *et al*., 2012). The earlier in the development of an embryo a rare somatic mutation arises, the more tissues it will be seen in (Weber-Lehmann *et al*., 2014). There is also potential for these somatic mutations to exist in the germline. Krawczak *et al* have proposed that SNVs, though rare, will exist and discriminate the genetic landscape of any pair of monozygotic twins (Krawczak *et al*., 2012). However, many have also failed to find any confirmable differences between MZ twins at the SNV level, presumably because of low Next Generation Sequencing (NGS) coverage (Baranzini *et al*., 2010; Handunnetthi *et al*., 2010; Miyake *et al*., 2013). Results also show that where DNA sequence differences exist between identical twins, genotypes usually show unexpected ratios, that is, instead of identifying expected heterozygote ratios between twins (e.g., 50/50 ratio of original and mutation genotype in one twin), more often a ratio that resembles for example an 80/20 split between the original and new mutation genotype, respectively (Weber-Lehmann *et al*., 2014) is ascertained by Sanger Sequencing. The best possible explanation for the identification of a newly arisen allele in only a small fraction of cells assayed (ie. 20%) is somatic mosaicism.

1.7 Advancing Technology in the Era of Genomics

As geneticists, our understanding of genomic change is driven by technology. Technological advances have been growing exponentially in recent years. As such, the resolution with which we can ascertain understanding of

genomic architecture is constantly increasing. Never before has the field had the ability to fine-map individual genomes as we currently can. The resolution has increased substantially and the methods available to analyze the datasets are becoming increasingly sophisticated.

This thesis utilizes three major technologies: the first, a comprehensive DNA Microarray; the second, Complete Genome Sequencing; and the third, a MeDIP Promoter Microarray. In addition, Real-Time PCR and Sanger Sequencing were used to confirm selected findings.

Although the specific technologies utilized will be discussed in further chapters, I briefly present here the rationale for the choice of each platform.

1.7.1 Affymetrix® Human SNP 6.0 Microarray

The Affymetrix SNP 6.0 microarray is one of the most comprehensive arrays available for hybridization with human DNA. At the time of purchase, this array had the highest resolution on the market with over 1.8 million genetic markers on each array. In addition, the average inter-marker distance of less than 700 base pairs makes it an excellent array for breakpoint estimation. In addition, this microarray is compatible with many third-party software packages and algorithms.

1.7.2 Complete Genomics Inc. Whole Genome Sequencing

Complete Genomics has significant experience in sequencing high-quality human genomes (Drmanac *et al.*, 2010). The platform utilizes efficient imaging with a combinatorial probe anchor ligation chemistry to independently assay each base from patterned nanoarrays of self-assembling DNA nanoballs. The high-fold coverage (~50 fold) that it achieves alongside high accuracy calls (99.999% accuracy) at single base pair resolution made this platform the ideal choice for this study.

1.7.3 NimbleGen Human DNA Methylation 3x720k CpG Island Plus RefSeq Promoter Microarray

This array utilizes a MeDIP (Methylated DNA immunoprecipitation) format and is a highly comprehensive assessment of CpG islands and promoter regions of RefSeq genes in humans. The array also includes a number of positive, negative and non-CpG control regions for quality control. In addition, the ability to ascertain over 700,000 regions of interest for methylation status allows for ideal comparison between samples in a study.

1.8 Hypothesis and Thesis Objectives

The collective research experiments presented in this thesis seek to assess the following hypothesis through the listed objectives:

Genome-wide de novo mutations and epimutations contribute to genomic individuality and may explain the discordance of monozygotic twins for schizophrenia.

1.8.1 Objectives

1. To use new and emerging technologies to identify Single Nucleotide Variants, Small Indels, Block Substitutions, Structural Variants (Insertions, Deletions, Tandem Duplications, Distal Duplications, Inversions, Interchromosomal Events and Copy Number Variants), and Methylation Changes in blood samples. Specifically through the use of the:
    a. Affymetrix® Human SNP 6.0 Microarray
    b. Complete Genomics Inc. Whole Genome Sequencing platform
    c. NimbleGen Human DNA Methylation 3x720k CpG Island Plus RefSeq Promoter Microarray
2. To analyze data obtained from the three platforms using multiple methods.

3. To identify differences between participants, specifically between monozygotic twins and their respective parents (where available) and annotate differences as *de novo* or inherited (where parental samples are available).
4. To characterize *de novo* differences between twins based on their potential involvement in schizophrenia discordance or predisposition.
5. To confirm a subset of novel results.
6. To test a model that may explain the results identified by the collective studies.

1.9 Overview of Participants in Thesis Experiments

This thesis focuses on the use of rare monozygotic twins who show discordance for schizophrenia. The research outlined in this thesis took place after ethics approval from The University of Western Ontario's Committee on research involving human subjects (Appendix B). All subjects provided written informed consent to participate in the study. Our close collaboration with Dr. Richard O'Reilly, Psychiatrist, has allowed for the acquisition of DNA from these rare twins and selected parental samples. Dr. O'Reilly has assured accurate diagnosis and has kept in contact with the families since sample collection (>5 years ago) to ensure that throughout the study the twins have remained discordant. However, it should be noted that previous studies have established that the vast majority of twin pairs that become concordant for schizophrenia will do so within five years of initial onset of the illness by the first affected twin (Belmaker *et al*., 1974).

These thesis experiments include six pairs of monozygotic twins discordant for schizophrenia and two sets of parents. The following table describes which participants were used in each chapter of this thesis (Table 1.1).

Table 1.1. Summary of participants included in the thesis experiments.

<table>
<tr><th>Twin Pair</th><th>Affymetrix 6.0 DNA Microarray (N=16) [Chapters 2,5]</th><th>Complete Genomics Sequencing (N=6) [Chapters 3,5]</th><th>Roche Nimblegen MeDIP Promoter Array (N=8) [Chapters 4,5]</th><th>Other Identifier [Chapters 3, 4, 5]</th></tr>
<tr><td>Twin Pair 1</td><td>YES</td><td></td><td></td><td></td></tr>
<tr><td>Twin Pair 2</td><td>YES</td><td>YES</td><td>YES</td><td rowspan="2">Family 2</td></tr>
<tr><td>Parents of Twin Pair 2</td><td>YES</td><td></td><td>YES</td></tr>
<tr><td>Twin Pair 3</td><td>YES</td><td>YES</td><td>YES</td><td rowspan="2">Family 1</td></tr>
<tr><td>Parents of Twin Pair 3</td><td>YES</td><td>YES</td><td>YES</td></tr>
<tr><td>Twin Pair 4</td><td>YES</td><td></td><td></td><td></td></tr>
<tr><td>Twin Pair 5</td><td>YES</td><td></td><td></td><td></td></tr>
<tr><td>Twin Pair 6</td><td>YES</td><td></td><td></td><td></td></tr>
</table>

Twin Pair 1

Members of twin pair 1 were assessed at age 20. These twins are male and self-declare their background to be Asian/Indian. The affected male patient of twin pair 1 was diagnosed with a psychotic disorder (not otherwise specified) and mild obsessive-compulsive disorder, at age 19. The unaffected co-twin has never been diagnosed with any mental health disorder. The twins remain discordant for psychosis >5 years after assessment.

Twin Pair 2 (Family 2)

Members of twin pair 2 were assessed at age 43. These twins are female and self-declare their background to be Caucasian. The affected female patient of twin pair 2 was diagnosed with schizoaffective disorder at age 27. The unaffected co-twin had a single episode of major depression at age 18. The twins remain discordant for psychosis >5 years after assessment.

Parents, Twin Pair 2

Father, Twin Pair 2: Caucasian male who was 80 when the samples for DNA testing were obtained. He has never been treated for any emotional or psychiatric disorders.

Mother, Twin Pair 2: Caucasian female who was 76 when the samples for DNA testing were obtained. She has never been treated for any emotional or psychiatric disorders.

Twin Pair 3 (Family 1)

Members of twin pair 3 were assessed at age 53. These twins are female and self-declare their background to be African American. The affected female patient of twin pair 3 was diagnosed with paranoid schizophrenia at age 22. The unaffected co-twin was diagnosed with bipolar I disorder at age 52. The twins remain discordant for psychosis >5 years after assessment.

Parents, Twin Pair 3

Father, Twin Pair 3: Afro-American male who was 82 when the samples for DNA testing were obtained. He had never been treated for any emotional or psychiatric disorders though he did exhibit a mild obsessive-compulsive personality upon assessment. He was diagnosed with Chronic Leukemia (CLL) at age 69.

Mother, Twin Pair 3: Afro-American female who was 74 when the samples for DNA testing were obtained. She has never been treated for any emotional or psychiatric disorders.

Twin Pair 4

Members of twin pair 4 were assessed at age 22. These twins are female and self-declare their background to be Caucasian. The affected female patient of twin pair 4 was diagnosed with paranoid schizophrenia at age 18. The unaffected co-twin had a single episode of major depression at age 20. The twins remain discordant for psychosis >5 years after assessment.

Twin Pair 5

Members of twin pair 5 were assessed at age 36. These twins are male and self-declare their background to be Caucasian. The affected male patient of twin pair 5 was diagnosed with undifferentiated schizophrenia at age 20. The unaffected co-twin has never been diagnosed with any mental health disorder. The twins remain discordant for psychosis >5 years after assessment.

Twin Pair 6

Members of twin pair 6 were assessed at age 42. These twins are male and self-declare their background to be Caucasian. The affected male patient of twin pair 6 was diagnosed with paranoid schizophrenia at age 16. The unaffected co-twin had a single episode of major depression at age 17. The twins remain discordant for psychosis >5 years after assessment.

1.10 References

Abdolmaleky, H.M., Nohesara, S., Ghadirivasfi, M., Lambert, A.W., Ahmadkhaniha, H., Ozturk, S., Wong, C.K., Shafa, R., Mostafavi, A., and Thiagalingam, S. (2014). DNA hypermethylation of serotonin transporter gene promoter in drug naïve patients with schizophrenia. Schizophr. Res. *152*, 373–380.

Allen, N.C., Bagade, S., McQueen, M.B., Ioannidis, J.P., Kavvoura, F.K., Khoury, M.J., Tanzi, R.E., and Bertram, L. (2008). Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. Nat. Genet. *40*, 827–834.

Alvarado, S., Fernald, R.D., Storey, K.B., and Szyf, M. (2014). The dynamic nature of DNA methylation: a role in response to social and seasonal variation. Integr. Comp. Biol. *54*, 68–76.

Arribas-Ayllon, M., Bartlett, A., and Featherstone, K. (2010). Complexity and accountability: the witches' brew of psychiatric genetics. Soc. Stud. Sci. *40*, 499–524.

Auta, J., Smith, R.C., Dong, E., Tueting, P., Sershen, H., Boules, S., Lajtha, A., Davis, J., and Guidotti, A. (2013). DNA-methylation gene network dysregulation in peripheral blood lymphocytes of schizophrenia patients. Schizophr. Res. *150*, 312–318.

Baillie, J.K., Barnett, M.W., Upton, K.R., Gerhardt, D.J., Richmond, T.A., De Sapio, F., Brennan, P.M., Rizzu, P., Smith, S., Fell, M., *et al.* (2011). Somatic retrotransposition alters the genetic landscape of the human brain. Nature *479*, 534–537.

Baranzini, S.E., Mudge, J., van Velkinburgh, J.C., Khankhanian, P., Khrebtukova, I., Miller, N.A., Zhang, L., Farmer, A.D., Bell, C.J., Kim, R.W., *et al.* (2010). Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. Nature *464*, 1351–1356.

Bassett, A.S., Bury, A., Hodgkinson, K.A., and Honer, W.G. (1996). Reproductive fitness in familial schizophrenia. Schizophr. Res. *21*, 151–160.

Bell, J.T., and Saffery, R. (2012). The value of twins in epigenetic epidemiology. Int. J. Epidemiol. *41*, 140–150.

Belmaker, R., Pollin, W., Wyatt, R.J., and Cohen, S. (1974). A follow-up of monozygotic twins discordant for schizophrenia. Arch. Gen. Psychiatry *30*, 219–222.

Betts, K.S., Williams, G.M., Najman, J.M., Scott, J., and Alati, R. (2014). Maternal prenatal infection, early susceptibility to illness and adult psychotic experiences: a birth cohort study. Schizophr. Res. *156*, 161–167.

Bignante, E.A., Rodriguez Manzanares, P.A., Mlewski, E.C., Bertotto, M.E., Bussolino, D.F., Paglini, G., and Molina, V.A. (2008). Involvement of septal CDK5 in the emergence of excessive anxiety induced by stress. Eur. Neuropsychopharmacol. *18*, 578–588.

Bonsch, D., Wunschel, M., Lenz, B., Janssen, G., Weisbrod, M., and Sauer, H. (2012). Methylation matters? Decreased methylation status of genomic DNA in the blood of schizophrenic twins. Psychiatry Res. *198*, 533–537.

Bruder, C.E.G., Piotrowski, A., Gijsbers, A.A.C.J., Andersson, R., Erickson, S., Diaz de Ståhl, T., Menzel, U., Sandgren, J., von Tell, D., Poplawski, A., *et al.* (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am. J. Hum. Genet. *82*, 763–771.

Byrne, M., Agerbo, E., Ewald, H., Eaton, W.W., and Mortensen, P.B. (2003). Parental age and risk of schizophrenia: a case-control study. Arch. Gen. Psychiatry *60*, 673–678.

Cañas, F., Ciudad, A., Gutiérrez, M., Gibert, J., Gascón, J., Carrasco, J.L., Bobes, J., Gómez, J.C., and Alvarez, E. (2005). Safety, effectiveness, and patterns of use of olanzapine in acute schizophrenia: a multivariate analysis of a large naturalistic study in the hospital setting. Med. Clin. (Barc). *124*, 481–486.

Castillo-Fernandez, J., Spector, T., and Bell, J. (2014). Epigenetics of discordant monozygotic twins: implications for disease. Genome Med. *6*.

Chen, J., Cao, F., Liu, L., Wang, L., and Chen, X. (2015). Genetic studies of schizophrenia: an update. Neurosci. Bull. *31*, 87–98.

Consortium, T.I.S., preparation, M., Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., Sklar, P., analysis, D., *et al.* (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature.

Dal, G.M., Ergüner, B., Sağıroğlu, M.S., Yüksel, B., Onat, O.E., Alkan, C., and Özçelik, T. (2014). Early postzygotic mutations contribute to *de novo* variation in a healthy monozygotic twin pair. J. Med. Genet. *51*, 455–459.

Delisi, L.E. (2009). Searching for the true genetic vulnerability for schizophrenia. Genome Med. *1*, 14.

Dempster, E.L., Pidsley, R., Schalkwyk, L.C., Owens, S., Georgiades, A., Kane, F., Kalidindi, S., Picchioni, M., Kravariti, E., Toulopoulou, T., *et al*. (2011). Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. Hum. Mol. Genet. *20*, 4786–4796.

Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., *et al*. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science *327*, 78–81.

Ehli, E.A., Abdellaoui, A., Hu, Y., Hottenga, J.J., Kattenberg, M., van Beijsterveldt, T., Bartels, M., Althoff, R.R., Xiao, X., Scheet, P., *et al*. (2012). *De novo* and inherited CNVs in MZ twin pairs selected for discordance and concordance on Attention Problems. Eur. J. Hum. Genet. *20*, 1037–1043.

First, M.B., Spitzer, R.L., Gibbon, M., and and Williams, J.B.W. (1996). Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV) (Washington, D.C.: American Psychiatric Press, Inc).

First, M.B., Gibbon, M., Spitzer, R.L., Williams, J.B.W., and Benjamin, L.S. (1997). Structured Clinical Interview for DSM-IV Axis II Personality Disorders, (SCID-II) (Washington, D.C.: American Psychiatric Press, Inc.).

Forsberg, L.A., Rasi, C., Razzaghian, H.R., Pakalapati, G., Waite, L., Thilbeault, K.S., Ronowicz, A., Wineinger, N.E., Tiwari, H.K., Boomsma, D., *et al*. (2012). Age-related somatic structural changes in the nuclear genome of human blood cells. Am. J. Hum. Genet. *90*, 217–228.

Goeree, R., Farahati, F., Burke, N., Blackhouse, G., O'Reilly, D., Pyne, J., and Tarride, J.E. (2005). The economic burden of schizophrenia in Canada in 2004. Curr. Med. Res. Opin. *21*, 2017–2028.

Handunnetthi, L., Handel, A.E., and Ramagopalan, S. V (2010). Contribution of genetic, epigenetic and transcriptomic differences to twin discordance in multiple sclerosis. Expert Rev. Neurother. *10*, 1379–1381.

Jablensky, A., Sartorius, N., Ernberg, G., Anker, M., Korten, A., Cooper, J.E., Day, R., and Bertelsen, A. (1992). Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization ten-country study. Psychol. Med. Monogr. Suppl. *20*, 1–97.

Karayiorgou, M., Morris, M.A., Morrow, B., Shprintzen, R.J., Goldberg, R., Borrow, J., Gos, A., Nestadt, G., Wolyniec, P.S., and Lasseter, V.K. (1995). Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. Proc. Natl. Acad. Sci. U. S. A. *92*, 7612–7616.

Kavanagh, D.H., Tansey, K.E., O'Donovan, M.C., and Owen, M.J. (2014). Schizophrenia genetics: emerging themes for a complex disorder. Mol. Psychiatry *20*, 72–76.

Keshavan, M.S., Prasad, K.M., and Pearlson, G. (2007). Are brain structural abnormalities useful as endophenotypes in schizophrenia? Int. Rev. Psychiatry *19*, 397–406.

Khandaker, G.M., Zimbron, J., Lewis, G., and Jones, P.B. (2013). Prenatal maternal infection, neurodevelopment and adult schizophrenia: a systematic review of population-based studies. Psychol. Med. *43*, 239–257.

Kim, K., Ban, H.J., Seo, J., Lee, K., Yavartanoo, M., Kim, S.C., Park, K., Cho, S.B., and Choi, J.K. (2014). Genetic factors underlying discordance in chromatin accessibility between monozygotic twins. Genome Biol. *15*, R72.

Kinoshita, M., Numata, S., Tajima, A., Shimodera, S., Ono, S., Imamura, A., Iga, J., Watanabe, S., Kikuchi, K., Kubo, H., *et al*. (2013). DNA methylation signatures of peripheral leukocytes in schizophrenia. Neuromolecular Med. *15*, 95–101.

Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., *et al*. (2012). *De novo* CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol. Psychiatry *17*, 142–153.

Knapp, M., Mangalore, R., and Simon, J. (2004). The global costs of schizophrenia. Schizophr. Bull. *30*, 279–293.

Kondo, S., Schutte, B.C., Richardson, R.J., Bjork, B.C., Knight, A.S., Watanabe, Y., Howard, E., Ferreira de Lima, R.L.L., Daack-Hirsch, S., Sander, A., *et al*. (2002). Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat. Genet. *32*, 285–289.

Kordi-Tamandani, D.M., Sahranavard, R., and Torkamanzehi, A. (2013). Analysis of association between dopamine receptor genes' methylation and their expression profile with the risk of schizophrenia. Psychiatr. Genet. *23*, 183–187.

Krawczak, M., Cooper, D.N., Fändrich, F., Engel, W., and Schmidtke, J. (2012). How to distinguish genetically between an alleged father and his monozygotic twin: a thought experiment. Forensic Sci. Int. Genet. *6*, e129–e130.

Laursen, T.M., and Munk-Olsen, T. (2010). Reproductive patterns in psychotic patients. Schizophr. Res. *121*, 234–240.

Lee, S.H., DeCandia, T.R., Ripke, S., Yang, J., Sullivan, P.F., Goddard, M.E., Keller, M.C., Visscher, P.M., and Wray, N.R. (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat. Genet. *44*, 247–250.

Levinson, D.F., Duan, J., Oh, S., Wang, K., Sanders, A.R., Shi, J., Zhang, N., Mowry, B.J., Olincy, A., Amin, F., *et al*. (2011). Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am. J. Psychiatry *168*, 302–316.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362–365.

Li, R., Montpetit, A., Rousseau, M., Wu, S.Y.M., Greenwood, C.M.T., Spector, T.D., Pollak, M., Polychronakos, C., and Richards, J.B. (2014). Somatic point mutations occurring early in development: a monozygotic twin study. J. Med. Genet. *51*, 28–34.

Lichtenstein, P., Yip, B.H., Björk, C., Pawitan, Y., Cannon, T.D., Sullivan, P.F., and Hultman, C.M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet *373*, 234–239.

Maiti, S., Kumar, K.H.B.G., Castellani, C.A., O'Reilly, R., and Singh, S.M. (2011). Ontogenetic *de novo* copy number variations (CNVs) as a source of genetic individuality: studies on two families with MZD twins for schizophrenia. PLoS One *6*, e17125.

Malhotra, D., and Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell *148*, 1223–1241.

McCarthy, M.I., and Hirschhorn, J.N. (2008). Genome-wide association studies: potential next steps on a genetic journey. Hum. Mol. Genet. *17*, R156–R165.

McClellan, J., and King, M.-C. (2010). Genomic analysis of mental illness: a changing landscape. JAMA *303*, 2523–2524.

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

Melka, M.G., Laufer, B.I., McDonald, P., Castellani, C.A., Rajakumar, N., O'Reilly, R., and Singh, S.M. (2014). The effects of olanzapine on genome-wide DNA methylation in the hippocampus and cerebellum. Clin. Epigenetics *6*, 1.

Mill, J., Tang, T., Kaminsky, Z., Khare, T., Yazdanpanah, S., Bouchard, L., Jia, P., Assadzadeh, A., Flanagan, J., Schumacher, A., *et al*. (2008). Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. Am. J. Hum. Genet. *82*, 696–711.

Miyake, K., Yang, C., Minakuchi, Y., Ohori, K., Soutome, M., Hirasawa, T., Kazuki, Y., Adachi, N., Suzuki, S., Itoh, M., *et al*. (2013). Comparison of Genomic and Epigenomic Expression in Monozygotic Twins Discordant for Rett Syndrome. PLoS One *8*, e66729.

Moran, P. a. P. (1965). Schizophrenia: Schizophrenia as a Genetic Morphism. Nature *206*, 1113–1113.

Mortensen, P.B., Pedersen, M.G., and Pedersen, C.B. (2010). Psychiatric family history and schizophrenia risk in Denmark: which mental disorders are relevant? Psychol. Med. *40*, 201–210.

Muotri, A.R., and Gage, F.H. (2006). Generation of neuronal variability and complexity. Nature *441*, 1087–1093.

Muotri, A.R., Marchetto, M.C.N., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K., and Gage, F.H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. Nature *468*, 443–446.

Nielsen, P.R., Laursen, T.M., and Mortensen, P.B. (2013). Association between parental hospital-treated infection and the risk of schizophrenia in adolescence and early adulthood. Schizophr. Bull. *39*, 230–237.

Nimgaonkar, V.L. (1998). Reduced fertility in schizophrenia: here to stay? Acta Psychiatr. Scand. *98*, 348–353.

Ollikainen, M., Smith, K.R., Joo, E.J., Ng, H.K., Andronikos, R., Novakovic, B., Abdul Aziz, N.K., Carlin, J.B., Morley, R., Saffery, R., *et al*. (2010). DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. Hum. Mol. Genet. *19*, 4176–4188.

Ott, J., and Wang, J. (2011). Multiple phenotypes in genome-wide genetic mapping studies. Protein Cell *2*, 519–522.

Perala, J., Suvisaari, J., Saarni, S.I., Kuoppasalmi, K., Isometsa, E., Pirkola, S., Partonen, T., Tuulio-Henriksson, A., Hintikka, J., Kieseppa, T., *et al*. (2007). Lifetime prevalence of psychotic and bipolar I disorders in a general population. Arch. Gen. Psychiatry *64*, 19–28.

Piotrowski, A., Bruder, C.E.G., Andersson, R., Diaz de Ståhl, T., Menzel, U., Sandgren, J., Poplawski, A., von Tell, D., Crasto, C., Bogdan, A., *et al*. (2008). Somatic mosaicism for copy number variation in differentiated human tissues. Hum. Mutat. *29*, 1118–1124.

Pompili, M., Amador, X.F., Girardi, P., Harkavy-Friedman, J., Harrow, M., Kaplan, K., Krausz, M., Lester, D., Meltzer, H.Y., Modestin, J., *et al*. (2007). Suicide risk in schizophrenia: learning from the past to change the future. Ann. Gen. Psychiatry *6*, 10.

Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S.E., Kähler, A., *et al*. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. Nature *506*, 185–190.

Rall, K., Eisenbeis, S., Barresi, G., Rückner, D., Walter, M., Poths, S., Wallwiener, D., Riess, O., Bonin, M., and Brucker, S. (2015). Mayer-Rokitansky-Küster-Hauser syndrome discordance in monozygotic twins: matrix metalloproteinase 14, low-density lipoprotein receptor-related protein 10, extracellular matrix, and neoangiogenesis genes identified as candidate genes in a tissue-specific . Fertil. Steril. *103*, 494–502.e3.

Rapoport, J.L., Giedd, J.N., and Gogtay, N. (2012). Neurodevelopmental model of schizophrenia: update 2012. Mol. Psychiatry *17*, 1228–1238.

Razin, A., and Riggs, A.D. (1980). DNA methylation and gene function. Science *210*, 604–610.

Rees, E., Walters, J.T.R., Georgieva, L., Isles, A.R., Chambert, K.D., Richards, A.L., Mahoney-Davies, G., Legge, S.E., Moran, J.L., McCarroll, S.A., *et al*. (2014). Analysis of copy number variations at 15 schizophrenia-associated loci. Br. J. Psychiatry *204*, 108–114.

Regier, D.A., Narrow, W.E., Rae, D.S., Manderscheid, R.W., Locke, B.Z., and Goodwin, F.K. (1993). The de facto US mental and addictive disorders service system. Epidemiologic catchment area prospective 1-year prevalence rates of disorders and services. Arch. Gen. Psychiatry *50*, 85–94.

Reumers, J., De Rijk, P., Zhao, H., Liekens, A., Smeets, D., Cleary, J., Van Loo, P., Van Den Bossche, M., Catthoor, K., Sabbe, B., *et al*. (2012). Optimized

filtering reduces the error rate in detecting genomic variants by short-read sequencing. Nat. Biotechnol. *30*, 61–68.

Ripke, S., Neale, B.M., Corvin, A., Walters, J.T.R., Farh, K.-H., Holmans, P.A., Lee, P., Bulik-Sullivan, B., Collier, D.A., Huang, H., *et al*. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

Rudan, I. (2010). New technologies provide insights into genetic basis of psychiatric disorders and explain their co-morbidity. Psychiatr. Danub. *22*, 190–192.

Rujescu, D., Ingason, A., Cichon, S., Pietiläinen, O.P.H., Barnes, M.R., Toulopoulou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E., *et al*. (2009). Disruption of the neurexin 1 gene is associated with schizophrenia. Hum. Mol. Genet. *18*, 988–996.

Saha, S., Chant, D., Welham, J., and McGrath, J. (2005). A systematic review of the prevalence of schizophrenia. PLoS Med. *2*, e141.

Sakuntabhai, A., Burge, S., Monk, S., and Hovnanian, A. (1999). Spectrum of novel ATP2A2 mutations in patients with Darier’s disease. Hum. Mol. Genet. *8*, 1611–1619.

Selten, J.-P., van der Ven, E., Rutten, B.P.F., and Cantor-Graae, E. (2013). The social defeat hypothesis of schizophrenia: an update. Schizophr. Bull. *39*, 1180–1186.

Shimabukuro, M., Jinno, Y., Fuke, C., and Okazaki, Y. (2006). Haloperidol treatment induces tissue- and sex-specific changes in DNA methylation: a control study using rats. Behav. Brain Funct. *2*, 37.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74–79.

Singh, S.M., Murphy, B., and O’Reilly, R. (2002a). Monozygotic twins with chromosome 22q11 deletion and discordant phenotypes: updates with an epigenetic hypothesis. J. Med. Genet. *39*, e71.

Singh, S.M., Murphy, B., and O’Reilly, R. (2002b). Epigenetic contributors to the discordance of monozygotic twins. Clin. Genet. *62*, 97–103.

Singh, S.M., Murphy, B., and O'Reilly, R.L. (2003). Involvement of gene-diet/drug interaction in DNA methylation and its contribution to complex diseases: from cancer to schizophrenia. Clin. Genet. *64*, 451–460.

Singh, S.M., McDonald, P., Murphy, B., and O'Reilly, R. (2004). Incidental neurodevelopmental episodes in the etiology of schizophrenia: an expanded model involving epigenetics and development. Clin. Genet. *65*, 435–440.

Stranger, B.E., Stahl, E.A., and Raj, T. (2011). Progress and promise of genome-wide association studies for human complex trait genetics. Genetics *187*, 367–383.

Sullivan, P.F., Kendler, K.S., and Neale, M.C. (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch. Gen. Psychiatry *60*, 1187–1192.

Sullivan, P.F., Daly, M.J., and O'Donovan, M. (2012). Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat. Rev. Genet. *13*, 537–551.

Vadlamudi, L., Dibbens, L.M., Lawrence, K.M., Iona, X., McMahon, J.M., Murrell, W., Mackay-Sim, A., Scheffer, I.E., and Berkovic, S.F. (2010). Timing of *de novo* mutagenesis--a twin study of sodium-channel mutations. N. Engl. J. Med. *363*, 1335–1340.

Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., Malhotra, D., Bhandari, A., *et al.* (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science *320*, 539–543.

Weber-Lehmann, J., Schilling, E., Gradl, G., Richter, D.C., Wiehler, J., and Rolf, B. (2014). Finding the needle in the haystack: differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. Forensic Sci. Int. Genet. *9*, 42–46.

Weinberger, D.R. (1996). On the plausibility of "the neurodevelopmental hypothesis" of schizophrenia. Neuropsychopharmacology *14*, 1S – 11S.

Wockner, L.F., Noble, E.P., Lawford, B.R., Young, R.M., Morris, C.P., Whitehall, V.L., and Voisey, J. (2014). Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients. Transl. Psychiatry *4*, e339.

Wong, C.C., Caspi, A., Williams, B., Craig, I.W., Houts, R., Ambler, A., Moffitt, T.E., and Mill, J. (2010). A longitudinal study of epigenetic variation in twins. Epigenetics *5*, 516–526.

Wray, N.R., and Gottesman, I.I. (2012). Using summary data from the danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. Front. Genet. *3*, 118.

Wright, J. (2014). Genetics: Unravelling complexity. Nature *508*, S6–S7.

Xu, B., Roos, J.L., Levy, S., van Rensburg, E.J., Gogos, J.A., and Karayiorgou, M. (2008). Strong association of *de novo* copy number mutations with sporadic schizophrenia. Nat. Genet. *40*, 880–885.

Xu, B., Roos, J.L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., Gogos, J.A., and Karayiorgou, M. (2011). Exome sequencing supports a *de novo* mutational paradigm for schizophrenia. Nat. Genet. *43*, 864–868.

Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Péquignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature *402*, 187–191.

Xue, Y., Wang, Q., Long, Q., Ng, B.L., Swerdlow, H., Burton, J., Skuce, C., Taylor, R., Abdellah, Z., Zhao, Y., *et al*. (2009). Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. Curr. Biol. *19*, 1453–1457.

Ye, K., Beekman, M., Lameijer, E.-W., Zhang, Y., Moed, M.H., van den Akker, E.B., Deelen, J., Houwing-Duistermaat, J.J., Kremer, D., Anvar, S.Y., *et al*. (2013). Aging as accelerated accumulation of somatic variants: whole-genome sequencing of centenarian and middle-aged monozygotic twin pairs. Twin Res. Hum. Genet. *16*, 1026–1032.

Yu, M., Li, W., Luo, S., Zhang, Y., Liu, H., Gao, Y., Wang, X., Wilson, J.X., and Huang, G. (2014). Folic acid stimulation of neural stem cell proliferation is associated with altered methylation profile of PI3K/Akt/CREB. J. Nutr. Biochem. *25*, 496–502.

Zhao, J., Wang, F., Xu, Z., and Fan, Y. (2014). [The epigenetic effect on pre-mRNA alternative splicing]. Yi Chuan *36*, 248–255.

Zhu, W.-L., Shi, H.-S., Wang, S.-J., Xu, C.-M., Jiang, W.-G., Wang, X., Wu, P., Li, Q.-Q., Ding, Z.-B., and Lu, L. (2012). Increased Cdk5/p35 activity in the dentate gyrus mediates depressive-like behaviour in rats. Int. J. Neuropsychopharmacol. *15*, 795–809.

Chapter 2 - Assessing Copy Number Variation (CNV) in Monozygotic Twins Discordant for Schizophrenia Using the Affymetrix Human SNP 6.0 Microarray

2.0 Overview of Chapter 2

Studies involving the analysis of structural variation including Copy Number Variation (CNV) have recently exploded in the literature. Furthermore, CNVs have been associated with a number of complex diseases and neurodevelopmental disorders (Firth *et al*., 2009; Jacquemont *et al*., 2011; Zhang *et al*., 2011). Copy number variants have the potential to alter dosage of gene(s) in regions they overlap, thereby having the potential to modify the expression of genes (Henrichsen *et al*., 2009). The discovery of CNVs from DNA microarrays works on the assumption that the majority of the diploid genome exists in two copies and therefore the signal intensity of markers along the chromosome should remain approximately the same. When the signal intensity changes over multiple markers, it can be inferred that a copy number change may be present (Chen *et al*., 2015). This chapter uses Affymetrix Human SNP 6.0 Arrays on 16 individuals to assess the copy number variation calls between identical twins and their parents (where available).

This chapter is divided into two main sections:

2A: Explores the use of four software programs for the analysis of Affymetrix Human SNP Array 6.0 data and leverages the unique nature of monozygotic twins to assess the calls made by each software program.

2B: Reports the results of the use of three of the four software programs to find *de novo* variants that may be associated with the disease in monozygotic twins discordant for schizophrenia.

A version of this chapter has been published in Castellani *et al.*, *BMC Bioinformatics* (2014a)

2A Biological relevance of CNV calling methods using familial relatedness including monozygotic twins

2A.0 Overview of Chapter 2A

Common methods for CNV detection use microarrays, where the signal intensities of consecutive probes are used to define the number of copies associated with a given genomic region. These practices pose a number of challenges that interfere with the ability of available methods to accurately call CNVs. It has, therefore, become necessary to develop experimental protocols to assess the reliability of CNV calling methods from microarray data so that researchers can properly discriminate biologically relevant data from noise.

I have developed a workflow for the integration of data from multiple CNV calling algorithms using the same array results. It uses four CNV calling programs: PennCNV (PC), Affymetrix® Genotyping Console (AGC), Partek® Genomics Suite (PGS) and Golden Helix™ SVS (GH) to analyze CEL files from the Affymetrix® Human SNP 6.0 Array. To assess the relative suitability of each program, I used individuals of known genetic relationships, including monozygotic twins. I found significant differences in CNV calls obtained by different CNV calling programs.

Although the programs showed variable patterns of CNVs in the same individuals, their distribution in individuals of different degrees of genetic relatedness has allowed me to offer two recommendations. The first involves the use of multiple algorithms (PC, AGC and PGS) for the detection of the largest possible number of CNVs, and the second suggests the use of PennCNV over other methods in this study, when the use of only one software program is desirable.

2A.1 Introduction

Copy number variants (CNVs) are defined as DNA segments (often outlined as 1 kb or larger) that are present in variable numbers in a genome (Feuk *et al*., 2006a; Iafrate *et al*., 2004; Sebat *et al*., 2004). Although common in the human genome, some CNVs have no apparent phenotypic effect (Freeman *et al*., 2006; Iafrate *et al*., 2004; Redon *et al*., 2006), while others are implicated in a variety of phenotypic effects including disease phenotypes (Firth *et al*., 2009; Jacquemont *et al*., 2011; Zhang *et al*., 2011). As such, the search for CNVs associated with disease phenotypes has emerged as a productive approach to identify genetic factors underlying a number of common and complex neurodevelopmental disorders (Conrad and Hurles, 2007; Glessner and Hakonarson, 2009; Kirov, 2010; Maiti *et al*., 2011; Sebat *et al*., 2004). There are two reasons for this productivity. Firstly, CNVs are a major contributor to genomic variation, with approximately 13% of the human genome affected by CNVs (Redon *et al*., 2006), and over 350,000 CNVs have been mapped to specific genomic locations that are documented in the Database of Genomic Variants (DGV) (Macdonald *et al*., 2013). Secondly, advances in technology, including microarrays, permit high-throughput methods to identify CNVs. Such technologies are now relatively common and are economically feasible alternatives to methods like whole genome sequencing. With a number of array platforms and bioinformatic algorithms available, it is necessary to identify optimal analytical pipelines to make inferences regarding specific genomic regions, their copy number identity, and their biological relevance.

On genome-wide microarrays, such as the Affymetrix Genome-Wide Human SNP Array 6.0, sets of probes are designed to determine which allele is present at genomic sites of known single nucleotide polymorphism (SNP probes). The arrays may also include additional probes designed for genomic sites where there is no known variance (known as copy number probes). Normally, CNVs are identified by fluorescent signals generated by SNP probes on the microarray. The fluorescent signals emitted by both SNP probes and copy number probes (if

present on the array) are summarized and analyzed for variance in signal intensity using bioinformatic tools, typically in comparison to a set of reference samples. Consecutive markers that exhibit altered signal intensity from the reference are interpreted as CNVs. There are a number of algorithms that have been developed to identify putative CNVs. Unfortunately, not all putative CNVs called by any existing algorithm can be viewed as biologically relevant. The application of multiple software programs that are designed to call CNVs from the same microarray data often yield differing results (Baross *et al.*, 2007; Kim *et al.*, 2012; Pinto *et al.*, 2007; Winchester *et al.*, 2009; Zhang *et al.*, 2011). The use of multiple algorithms has been shown to increase the reliability of observations with different degrees of confidence. For example, Kim *et al.* used three calling algorithms (PennCNV, QuantiSNP, and Birdsuite) on a set of results from Affymetrix arrays. They found that only 1.5% of total CNV calls could be identified by all three distinct algorithms (Kim *et al.*, 2012). Furthermore, their attempt to confirm putative CNV calls using qPCR produced differing results; 38.3% of CNVs called by a single algorithm, 57.6% of CNVs called by two algorithms and 71.4% of CNVs called by three algorithms could be confirmed by qPCR (Kim *et al.*, 2012).

Although SNP arrays have become popular for ascertaining copy number data in addition to SNP genotypes, there are many issues intrinsic to the use of SNP arrays for the identification of CNVs. Theoretically, it is possible to resolve some of these issues through the use of sensitive analytical methods. In fact, the past ten years have seen a boom in the development of algorithms and technical resolution that have been applied across platforms and programs (Valsesia *et al.*, 2012). Quality control measures such as batch effect correction, normalization methods, and reference group choice seem like simple considerations when compared to the choices available in both algorithms and CNV identification software, as well as in post-analysis filters like marker density and minimum marker thresholds. The reality of limited biological validity in the use of microarrays to call copy number variable regions is concerning for this area of

research that may hold exceptional promise in clinical applications (Duclos *et al.*, 2011). The need for best practices in the workflow for CNV calling protocols has never been more essential.

This study is aimed at assessing putative CNV calls made using the Affymetrix Human SNP 6.0 Array using four CNV calling programs: PennCNV (Wang *et al.*, 2007), Affymetrix Genotyping Console (Korn *et al.*, 2008), Partek Genomics Suite (Partek Inc., St. Louis, MO, USA), and Golden Helix SNP and Variation Suite (Golden Helix, Bonzeman, MT, USA). Using individuals of known relatedness, I have identified overlapping copy number variants across the four algorithms. Using the dataset generated, I have assessed the relative sensitivity of each of the four methods from the following comparisons: between unrelated individuals, between parents and offspring, and between monozygotic twins, that are thought to share 0%, 50%, and 100% genetic relatedness, respectively. The most biologically relevant CNVs will be expected to follow this relationship, with the exception of *de novo* events. The results showed that overall, Affymetrix Genotyping Console identified the most differences between unrelated individuals, while Partek yielded the most similarity between identical twins. On average, PennCNV called CNVs that were comparable to Affymetrix Genotyping Console across unrelated individuals and CNVs that were similar to the Partek results. Assessments using Golden Helix did not follow the trends expected from the known genetic relatedness of individuals. I propose that a combination of three programs (Affymetrix Genotyping Console, Partek, and PennCNV) may be optimal to identify biologically relevant CNV calls due to their ability to resolve copy number variations across different biological relatedness.

2A.2 Methods

2A.2.1 Subjects

This study received ethics approval by the University of Western Ontario’s Committee on Research Involving Human Subjects (Appendix B). Written informed consent was obtained from all participants. Sixteen genomic DNA

samples were isolated from whole blood representing the study participants that included six pairs of MZ twins (three female pairs and three male pairs) and two sets of parents for two of the twin pairs (N = 16). The six twin pairs ranged in age from 20 to 53 years at the time of sample collection. Genomic DNA was extracted from whole blood using the PerfectPure DNA Blood Kit following the manufacturer's protocol (Invitrogen, Carlsbad, CA).

2A.2.2 Microarray Hybridization and Analysis

Whole genome microarray analysis was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 at the London Regional Genomics Centre (London, ON) following the manufacturer's protocol. Sixteen arrays (one array per sample) were processed and analyzed as a single batch and scanned to produce CEL files. The CEL files were used to generate CNV calls on all 16 individuals using four programs: Affymetrix Genotyping Console 4.1.1 (AGC), Partek Genomics Suite (PGS) PennCNV (PC), and Golden Helix SVS Suite 7.0 (GH). In Affymetrix Genotyping Console, I used both the Birdsuite package (version 2) and the Canary algorithm for CNV detection. Birdseye, which is found in Birdsuite, was used for the detection of rare CNVs via a Hidden Markov Model (HMM) and Canary was used to call copy number state in genomic regions with known copy number polymorphism. In Partek, HMM Region Detection using default parameters was selected. In PennCNV, the default HMM algorithm was selected. In Golden Helix I used the Copy Number Analysis Method (CNAM) optimal segmenting algorithm. HMM- based algorithms use prior probabilities of copy number states in conjunction with array-derived normalized fluorescent intensity values to call the most likely copy number state in a given genomic region. The copy number states determined by HMM are discrete and they include 0, 1, 2, 3, and 4+. On the other hand, genomic segmentation scans two adjacent regions of the genome to find differences in copy number using two specific t-tests. CNAM optimal segmenting uses genetic marker map information alongside log2 ratios to discover regions in which log2 ratios vary significantly

between adjacent segments. I used the univariate method of optimal segmenting which segments each sample in the study separately. Canary, which was used in Affymetrix Genotyping Console, calculates Copy Number Polymorphism (CNP) copy number state for over 1,100 regions of known copy number polymorphism (frequency in the population greater than one percent).

The user-defined analytical parameters were kept consistent across the analysis. Specifically, the HapMap 270 6.0 Array reference was used as a reference file and variants were identified as DNA regions, which were called as copy number state of 0, 1, 3, or 4+ covering a minimum of 10 consecutive markers on the array. Only variants greater than or equal to 1 kb in size were included in subsequent analysis.

2A.2.3 CNV Call Merging

The CNV calls made by each of the four software programs were merged with adjacent CNV calls that may represent the same CNV event. The criteria used to merge were 1) CNVs had to be adjacent on the same chromosome (no other CNV call between them); 2) CNVs had to share the same gain/loss status; and 3) adjacent calls were ≤ 20% of the total length, that is, if there were three consecutive genomic segments A, B and C, where A and C are both losses and B is unchanged, I divided the length of the gap B by the length of A + B + C. If this fraction was ≤20%, then I merged A + B + C as a single CNV call. If there were multiple consecutive CNVs, each with 20% or less length between one and all of the others, then I extended the formula to the next CNV and merged all of the CNVs into one event. When multiple smaller CNVs were merged into one large CNV event, I identified the event as a merged CNV. I then labeled the newly merged CNVs and any CNVs that remained unmerged as either “CNV-Gain” or “CNV-Loss” within the calls from all four software programs in all individuals.

2A.2.4 Comparison of CNV Calls Between Software Programs

To compare CNV calls made by different software programs, I used a 50% reciprocal overlap (RO) criterion to compare the calls made within an individual from the four software programs. The use of 50% RO for comparing calls is consistent with other reports (Pang *et al*., 2010; Wain *et al*., 2009; Yavas *et al*., 2009). Two CNV events were considered to pass the 50% RO criterion if at least half of the length of the first CNV overlapped with the second CNV and vice versa. If the 50% RO criterion was met, the two events were then considered to be the same event (called by different algorithms but identified in the same individual) as long as their call states also matched. I calculated the reciprocal overlap (O) (≥50% criteria) as follows:

Where x and y are both CNVs, L is length in base pairs that the two CNVs (x and y) overlap, end indicates the end base pair position of the given CNV, and start indicates the start base pair position of the given CNV.

$$O(A) = \frac{L}{x_{end} - x_{start} + 1}$$

$$O(B) = \frac{L}{y_{end} - y_{start} + 1}$$

CNVs met the ≥50% RO criteria if O(A) and O(B) were both ≥50%. CNVs that did not meet this criterion were considered to be different events.

2A.2.5 Comparison of CNV Calls Between Subjects

The same RO definition (≥50%) was used to compare shared and unshared calls in a pairwise comparison between individuals in the following categories of genetic relatedness: between MZ twins in a twin pair, between parent and child, and between unrelated individuals. Following RO comparisons, I calculated the average difference (*d*) within each group (where *d* is the total number of unshared CNVs across the two compared individuals, divided by the total number of CNVs called across the two compared individuals). Specifically, I

looked at three comparisons within each group, that is, three MZ twin comparisons, three parent–child comparisons and three unrelated pair comparisons. This calculation was used to test the relationship within each group in relation to their expected genetic relatedness. To perform the reciprocal overlap formula, HD-CNV (Hotspot Detector for Copy Number Variants) was used with 50 as the identified RO merge criteria (Butler *et al.*, 2013). The *d* value was averaged for each type of relatedness in each individual software program. The results were assessed to compare the effectiveness of each individual algorithm in the identification of biologically relevant CNVs.

2A.3 Results

Table 2.1 shows the total number of unmerged CNVs representing gains and losses identified by each of the four software programs for 16 individuals using the same CEL files from Affymetrix Human SNP 6.0 Arrays. The 16 individuals included in this analysis represent six pairs of MZ twins and the two parents for twin pairs 2 and 3. The results show that the number of raw CNVs identified in each individual varies depending on the program used. This variability is apparent in the numbers of gains and losses as well as the total numbers. Although AGC, PGS, and PC identified similar numbers of CNVs for most individuals (average of ~78 CNVs per individual), GH yielded more CNVs in each individual (average of ~317 CNVs per individual). While PGS yielded relatively more gains than losses, the other programs (AGC, PC and GH) yielded relatively more losses than gains. The differences in the number of gains and losses called between programs have suggested that each method may highlight some aspects of CNV calling but not others. I attempted to gain an insight into this variability by assessing the distribution of CNV calls in different contexts.

Table 2.1. Raw (pre-merge) Copy Number Variant Calls by Program.

ID	Affymetrix Genotyping Console			PennCNV			Partek Genomics Suite			Golden Helix SVS		
	Gain	Loss	Total	Gain	Loss	Total	Gain	Loss	Total	Gain	Loss	Total
1A	12	29	41	12	24	36	22	31	53	47	121	168
1B	19	31	50	20	21	41	44	28	72	107	102	209
2A	19	29	48	17	24	41	49	17	66	79	121	200
2B	17	29	46	17	25	42	41	24	65	77	114	191
3A	19	28	47	9	16	25	34	20	54	72	120	192
3B	18	26	44	6	15	21	29	14	43	73	127	200
4A	13	31	44	7	18	25	36	13	49	78	122	200
4B	17	35	52	6	16	22	25	12	37	63	107	170
5A	15	41	56	8	37	45	15	39	54	82	122	204
5B	12	42	54	7	34	41	13	45	58	73	99	172
6A	30	31	61	25	25	50	20	27	47	82	104	186
6B	26	27	53	26	27	53	21	27	48	73	108	181
Father 2A/2B	1065	101	1166	49	42	91	105	32	137	1009	297	1306
Mother 2A/2B	175	34	209	13	35	48	19	96	115	496	301	797
Father 3A/3B	24	41	65	14	13	27	16	11	27	113	263	376
Mother3A/3B	21	48	69	10	25	35	9	89	98	62	262	324

Note: A=Affected Twin, B=Unaffected Twin

First, I assessed the size distribution of CNVs called by the four programs. I found that the four programs vary in the number of CNVs called and that CNVs fall into different size categories (Figure 2.1). CNVs in the range of 1-100 kb were most frequent in AGC calls (>80% of total calls) and least frequent in PGS calls (<60% of total calls). Similarly, the largest CNVs (1-10 Mb) were observed at higher frequency in PGS calls (>10%) as compared to the other three programs (range 1-5%).

The chromosomal distribution of CNVs identified by the four programs showed that GH calls more CNVs on all chromosomes as would be expected with the higher number of calls overall (Figure 2.2). Also, this number is closely followed by PGS calls particularly on chromosome 2, 9, 14 and 15. Otherwise, the distribution of CNVs across chromosomes is proportional to chromosome size, as expected.

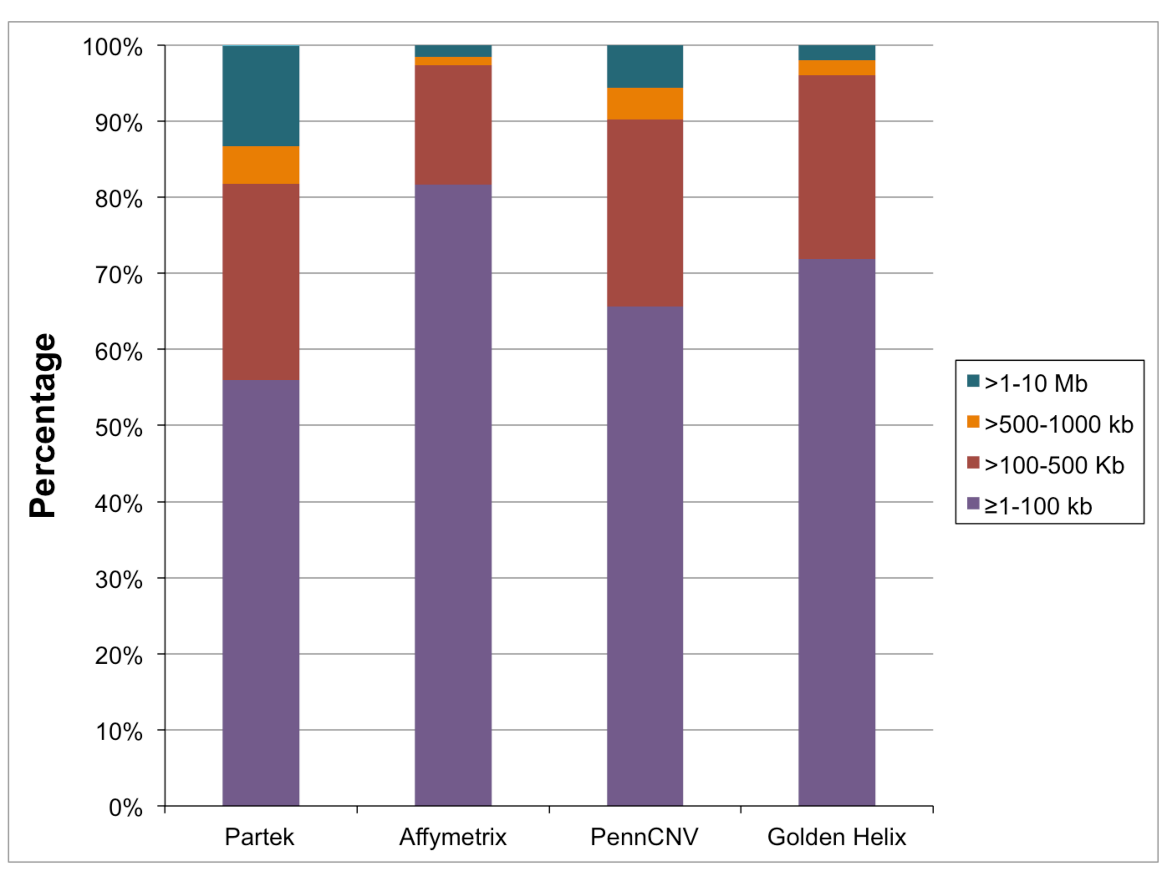


Figure 2.1. Distribution of raw copy number variant calls (3957) by size generated by each software package.

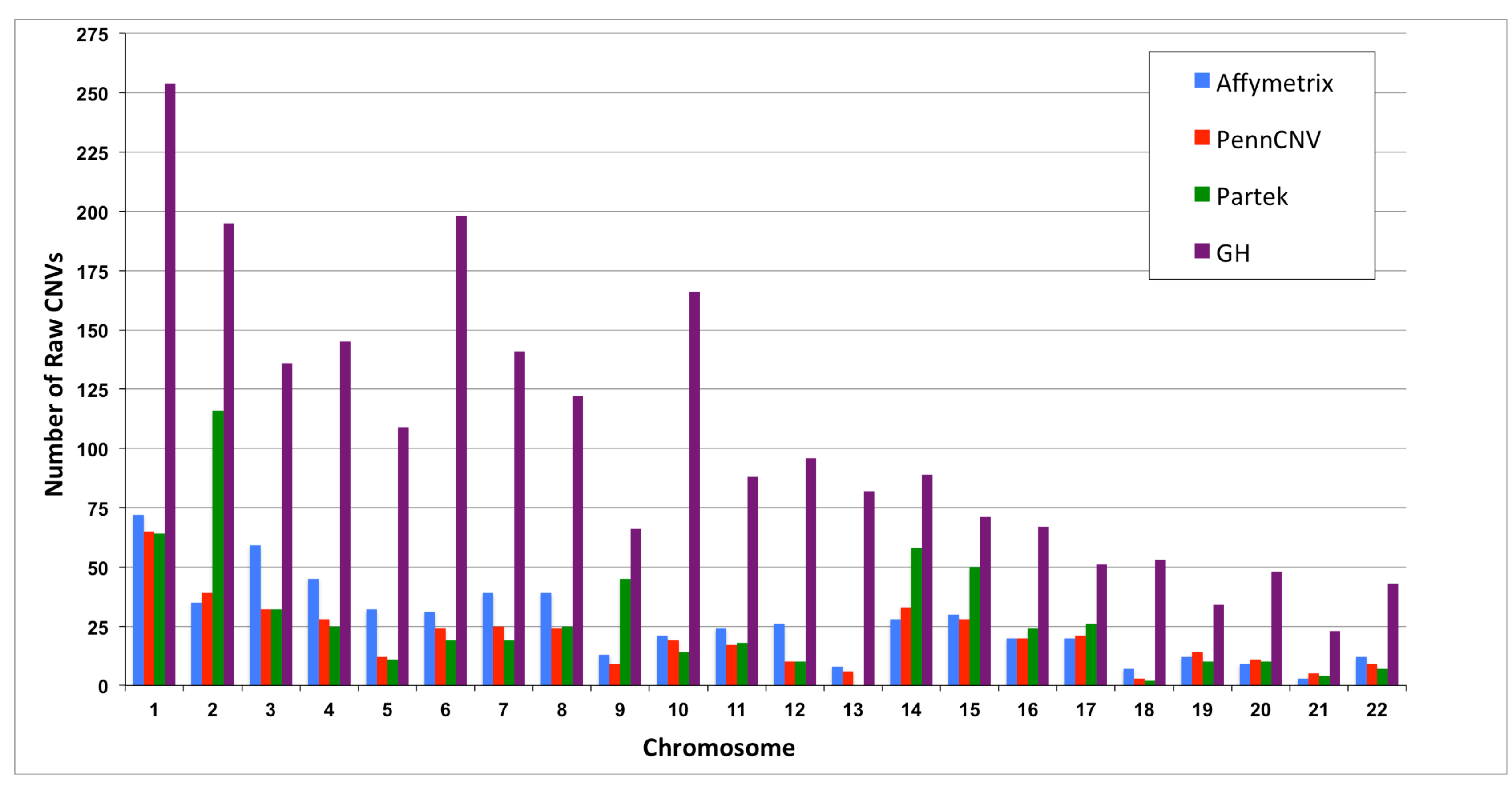


Figure 2.2. Chromosomal distribution of raw copy number variants in autosomes (3957) generated by each software package.

Next, I assessed the overlaps in CNV calls made by each of the four programs in the twelve MZ twins (Figure 2.3). Most CNVs called by GH were not shared by calls made by any other program. The low degree of overlap suggests that the underlying assumptions of CNV calling by GH are different from the other three methods. Also, a significant number of CNV calls called by the other three methods (AGC, PGS and PC) showed overlaps. The CNV calls by PGS and AGC showed the most overlap (59%), closely followed by the overlap between PC and PGS (54%), and between AGC and PC (46%). CNV calls that overlapped between AGC, PGS, and PC represented 27% of the total number of CNV calls made by the three programs. When calls made by GH were included, all four programs shared only 0.32% (12/3713) of total CNV calls made. These results are similar to other reports involving comparison of different CNV calling programs (Kim *et al*., 2012; Pinto *et al*., 2007).

I expected that differences and similarities in CNV calls would follow the genetic relatedness of individuals. For example, monozygotic twins would be expected to share the most CNV calls, while CNV calls for two unrelated individuals will show the highest amount of divergence. Similarly, pairs of individuals (parent and child) with presumed genetic relatedness would be expected to fall between 100% (MZ twins) and 0% (unrelated individuals).

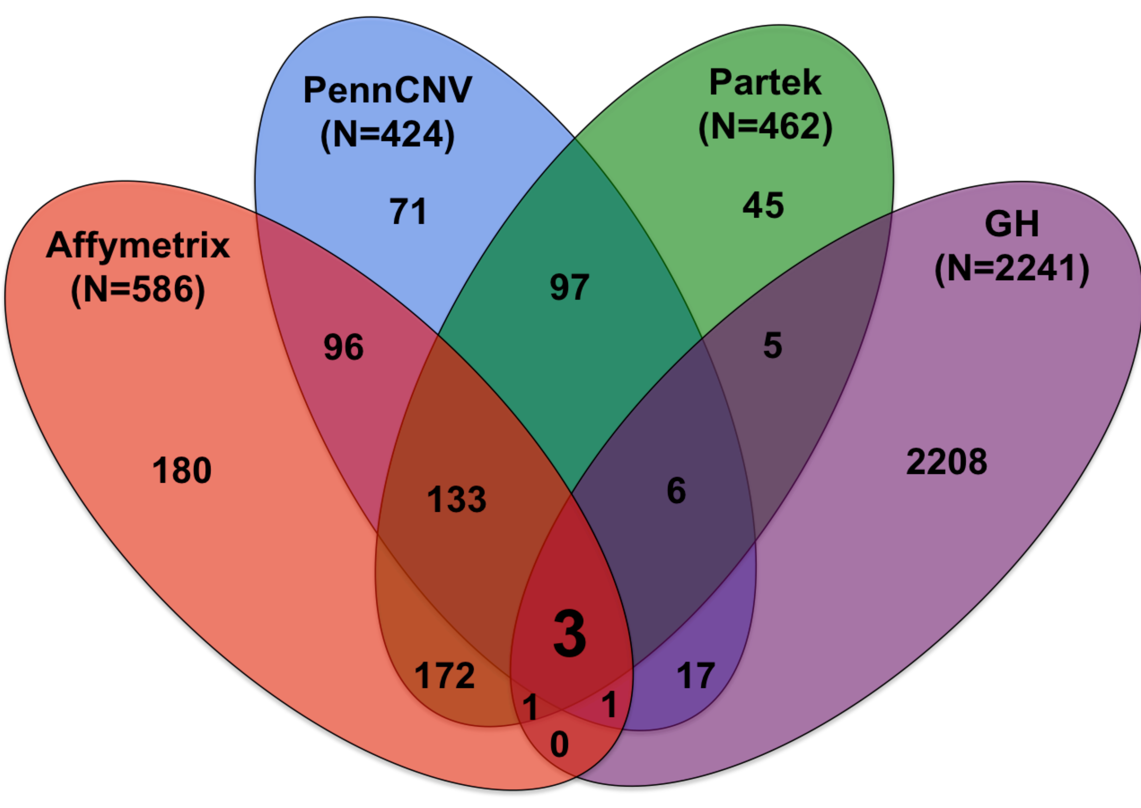


Figure 2.3. Overlap of copy number calls across programs.

Venn diagram showing CNV (post merging of adjacent calls into one event) calls made by each software program across six pairs of monozygotic twins (numbers in parenthesis), as well as the number of CNVs shared in common between three software programs and between four software programs (3 CNVs overlapping *KGFLP1, CACNB2, ST8SIA6*). The total number of non-unique CNVs identified post-merge across four software programs was 3713.

Figure 2.4 shows the degree of CNV difference (*d*) involving randomly selected pairs of unrelated individuals (N = 3), parent–child pairs (N = 3), and MZ twins (N = 3) for each of the four software programs. The estimate of difference (*d*) for CNVs called by GH has no relationship to the genetic relatedness between individuals. Conversely, CNV calls made by the other three programs (AGC, PGS, and PC) follow the expected genetic relatedness. The largest difference between unrelated individuals was identified by AGC (82%) and followed closely by PC (80%). The smallest difference found between MZ twins is reflected by PGS (18%) followed by PC (21%). Interestingly, the parent–child differences for the three methods PGS, PC, and AGC were estimated to be ~56%, ~61%, and ~72%, respectively. Even though the standard error associated with these means (based on only 3 comparisons) are relatively large, the overall trends for the pairwise comparisons made by PGS, PC, and AGC follow the expected pattern based on the known genetic relatedness.

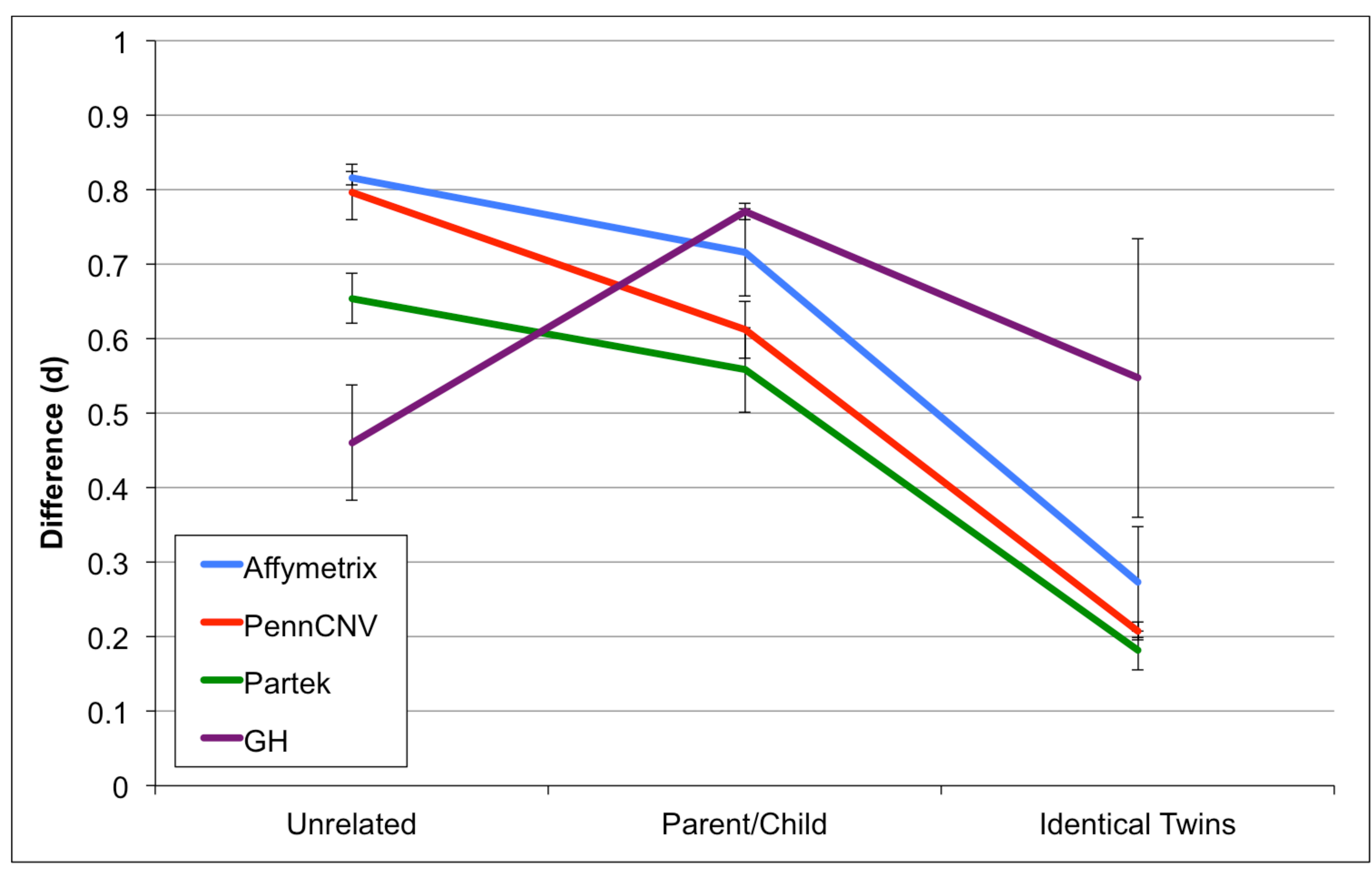


Figure 2.4. Copy number call differences made by each program and categorized by relatedness.

Mean difference (*d*) ± SEM, of three pairwise comparisons between two unrelated individuals, parent and child, and monozygotic twins as determined by each software program (PennCNV, Affymetrix: Affymetrix Genotyping Console, Partek: Partek Genomics Suite, GH: Golden Helix SNP and Variation Suite).

2A.4 Discussion

The results support the likelihood that PGS, PC, and AGC are identifying biologically relevant CNV calls. Calls made by multiple programs provide a greater likelihood that the underlying CNV may be real. At the same time, if one is forced to choose only one method, my analysis based on the ability to resolve varying degrees of genetic relatedness favours the use of PC. The reason for this choice is based on the fact that it has a relatively high *d* value for unrelated individuals, a low *d* value between MZ twins, and an intermediate degree of difference involving parent and child, as would be expected. On average, the number of CNVs found in parental samples was higher than the number found in twin samples; this finding may reflect the age of sample collection. Environmental effects may also be contributing to the increased number, particularly in Twin Pair 2 where the Father has undergone chemotherapy treatment.

I compared four programs to call CNVs (AGC, PGS, PC, and GH) from the same microarray data for 16 individuals using the Affymetrix Human SNP 6.0 Array. The CNV calls are different across the four methods, but some overlap was observed. It follows a number of other reports in the literature that have also reported similar discrepancies (Eckel-Passow *et al.*, 2011). It is therefore not surprising that the results are different across the four methods and particularly between the three methods that use a HMM (AGC, PGS, and PC) based approach versus the method that uses a segmentation approach (GH). Unlike the HMM that assume the means of different copy number states to be consistent, optimal segmenting delineates CNV boundaries with increased sensitivity.

Overall, I conclude that some CNV calling methods can appropriately distinguish known levels of genetic relatedness and others have more difficulty doing so. I also note that some differences between related individuals, including monozygotic twins would be expected due to somatic mosaicism; however, these differences would be expected to be relatively small in number.

The results presented in this report suggest that microarray experiments

are prone to errors in CNV calls. Further, my results are likely to include false positive as well as false negative calls found in the arrays. Most CNV databases are populated with entries from the past half-decade that include results that have not been confirmed and may represent error (Eckel-Passow *et al*., 2011; Lin *et al*., 2011; Pawitan *et al*., 2005; Zhang *et al*., 2011). In fact, research involving CNV calling from microarray results would benefit from better microarray technologies, better algorithms for CNV calling and better standards for independent confirmation.

All steps in a microarray experiment, from the isolation of DNA from tissue samples to the calling of CNVs from CEL files, are points at which error can be introduced. These and other confounding factors could affect the accuracy of biologically-significant CNV detection. Inferring copy number from array data is notoriously plagued with false-positive rates that may vary depending on the algorithm used (Eckel-Passow *et al*., 2011). Given the scientific and clinical implications for accurate CNV detection (Valsesia *et al*., 2012) as well as the introduction of ascertainment bias into future studies via microarray design, algorithm parameters, and database entries (Teo *et al*., 2012; Wineinger and Tiwari, 2012), it is necessary to identify the most reliable CNV calling programs.

2A.5 Conclusion

The results presented here offer a number of insights with respect to the total number of CNVs across individuals including gains and losses, program-specific distributions of CNV size, and CNV distribution across chromosomes. As expected, the number of CNVs was related to the size of the chromosome. The major observation from the collective results is that the number of CNVs identified by Golden Helix (GH) software was much higher than the number called by the other three programs (AGC, PC, PGC). Also, Golden Helix identified vastly different CNVs when compared to those identified by Affymetrix Genotyping Console (AGC), Partek Genomic Suite (PGS), and PennCNV (PC). Excluding GH as an outlier, AGC yielded more CNV calls than PGS and PC. As

expected, AGC, PGS, and PC yielded a relatively large number of overlapping CNVs. These results are consistent with previous reports (Pang *et al.*, 2010).

Further assessment of the four CNV calling methods was considered in the context of similarities and differences involving individuals of differing genetic relatedness. As it stands, three of the four methods met such expectations to different degrees. The results offer two conclusions. First, overlapping CNV calls by three of the programs (AGC, PGS, and PC) will offer the highest likelihood of discovering biologically relevant calls. The combination of AGC and PC identified the most differences among unrelated individuals whereas PGS and PC showed the least differences between MZ twins. The results from GH showed a higher number of CNVs than would be expected and also did not follow the expected pattern when groups of known relatedness were compared. For this reason, I do not recommend GH and suggest that further research should explore the unexpected profile generated from the software. Secondly, the PC calls best reflect the expectations at all three levels involving unrelated individuals, parent–child, and MZ twins. My results and conclusions are in agreement with other groups, which have found that without independent validation using bench confirmation techniques such as qPCR, CNVs calls should not be assumed to be truly valid variants (Kim *et al.*, 2012; Pinto *et al.*, 2011). Finally, I suggest that incorporation of family data will help in improving the quality of CNV calls alongside the use of multiple CNV calling methods.

A version of this chapter has been published in Castellani *et al.*, *Twin Res. Hum. Genet.* (2014b)

2B Assessment of Copy Number Variation in Monozygotic Twins Discordant for Schizophrenia

2B.0 Overview of 2B

As mentioned above, when the Affymetrix Human SNP 6.0 arrays were analyzed using Affymetrix Genotyping Console™, Partek Genomics Suite, PennCNV, and Golden Helix SVS the results yielded both program-specific and overlapping variants. In the further assessment of CNVs that may contribute to disease discordance, only CNVs called by Affymetrix Genotyping Console, Partek Genomics Suite, and PennCNV were used in further analysis. This analysis included an assessment of calls in each of the six twin pairs towards identification of unique CNVs in affected and unaffected co-twins. Real time polymerase chain reaction (PCR) experiments confirmed two CNVs that were found in the affected patient but not in the unaffected twin. The results identified CNVs and genes that were previously implicated in mental abnormalities in four of the six twin pairs. Specifically the genes PYY (twin pairs 1 and 5), EPHA3 (twin pair 3), KIAA1211L (twin pair 4), and GPR139 (twin pair 5) were identified as patient specific. They represent potential candidate genes and CNV regions that may contribute to the discordance of these monozygotic twin pairs for this heterogeneous neurodevelopmental disorder. These findings suggest that phenotypic differences between monozygotic twins may arise from ontogenetic *de novo* events arising after the splitting of the zygote.

2B.1 Introduction

Schizophrenia clusters in some families and has a high heritability estimate (80%) (Sullivan *et al.*, 2003). In fact, the best predictor of the occurrence of this disease is family history. The inheritance pattern of schizophrenia is complex. This complexity is reflected in the observation that monozygotic twins, who are said to share 100%, and dizygotic twins who are said to share 50%, of

their genetic makeup are concordant in only 48% and 17% of cases, respectively (McGuffin *et al*., 1994). These observations suggest a role for non-genetic and random genetic factors (O'Reilly and Singh, 1996; Singh and O'Reilly, 2009), including random developmental events (Singh *et al*., 2004), epigenetic mechanisms (Singh *et al*., 2002), and environmental factors (Torrey *et al*., 1997). Over 30 years of genetic research using linkage and association analysis have identified a number of promising linkages (Sullivan, 2005) and candidate genes (Hamilton, 2008; Karayiorgou and Gogos, 2006). Most of these results have been difficult to reliably replicate except in the case of a few variants, which have been associated across multiple studies, typically when large sample sizes are employed (Ripke *et al*., 2013; Torkamani *et al*., 2010). This difficulty in identifying causal genes for schizophrenia has been attributed to extensive heterogeneity, including genetic differences among patients from the same family (Beckmann and Franzek, 2000). Application of genome-wide expression arrays in schizophrenia has identified a long list of genes with altered expression in the brain (McInnes and Lauriat, 2006; Verveer *et al*., 2007) and blood (Gladkevich *et al*., 2004; Tsuang *et al*., 2005). However, altered expression of these genes cannot always be replicated.

Recent advances have helped in the identification of CNVs and opened a new direction in schizophrenia genetics research (Kirov, 2010). By virtue of their variable size, they may directly disrupt multiple genes that are co-located (Feuk *et al*., 2006b). In addition to having a direct effect on the expression of the amplified or deleted genes (Stranger *et al*., 2007), they may have indirect effects on gene expression extending upstream and downstream of the CNV region (Henrichsen *et al*., 2009). While most CNVs are polymorphic, some are generated *de novo* (Zogopoulos *et al*., 2007). Common CNVs in humans are believed to play a role in evolution (Lee and Scherer, 2010). They also underlie a significant proportion of variation in humans, including differences in cognitive, behavioral, and psychological features (Lee and Lupski, 2006). Further, they have been implicated across a wide variety of common disorders (Buchanan and

Scherer, 2008; Consortium *et al*., 2010; Stankiewicz and Lupski, 2010), including mental disorders (Feuk *et al*., 2006a; Lee and Lupski, 2006; McCarroll and Altshuler, 2007), particularly autism (OMIM: 209850) (Consortium *et al*., 2007; Glessner *et al*., 2009; Moessner *et al*., 2007; Sebat *et al*., 2007; Wang *et al*., 2009) and schizophrenia (Glessner *et al*., 2010; Kirov *et al*., 2008; Need *et al*., 2009; Stefansson *et al*., 2008; Walsh *et al*., 2008; Xu *et al*., 2008). The results, generated with increasing genomic coverage and numbers of patients, have identified a set of candidate CNVs. These include rare deletions at 1q21.1, 15q13.3, 15q11.2, and 22q11.2, as well as duplications at 16p11.2, 16p13.1, and 7q36.3 (Kirov *et al*., 2008). In addition, various gene regions have been associated with copy number variation in schizophrenia, namely deletions of NRXN1, APBA2, and CNTNAP2 (Friedman *et al*., 2008; Liu *et al*., 2002).

The findings in the field also suggest that, with few exceptions, schizophrenia is caused by aberrations in a relatively large number of genes, most with relatively small effects, that cumulatively produce a genetic predisposition. Some of these aberrations may be inherited while others may represent *de novo* events (Singh *et al*., 2009). The field is starting to recognize that rare variants likely play a role in the causation of schizophrenia. This model is not compatible with traditional experiments in which a group of patients are compared with an equally large group of unaffected controls. In such an approach, adding more patients will add additional genetic heterogeneity across cases. This complexity is likely better approached by the precise genetic matching of patients with unaffected controls that can be achieved using monozygotic twins. Even if rare variants identified using this approach are limited to a given set of twins or a given family, they are likely to help in identifying the underlying pathways and genes involved in this disorder.

In this research, I used six pairs of monozygotic twins discordant (MZD) for schizophrenia and assessed the CNV differences between twin pairs. The resulting CNV differences are of interest in identifying patient-specific differences, including gene dosage changes that may differ in a MZD pair. Previous studies

utilizing monozygotic twins have associated CNV differences between twins with various diseases. Using MZ twins, three somatic CNV events were found to be associated with discordance for congenital heart defects (Breckpot *et al*., 2012). Similarly, two *de novo* CNVs — a pre-twinning duplication and a post-twinning deletion were found to be associated with attention problems (Ehli *et al*., 2012). The results are twin-specific and trends are not always consistent (Bloom *et al*., 2013; Halder *et al*., 2012; Maiti *et al*., 2011). Some studies call CNV differences (Maiti *et al*., 2011) while others call no difference in CNVs between MZ twins discordant for schizophrenia (Bloom *et al*., 2013). In either case, the MZD strategy is effective in the identification of previously undiscovered genes in schizophrenia, particularly when combined with the use of multiple software programs for CNV analysis. Given the high heterogeneity of this disorder, I would *a priori* expect many aberrations to be patient-specific. These patient-specific genetic changes can be best identified using nature's best match for each patient - their monozygotic twin. I have hypothesized that the discordance of monozygotic twins for schizophrenia may involve *de novo* mutations (DNM) (Singh *et al*., 2009) contributing to a disease liability threshold. If that is so, I should be able to identify differences between MZD twins for schizophrenia that are *de novo* in nature and do not apply to all twin pairs, but instead show twin-pair specificity. In this report, I have employed a stringent CNV detection protocol using multiple CNV calling methods, and identified CNV differences between MZD for schizophrenia. The results support the potential presence of *de novo* CNVs that are compatible with the development of schizophrenia.

2B.2 Methods

2B.2.1 Ethics Statement and Clinical Background

As mentioned in the methods for section 2A, this study received ethics approval by the University of Western Ontario's Committee on research involving human subjects (Appendix B). All subjects provided written informed consent to participate in this study and were interviewed by a psychiatrist (ROR) using the SCID-I (First *et al*., 1996) and the SCID-II (First *et al*., 1997). All of the patients were adults at the time of consent. Past clinical notes were obtained to aid diagnosis. Whole blood samples were obtained from each individual. The twin pairs studied ranged in the age from 20 to 53 years at the time of sample collection. Three of the pairs were female and three of the pairs were male.

The overall strategy used to generate and interpret the results is outlined in a flowchart in Figure 2.5.

2B.2.2 DNA Preparation, Hybridization and CEL File Analysis

Deoxyribonucleic acid was extracted from whole blood using the PerfectPure DNA Blood Kit (http:// www.5prime.com) following the manufacturer's protocol. Whole genome microarray analysis using the Affymetrix Genome-Wide Human SNP Array 6.0 was performed at the London Regional Genomics Centre (LRGC) following the manufacturer's protocol. For downstream analysis of CEL files, Affymetrix Genotyping Console 4.1.1 (A), Partek Genomics Suite (P), PennCNV (p), and Golden Helix SVS Suite 7.0 (G) were used.

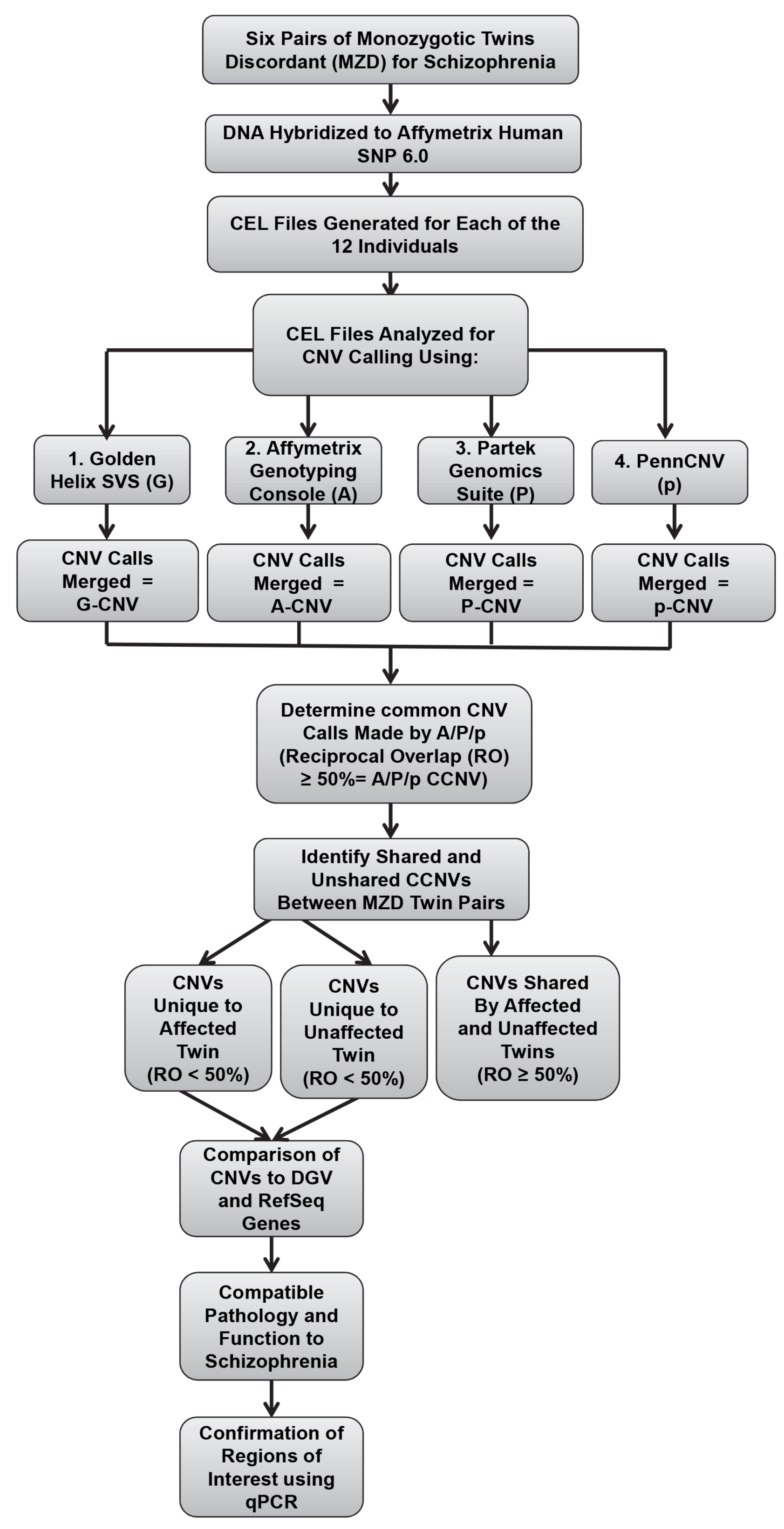


Figure 2.5. Flowchart of Chapter 2B experimental design

2B.2.3 Calling and Merging of CNVs for Individual Genomes

The Hap Map 270 6.0 Array reference was utilized as a model reference file. Variants were identified as those DNA regions, which were called as copy number state 0, 1, 3, or 4+ by 10 or more consecutive markers on the chip. Recent literature suggests that a baseline of at least seven consecutive probes is necessary for reliable CNV detection (Wineinger and Tiwari, 2012). In addition, only variants that were greater than 1 kb in size were classified as CNVs for the purposes of this study and only those identified by at least three software programs in the same individual were included in subsequent analysis. I used quantile normalization across all four software programs. Overlapping genes were identified using the UCSC (University of California, Santa Cruz) genome browser table view (NCBI36/hg18). Identification of CNVs was followed by merging of CNV calls within software programs and comparison of calls between software programs to identify CNVs called by the three software programs. I identified CNVs within each of the four software programs that were likely to be the same event using the following criteria: (1) CNVs had to be adjacent on the same chromosome (no other CNV call between them); (2) had to share the same gain/loss status; (3) adjacent calls were merged together into one single call, using gap ≤20% of total length. That is, if there are three genomic segments, A, B, and C, where A and C are both losses, I divided the length of gap B by the length of A+B+C, and if this fraction is ≤20%, then I merged A + B + C as a single CNV call.

2B.2.4 Identification of Overlapping CNV Calls for Individual Genomes

I used the 50% reciprocal overlap (RO) formula as discussed in Chapter 2A. Copy number variants that shared 50% or more similarity with one another were classified as overlapping. This is consistent with the definition of an overlapping CNV identified by other groups (Pang *et al*., 2010; Wain *et al*., 2009; Yavas *et al*., 2009). In other words, if at least half of the first CNV overlapped with the second CNV and vice versa they were considered to be the same event.

2B.2.5 Comparison of CNV Calls Within Monozygotic Twin Pairs

The same RO definition was used to compare calls between monozygotic twin pairs. CNVs were compared between affected and unaffected twin pairs to determine which CNVs were shared and unshared between twins. Unshared CNVs between twins were then annotated with gene information and compared to CNVs in the Database of Genomic Variants (DGV). The genes overlapping CNVs that were different between twin pairs and called by at least three software programs were further characterized using Ingenuity Pathway Analysis (IPA; Ingenuity Systems, California) and GeneMania (Toronto, ON) to identify gene networks and canonical pathways. Finally, I compared the genes identified in this study to those genes listed in the Schizophrenia Gene Database (http://www.schizophreniaforum.org/res/sczgene/default.asp) to determine the genes that appear to be most likely to play a role in schizophrenia discordance. Additional searches from PubMed covering the most recent results helped update any connection between genes of interest and disease pathology.

2B.2.6 Confirmation of Unique CNVs

A subset (10) of the CNV differences identified between monozygotic twins were confirmed using TaqMan Quantitative PCR (qPCR) Copy Number Assays from Life Technologies, Inc (Grand Island, NY). In particular, CNVs which were unique to co-twin and overlapped regions of interest to disease were chosen. The control used for comparison of copy number in the TaqMan experiments was RNAseP and the calibrator used was the individual's unaffected co-twin. Each Real Time experiment was repeated four times. The delta-delta Ct method was used and significance was defined as a Z-Score <1.75 and 95% confidence level. The Ct value is the Cycle Threshold, and is defined as the number of cycles required for the fluorescent signal to cross the threshold value. CNV calls were determined using Copy Caller (Life Technologies).

2B.3 Results

First, the number of unfiltered CNVs identified by each software program in each of the six twin pairs varied by program (Table 2.1). Golden Helix's SVS identified the highest number of CNVs in each individual, with a range of 168 to 209 variants. Affymetrix Genotyping Console and Partek Genomics Suite called similar numbers of CNVs in each individual, with ranges of 41 to 61 and 37 to 72 variants, respectively. PennCNV called the smallest number of CNVs in each individual, with a range of 21 to 53 variants. The smallest number of unfiltered variants called in any one individual was 21 (PennCNV) and the largest number was 209 (Golden Helix SVS). The majority of CNV calls across all four software programs were between 1 kb and 100 kb in size, with at least 56% of the total CNVs in each analysis falling into this range. Partek Genomics Suite yielded the highest number of large calls (1-10 Mb) and Affymetrix Genotyping Console yielded the greatest number of small CNV calls (1–100 kb).

Second, the percentage of Affymetrix Genotyping Console CNV calls detected by three or more algorithms was 23.04%, the percentage of Partek CNV calls detected by three or more algorithms was 30.30%, the percentage of PennCNV calls detected by three or more algorithms was 33.02% and the percentage of SVS CNV calls detected by three or more algorithms was only 0.36%. This summary of overlapping CNVs across four programs strongly suggests that the most reliable calls may represent overlapping CNVs involving Affymetrix (A), Partek (P), and PennCNV (p), termed as A/P/p. A combination that includes Golden Helix along with any other two methods yielded rare CNVs only and was considered too restrictive. More importantly, the overlap generated by A/P/p calls was less restrictive across the twin pairs. Consequently, the CNVs identified by this combination were further assessed in the follow-up analysis involving shared and unshared CNVs between members of the six twin pairs studied (Table 2.2).

Third, I found a total of 38 CNV events called by the three (A/P/p) software programs that were not shared with their co-twin across the six pairs.

Specifically, 14 unique CNV events were observed in co-twins affected with schizophrenia (Table 2.3), while 22 were unique to the six unaffected co-twins of the six MZD pairs (Table 2.4). Some of the CNVs in both categories contained genes, while others were located in non-coding regions of the genome. In fact, there was a total of 12 unique genes overlapping the 14 CNVs that were found in affected members only. Similarly, there was a total of 28 unique genes overlapping the 22 CNVs that were found in unaffected twins only. The results confirm that monozygotic twins do differ for rare CNVs. I have undertaken a pair-specific analysis in an effort to explain the discordance of the monozygotic twin pairs for schizophrenia which is described below.

Table 2.2. Shared and unshared status of CNVs called by 3+ software programs in identical twins.

Twin	Called by Affymetrix GC, Partek and Golden Helix SVS			Called by Partek, Golden Helix SVS and PennCNV			Called by Affymetrix GC, Golden Helix SVS and PennCNV			Called by Affymetrix GC, Partek and PennCNV (A/P/p)		
	Shared	Unique	Total	Shared	Unique	Total	Shared	Unique	Total	Shared	Unique	Total
1A	0	0	0	0	0	0	0	0	0	10	4	14
1B		3	3		4	4		3	3		2	12
2A	0	0	0	1	2	3	0	0	0	11	1	12
2B		0	0		0	1		0	0		4	15
3A	0	0	0	0	1	1	0	0	0	6	4	10
3B		0	0		0	0		0	0		1	7
4A	0	0	0	0	0	0	0	0	0	7	3	10
4B		0	0		0	0		0	0		0	7
5A	0	0	0	0	0	0	0	0	0	8	2	10
5B		1	1		0	0		1	1		10	18
6A	0	0	0	0	0	0	0	0	0	8	0	8
6B		0	0		0	0		0	0		5	13

Table 2.3. CNVs identified by Affymetrix Genotyping Console, Partek Genomics Suite and PennCNV (A/P/p) and only present in the affected co-twin. (NCBI36/hg18)

ID	Chr	Region	Start Position	End Position	Length (kb)	State	# of Probes	Genes	DGV
1A	6	6q14.1	77016905	77027694	11	Loss	33	None	No
	7	7q31.1	109441200	109453946	13	Loss	13	None	No
	14	14q21.2	41628351	41657337	29	Loss	11	None	No
	17	17q21.31	39423041	39430053	7	Loss	17	*PYY*	No
2A	11	11p15.1	18949072	18961778	13	Gain	55	None	Yes
3A	3	3p11.2-3p11.1	89394600	89419369	25	Loss	51	*EPHA3*	No
	7	7q11.21	64594329	64955220	361	Loss	174	*LOC441242, INTS4L2, CCT6P1, SNORA22*	Yes
	11	11p15.4	5789589	5809449	20	Loss	30	*OR52N2*	Yes
	16	16p11.1	34459037	34624994	166	Gain	128	*LOC283914, LOC146481, LOC100130700*	Yes
4A	2	2q11.2	98858308	98879625	21	Loss	52	*KIAA1211L*	No
	3	3q25.1	151511085	151547185	36	Loss	30	None	No
	16	16q23.2	78372440	78377393	5	Loss	11	None	No
5A	16	16p12.3	19945650	19965863	20	Loss	30	*GPR139*	No
	17	17q21.31	39423041	39430053	7	Loss	17	*PYY*	No
6A	NONE								

Note: Chr= Chromosome, DGV= Database of Genomic Variants.

Table 2.4. CNVs identified by Affymetrix Genotyping Console, Partek Genomics Suite and PennCNV (A/P/p) and only present in the unaffected co-twin. (NCBI36/hg18)

ID	Chr	Region	Start Position	End Position	Length (kb)	State	# of Probes	Genes	DGV
1B	2	2q22.1	138045294	138065759	20	Loss	24	*THSD7B*	No
	7	7q35	143917601	144051219	134	Gain	31	*TPK1*	Yes
2B	1	1q21.2	149086185	149202838	117	Loss	76	*ARNT,SETDB1*	Yes
	7	7q21.2	91031620	91040715	9	Loss	29	None	Yes
	17	17q21.32	44535027	44739523	204	Gain	30	*B4GALNT2,GNGT2, ABI3,PHOSPHO1, FLJ40194,ZNF652*	No
	19	19q13.13-19q13.2	43294390	43538078	244	Loss	117	*SIPA1L3,DPF1,PPP1R14, SPINT2,YIF1B,C19orf33, KCNK6, CATSPERG*	Yes
3B	3	3q26.1	165046890	165083684	37	Loss	65	None	No
4B	NONE								
5B	1	1p36.33	61735	86161	24	Loss	14	None	Yes
	1	1p21.1	104155790	104248433	93	Loss	31	None	Yes
	1	1q25.2	174796556	174801847	5	Loss	17	*PAPPA2*	No
	2	2p22.3	34709689	34727855	18	Loss	21	None	Yes
	3	3p14.1	68746577	68747401	1	Loss	11	None	No
	3	3q11.2	98944458	98949409	5	Loss	38	*EPHA6*	No

	4	4p15.1	34779042	34822761	44	Loss	40	None	No
	4	4q24	104742390	104761153	19	Loss	37	*TACR3*	No
	8	8p11.23	39301780	39384688	83	Loss	28	*ADAM5*	Yes
	9	9p11.2	44247866	44705656	458	Loss	16	*LOC643648*	Yes
6B	2	2p22.3	34709689	34727867	18	Loss	21	None	Yes
	2	2q14.3	129638490	129640285	2	Loss	11	None	Yes
	3	3q26.1	162551776	162619878	68	Gain	39	*SPTSSB*	No
	12	12p11.1	33301406	33306843	5	Loss	10	None	Yes
	17	17q21.32	44214888	44362186	147	Gain	89	*TTLL6,CALCOCO2, ATP5G1, UBE2Z*	Yes

Note: Chr=Chromosome, DGV=Database of Genomic Variants.

2B.3.1 Twin Pair 1

The affected male patient in twin pair 1 was diagnosed with a psychotic disorder at age 19. He had 41, 53, and 36 raw CNV calls by each of the three (A, P, or p) programs, respectively. In comparison, his unaffected co-twin had 50, 72, and 41 raw CNV calls by the three methods. After CNV merging and discarding of CNVs that were not called by the combination of A/P/p, the affected and unaffected members of this twin pair yielded 14 and 12 CNVs, respectively. These CNVs fell into three categories; shared between the twin pair (10), unique to unaffected (2), and unique to the affected member of twin pair 1 (4). The two CNVs that were found to be unique to the unaffected twin were a loss at 2q22.1 and a gain at 7q35. These CNVs overlapped the genes THSD7B and TPK1, respectively (Table 2.4). Of the four CNVs that were found to be unique to the affected member, none are reported in the DGV. Three of these CNVs (6q14.1, 7q31.1, 14q21.2) cover no gene overlaps while one (17q21.31) covers the PYY gene (Table 2.3). Interestingly, the 14q21 region has been previously implicated in bipolar disorder (Liu *et al*., 2003).

2B.3.2 Twin Pair 2

The affected female patient in twin pair 2 was diagnosed with schizoaffective disorder at age 27. She had 48, 66, and 41 raw CNV calls by each of the three (A, P, or p) programs, respectively. In comparison, her unaffected co-twin had 46, 65, and 42 raw CNV calls. After CNV merging and discarding of CNVs that were not called by the combination of A/P/p, 12 and 15 CNVs remained in the affected and unaffected twins, respectively. Of these, 11 were shared between the twin pair, 4 were unique to unaffected, and 1 was unique to the affected member of this twin pair. The four CNVs that were found to be unique to the unaffected member were a loss at 1q21.1 that overlapped two genes, a loss at 7q21.2 that overlapped no genes, a gain at 17q21.32 that overlapped six genes, and a loss at 19q13.13-19q13.2 that overlapped eight

genes (Table 2.4). The CNV that was found to be unique to the affected member was a gain found in the region 11p15.1 that does not overlap with any known genes and has been previously reported in the DGV (Table 2.3). No CNVs identified in twin pair 2 have been reported for any neurodevelopmental disorder and the observations do not seem to be likely candidates to explain the discordance for schizophrenia seen in this twin pair.

2B.3.3 Twin Pair 3

The affected female in twin pair 3 was diagnosed with paranoid schizophrenia at age 22. She had 47, 54, and 25 raw CNV calls by each of the three (A, P, or p) programs, respectively. In comparison, her twin sister had 44, 43, and 21 raw CNV calls. The merging and discarding of non-overlapping CNVs yielded 10 and 7 CNVs in the affected and unaffected twins, respectively. Of these, six of the CNVs were shared between twin pair 3, one was unique to unaffected, and four were unique to the affected member. The CNV that was found to be unique to the unaffected member was a loss at 3q26.1 that did not overlap any gene (Table 2.4). The four CNVs that were found to be unique to the affected twin overlapped four regions: 3p11.2-3p11.1, 7q11.21, 11p15.4, and 16p11.1 (Table 2.3). Of these, the loss at 3p11.2-3p11.1 is the only one that was not previously reported in the DGV. This CNV overlapped the EPHA3 gene. The other genes unique to the affected twin were LOC441242, INTS4L2, CCT6P1, SNORA22, OR52N2, LOC283914, LOC146481, and LOC100130700. Yet another identified gene, SNORA22, encodes a small nucleolar ribonucleic acid (RNA), which may guide chemical modifications of other RNAs (Kiss, 2001). Interestingly, CCT6P1 is highly expressed in brain (Velculescu *et al*., 1995). Further, the 16p11 region has been implicated in mental disorders, including psychosis (Steinberg *et al*., 2012).

2B.3.4 Twin Pair 4

The female patient in twin pair 4 was diagnosed with paranoid schizophrenia at age 18. She had 44, 49, and 25 raw CNV calls identified by each of the A, P, or p programs, respectively. In comparison, her unaffected co-twin had 52, 37, and 22 raw CNV calls. After CNV merging and discarding of CNVs that were not called by A/P/p, 10 and 7 CNVs remained in the affected and unaffected members of this twin pair, respectively; seven were shared between the twin pair and three were unique to the affected member. Interestingly, there was no CNV that was unique to the unaffected twin (Table 2.4). Of the three CNVs that were found to be unique to the patient, none are reported in the DGV. Two of the CNVs (3q25.1, 16q23.2) cover no genes while one (2q11.2) covers the KIAA1211L gene (Table 2.3). Interestingly, the 3q25 region has been previously implicated in autism-spectrum disorders (Auranen *et al.*, 2002). Also, KIAA1211L is expressed in the brain and has been reported in bipolar disorder (Scott *et al.*, 2009). Also of interest, this CNV has not been previously identified in the DGV, making this CNV loss a potential candidate for the discordance for schizophrenia of this twin pair.

2B.3.5 Twin Pair 5

The affected male in twin pair 5 was diagnosed with undifferentiated schizophrenia at age 20. He had 56, 54, and 45 raw CNV calls and his unaffected co-twin had 54, 58, and 41 calls identified by the A, P, or p software programs, respectively. After CNV merging and discarding of CNVs that were not called by all three programs, 10 and 18 CNVs remained in the affected and unaffected twins, respectively. These CNVs fell into three categories: shared between the twin pair (8), unique to unaffected (10), and unique to the affected member of twin pair 5 (2). The 10 CNVs that were found to be unique to the unaffected member were found in the regions 1p36.33, 1p21.1, 1q25.2, 2p22.3, 3p14.1, 3q11.2, 4p15.1, 4q24, 8p11.23, and 9p11.2, and overlapped the

PAPPA2, EPHA6, TACR3, ADAM5, LOC643648, LOC283914, LOC146481, and LOC100130700 genes (Table 2.4). Of the two CNVs that were found to be unique to the patient (16p12.3 and 17q21.31), the first, a loss at 16p12.3, overlaps the GPR139 gene and the second, a loss at 17q21.31, overlaps the PYY gene (Table 2.3). PYY, as also discussed in the pair-specific results shown for twin pair 1 above, is a potential candidate for the discordance for schizophrenia identified in this twin pair as well. GPR139 is a gene that is an important mediator of signal transduction. G-protein receptors are predominantly expressed in the brain and are known to play important roles in the central nervous system. GPR139 has been previously reported to be associated with attention deficit hyperactivity disorder (ADHD; OMIM: 143465) (Ebejer *et al*., 2013). Both the CNV loss overlapping the *GPR139* gene and the CNV loss overlapping the PYY gene have not been listed in the DGV. They may represent *de novo* events and candidates for the disease discordance of this twin pair.

2B.3.6 Twin Pair 6

The male patient in twin pair 6 was diagnosed with paranoid schizophrenia at age 16. He had 61, 47, and 50 raw CNV calls as identified by the A, P, or p software programs, respectively. In comparison, his unaffected co-twin had 53, 48, and 53 raw CNV calls. After CNV merging and discarding of CNVs that were not called by all three (A/P/p) programs, 8 and 13 CNVs remained in the affected and unaffected members, respectively. These CNVs fell into three categories: shared between the twin pair (8), unique to unaffected (5), and unique to the affected member of twin pair 6 (0). The five CNVs that were found to be unique to the unaffected member were found at 2p22.3, 2q21.1, 3q26.1, 12p11.1, and 17q21.32. Only the CNV gains at 3q26.1 and 17q21.32 overlapped genes. The CNV at 3q26.1 overlapped the SPTSSB gene and the CNV at 17q21.32 overlapped TTLL6, CALCOCO2, ATP5G1, and UBE2Z (Table 2.3). No CNVs were found to be unique to the affected member of this twin pair. Consequently,

no CNVs identified in twin pair 6 seem to represent candidates to explain their discordance for schizophrenia.

The results outlined above have identified rare and pair-specific CNV differences between monozygotic twins in each of the six twin pairs discordant for schizophrenia. Some CNVs involve single or multiple genes and others represent non-coding genomic regions. Also, a number of the CNVs are not reported in DGV, specifically 10 of 14 events found to be unique to affected twins and 9 of 22 events found to be unique to unaffected twins.

2B.4 Discussion

It has become apparent that CNVs are common in human populations and play a significant role in the etiology of complex diseases, including schizophrenia (Ahn *et al*., 2013; St Clair, 2013). However, it is not easy to identify disease-specific CNVs and establish their mode of action in the causation of the disease. Of special concern is the use of arrays with different degrees of genome coverage and the large number of algorithms available to call CNVs. Although the Affymetrix Human Array 6.0 appears to meet most of the platform criteria, including coverage for CNV calling in humans, a gold standard algorithm for the analysis of data has not been established (Zhang *et al*., 2011). There is a likelihood of false positive results. Despite this, such experiments have generated and continue to generate valuable insights. Reports assessing the use of different software algorithms to analyze the same microarrays have identified a low concordance rate between software programs (Kim *et al*., 2012; Pinto *et al*., 2011). This is likely due to the substantial background noise, which contributes to a false discovery rate of variants (Grayson and Aune, 2011). To avoid this, often two programs are used to call the CNVs and the resulting shared CNVs are considered to be reliable. Although logical, this approach is not totally satisfactory as it may ignore and miss out on some critical results. In this analysis I have focused on shared results and specifically I have used four different software programs to call CNVs. I found a relatively low percentage of concordance

between the calls. This is consistent with findings in the literature (Kim *et al*., 2012; Pinto *et al*., 2011) and highlights the necessity for more stringent guidelines for CNV calling from microarrays. A study conducted by Kim *et al*. (2012) suggested that at least three calling algorithms should be used to ensure the reliability of results (Kim *et al*., 2012).

2B.4.1 Selected Regions/Genes

As stated, I used four CNV calling programs (Golden Helix's SVS, Affymetrix Genotyping Console, Partek Genomics Suite, and PennCNV) and selected CNVs that were called by three methods (Affymetrix Genotyping Console, Partek Genomics Suite, and PennCNV) referred to as A/P/p. I chose 10 CNVs to confirm the results by Real Time PCR. The qPCR results established that one CNV, a CNV loss at 7q11.21 in twin pair 3 was significantly different between twins and matched the expected result from the microarray (Figure 2.6). Further, three CNVs (CNV loss at PYY in twin pair 5, CNV loss at OR52N2 in twin pair 3 and CNV gain at 5q11.2 in twin pair 6) showed the expected trend when analyzed by qPCR but were not statistically significant (Table 2.5).

Taken together, the genes/regions that were either confirmed as different between twins or not confirmed but showed the expected direction were 7q11.21, *EPHA3*, *OR52N2,* 5q11.2, and *PYY*. These genes/regions have been previously identified in the literature, as discussed below:

The 7q11.21 region had a 361 kb loss in twin pair 3 that was confirmed by real time PCR. What is of particular interest is the proximity of this region to the deletion on 7q11.23 responsible for Williams-Beuren syndrome and previously associated with schizophrenia (Mulle *et al.,* 2014).

The *EPHA3* gene showed a significant difference between the members of twin pair 3, however the direction of the difference in the Real Time PCR result was a gain when a loss was expected from the microarray result (Table 2.5). *EPHA3* belongs to the ephrin receptor subfamily of protein-tyrosine kinases that have been implicated in mediating developmental events, particularly in the

nervous system, and have been previously associated with neurodegenerative diseases (Martinez and Soriano, 2005). Also, the *EPHA3* gene is the only CNV identified in twin pair 3 that has not been previously reported in the DGV and therefore has the potential to explain discordance for schizophrenia in this twin pair.

Results on *OR52N2*, an olfactory receptor, suggested a loss in the patient of twin pair 3. Olfactory receptors share a seven-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals (Malnic *et al.*, 2004). Recent reports suggest robust olfactory deficits in schizophrenia patients (Moberg *et al.*, 2013).

Similarly, a loss in the PYY gene was identified by the microarray in twin pairs 1 and 5. However, only in twin pair 5 was a loss suggested by real time PCR. PYY encodes a protein that releases a peptide that inhibits pancreatic secretion and mobility in the gut and has been previously identified to be a cerebrospinal fluid marker for mental illness (Widerlov *et al.*, 1988) and autism spectrum disorders (de Krom *et al.*, 2009). This CNV loss at the PYY gene has not been previously reported in the DGV and therefore may be a potential candidate for the observed discordance.

These results highlight that genes identified by this DNA Microarray and also found to show expected trends in qPCR, are compatible with candidate genes implicated in mental disorders, including schizophrenia.

Table 2.5. TaqMan Real Time PCR Results.

The asterisk refers to a significant difference relative to unaffected co-twin. Significance was defined as a Z-Score <1.75 and 95% confidence level. ^=Right direction, not significant. CNV call refers to Copy Caller Result. (NCBI36/hg18)

Gene/Region	Life Technologies Product Number	Chr	Start	End	Size (kb)	Twin	Expected Based on Microarray	CNV Call in Affected Twin (qPCR)	Result
EPHA3	Hs06670176	3	89394600	89419369	25	3A	Loss	2.3*	Significant, Wrong direction
GPR139	Hs01989810	16	19945650	19965863	20	5A	Loss	2.0	Not significant
OR52N2	Hs01320916	11	5789589	5809449	20	3A	Loss	1.9^	Not significant, Same direction
CACNB2	Hs05191328	10	18444576	18790807	346	1**A**	Gain	1.9	Not significant, Wrong direction
ST8SIA6	Hs05146368	10	17428699	17652707	224	1A	Gain	2.0	Not significant
PYY	Hs05489586	17	39423041	39430053	7	1A	Loss	2.0	Not significant
		17	39423041	39430053	7	5A	Loss	1.9^	Not significant, Same direction
KIAA1211L	Hs04657032	2	98858308	98879625	21	4A	Loss	2.1	Not significant, Wrong direction
5q11.2	Hs06079478	5	57326027	57333521	7	6A	Gain	2.2^	Not significant, Same direction
7q11.21	Hs07521783	7	64594329	64955220	361	3A	Loss	1.5*	Significant, Same direction

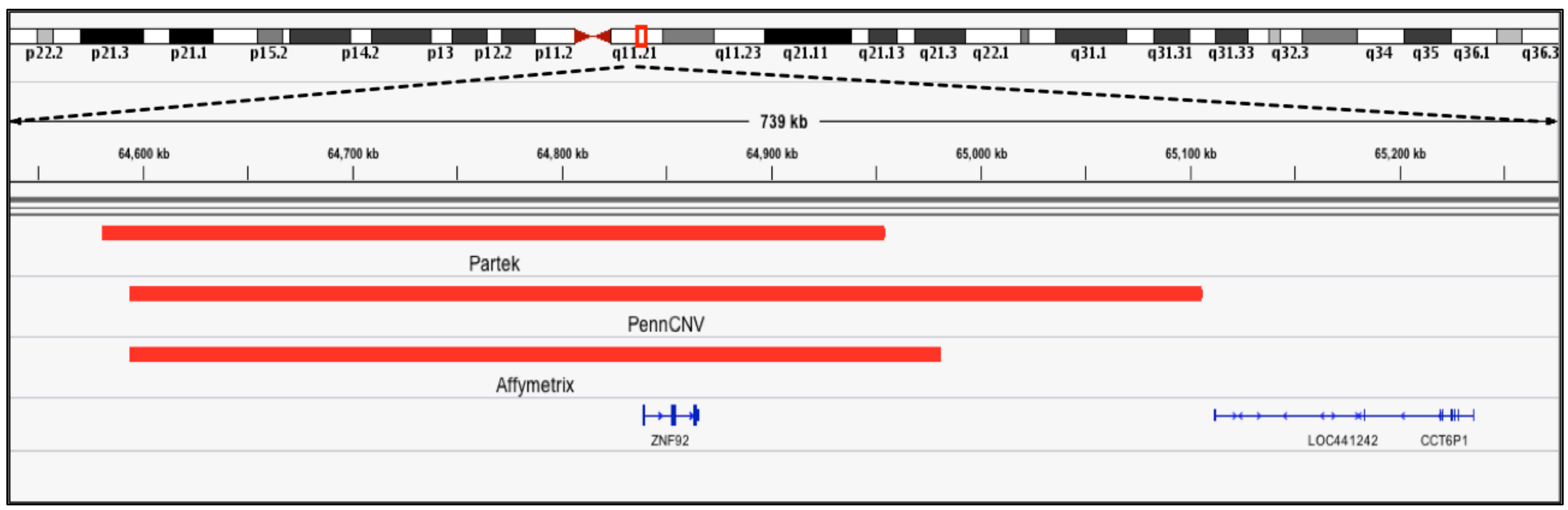


Figure 2.6. Copy Number Variation Loss at 7q11.21 in Twin 3A.

A variant in this location was independently identified using all three programs (Partek, PennCNV and Affymetrix Genotyping Console) (NCBI36/hg18).

2B.4.2 *De novo* Mutation and Discordance

Another challenge with genetic studies in schizophrenia is the extensive heterogeneity that may include multiple genetic, and environmental factors (van Dongen and Boomsma, 2013; Singh and O'Reilly, 2009). Two observations are of particular relevance to this discussion. First, a number of genomic regions and genes including CNVs, both inherited (relatively common) and *de novo* (extremely rare), have been implicated in this complex neurodevelopmental disease (Van Den Bossche *et al*., 2013; Kirov, 2010; Maiti *et al*., 2011). Also, there is less than 100% concordance (48%) between monozygotic twins (McGuffin *et al*., 1994). Consequently, it is not always possible to genetically match controls with patients. This research is exceptional in that it uses monozygotic twins discordant for schizophrenia where the unaffected member represents nature's best-matched genetic control. Further, in some cases the discordance of monozygotic twins for schizophrenia could be attributed to differences in their CNVs potentially caused by *de novo* mutations (DNM) (Maiti *et al*., 2011; Singh *et al*., 2009). The results suggest that DNM are not limited to the germlines alone. Rather they are ongoing throughout life, including stages of differentiation, development, and aging (Lupski, 2010). The occurrence of DNMs has now been demonstrated using a variety of strategies including MZ twins (Bruder *et al*., 2008; Singh *et al*., 2009), trios (Vissers *et al*., 2010), and MZ twins compared with both parents (Maiti *et al*., 2011). Quantitatively, Maiti *et al.* identified one and two DNMs in two pairs of MZ twins respectively, based on parental genotypes, while Vissers *et al.* identified one to two DNMs per trio in eight trios with a mentally retarded proband using family-based exome sequencing (Maiti *et al*., 2011; Vissers *et al*., 2010). The results suggest that DNMs can account for phenotypic discordance between MZ twin pairs. Also, the degree of difference may vary from pair to pair. Their phenotypic impact will depend not only on the genomic region involved but also on the background genotype, and the timing of DNMs during ontogeny. In addition, the mechanism

that is responsible for genomic discordance of MZ twins may generate mosaics, with or without significant phenotypic manifestation (Ruderfer *et al.*, 2013).

My results on six MZD pairs show that each MZD pair differs in rare CNVs. Also, the discordance of some of the pairs could be attributed to CNVs identified in this analysis. For this, I have used the following criteria. First, the CNV should be present in the affected member(s) of the twin pair only. Second, the gene must be expressed in the brain and/or the gene must have relevance to the neurodevelopment and physiological outcomes associated with schizophrenia. Finally, the CNV of interest must not have been identified previously in normal healthy individuals (DGV). The use of these criteria has allowed me to identify potential genes implicated in schizophrenia in four of the six pairs studied. The four pairs that do meet this criterion have their own twin-pair-specific CNVs. Given extensive heterogeneity and the rare nature of *de novo* events; these patterns are expected to be variable. Not surprisingly, the observed differences are pair-specific with respect to the genomic region(s) and gene(s) involved. For example, the region 16p11 was identified as uniquely disrupted in the affected member of twin pair 3 only - this is particularly interesting as this region has been previously associated with psychosis (Steinberg *et al.*, 2012). Although the results are patient and pair-specific, I did find some genomic regions and genes that were common across unrelated patients. An example of a region identified in more than one sample was 2p22.3, the same CNV was uniquely identified in this region in the unaffected member of twin pair 5 and twin pair 6 only — this may suggest a possible protective or mediating effect on the disease from a copy number variation loss in this region.

2B.5 Conclusion

In conclusion, the MZD twin-based genomic (CNV) strategy to identify candidate genes and regions that may be involved in schizophrenia is logical. It has the potential to serve as an effective strategy in identification of genes and genetic mechanisms that may cause complex disorders. Specifically, individuals

with schizophrenia have CNV gains and losses that are likely to contribute to the disease. Here, the inherited mutations may provide predisposition that may not be sufficient for disease manifestation. Occurrence of any/some additional *de novo* events (CNV or mutation) may add to the disease liability threshold and manifest the disease. This model (Maynard *et al*., 2001; Singh *et al*., 2004) of disease development may explain a number of observations on schizophrenia. First, in most cases of discordance in monozygotic twins, even the normal twin may have some or delayed manifestation of some or all symptoms. Second, in most cases of familial schizophrenia, the additional hits may or may not be needed depending on the nature of familial predisposition. Also, in some cases, environmental components may add to the predisposition or be enough by itself to affect neurodevelopment and result in the disease. What is needed to accurately assess this hypothesis are precise and reliable results that unfortunately are not always available. This challenge is apparent from my results where only one of the 10 CNVs identified by the microarray could be statistically confirmed, while three CNVs showed the expected trend but failed to reach the level of significance in qPCR. This follows a number of recent reports (Ehli *et al*., 2012). Despite these limitations, the MZD strategy outlined appears realistic. Specifically, the use of strict criteria for the assessment of CNVs in monozygotic twin genomes discordant for schizophrenia has identified a novel CNV (7q11.21) that is surrounded by low copy repeats with the potential to undergo mechanisms that generate CNVs *de novo*. This confirmed CNV was seen exclusively in the affected patient of twin pair 3 and deserves further investigation as a candidate region for schizophrenia and related disorders in this twin pair and beyond.

2.6 References

Ahn, K., Gotay, N., Andersen, T.M., Anvari, A.A., Gochman, P., Lee, Y., Sanders, S., Guha, S., Darvasi, A., Glessner, J.T., *et al*. (2013). High rate of disease-related copy number variations in childhood onset schizophrenia. Mol. Psychiatry.

Auranen, M., Vanhala, R., Varilo, T., Ayers, K., Kempas, E., Ylisaukko-Oja, T., Sinsheimer, J.S., Peltonen, L., and Jarvela, I. (2002). A genomewide screen for autism-spectrum disorders: evidence for a major susceptibility locus on chromosome 3q25-27. Am. J. Hum. Genet. *71*, 777–790.

Baross, A., Delaney, A.D., Li, H.I., Nayar, T., Flibotte, S., Qian, H., Chan, S.Y., Asano, J., Ally, A., Cao, M., *et al*. (2007). Assessment of algorithms for high throughput detection of genomic copy number variation in oligonucleotide microarray data. BMC Bioinformatics *8*, 368.

Beckmann, H., and Franzek, E. (2000). The genetic heterogeneity of "schizophrenia." World J. Biol. Psychiatry *1*, 35–41.

Bloom, R.J., Kahler, A.K., Collins, A.L., Chen, G., Cannon, T.D., Hultman, C., and Sullivan, P.F. (2013). Comprehensive analysis of copy number variation in monozygotic twins discordant for bipolar disorder or schizophrenia. Schizophr. Res. *146*, 289–290.

Van Den Bossche, M.J., Strazisar, M., Cammaerts, S., Liekens, A.M., Vandeweyer, G., Depreeuw, V., Mattheijssens, M., Lenaerts, A.S., De Zutter, S., De Rijk, P., *et al*. (2013). Identification of rare copy number variants in high burden schizophrenia families. Am. J. Med. Genet. B, Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet. *162B*, 273–282.

Breckpot, J., Thienpont, B., Gewillig, M., Allegaert, K., Vermeesch, J.R., and Devriendt, K. (2012). Differences in Copy Number Variation between Discordant Monozygotic Twins as a Model for Exploring Chromosomal Mosaicism in Congenital Heart Defects. Mol. Syndromol. *2*, 81–87.

Bruder, C.E.G., Piotrowski, A., Gijsbers, A.A.C.J., Andersson, R., Erickson, S., Diaz de Ståhl, T., Menzel, U., Sandgren, J., von Tell, D., Poplawski, A., *et al*. (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am. J. Hum. Genet. *82*, 763–771.

Buchanan, J.A., and Scherer, S.W. (2008). Contemplating effects of genomic structural variation. Genet. Med. *10*, 639–647.

Butler, J.L., Osborne Locke, M.E., Hill, K.A., and Daley, M. (2013). HD-CNV: hotspot detector for copy number variants. Bioinformatics *29*, 262–263.

Chen, J., Cao, F., Liu, L., Wang, L., and Chen, X. (2015). Genetic studies of schizophrenia: an update. Neurosci. Bull. *31*, 87–98.

Conrad, D.F., and Hurles, M.E. (2007). The population genetics of structural variation. Nat. Genet. *39*, S30–S36.

Consortium, A.G.P., Szatmari, P., Paterson, A.D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X.Q., Vincent, J.B., Skaug, J.L., Thompson, A.P., *et al*. (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat. Genet. *39*, 319–328.

Consortium, W.T.C.C., Craddock, N., Hurles, M.E., Cardin, N., Pearson, R.D., Plagnol, V., Robson, S., Vukcevic, D., Barnes, C., Conrad, D.F., *et al*. (2010). Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature *464*, 713–720.

Van Dongen, J., and Boomsma, D.I. (2013). The evolutionary paradox and the missing heritability of schizophrenia. Am. J. Med. Genet. B, Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet. *162B*, 122–136.

Duclos, A., Charbonnier, F., Chambon, P., Latouche, J.B., Blavier, A., Redon, R., Frebourg, T., and Flaman, J.M. (2011). Pitfalls in the use of DGV for CNV interpretation. Am. J. Med. Genet. A *155A*, 2593–2596.

Ebejer, J.L., Duffy, D.L., van der Werf, J., Wright, M.J., Montgomery, G., Gillespie, N.A., Hickie, I.B., Martin, N.G., and Medland, S.E. (2013). Genome-wide association study of inattention and hyperactivity-impulsivity measured as quantitative traits. Twin Res. Hum. Genet. *16*, 560–574.

Eckel-Passow, J., Atkinson, E.J., Maharjan, S., Kardia, S.L.R., and de Andrade, M. (2011). Software comparison for evaluating genomic copy number variation for Affymetrix 6.0 SNP array platform. BMC Bioinformatics *12*, 220.

Ehli, E.A., Abdellaoui, A., Hu, Y., Hottenga, J.J., Kattenberg, M., van Beijsterveldt, T., Bartels, M., Althoff, R.R., Xiao, X., Scheet, P., *et al*. (2012). *De novo* and inherited CNVs in MZ twin pairs selected for discordance and concordance on Attention Problems. Eur. J. Hum. Genet. *20*, 1037–1043.

Feuk, L., Marshall, C.R., Wintle, R.F., and Scherer, S.W. (2006a). Structural variants: changing the landscape of chromosomes and design of disease studies. Hum. Mol. Genet. *15 Spec No*, R57–R66.

Feuk, L., Carson, A.R., and Scherer, S.W. (2006b). Structural variation in the human genome. Nat Rev Genet *7*, 85–97.

First, M.B., Spitzer, R.L., Gibbon, M., and and Williams, J.B.W. (1996). Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV) (Washington, D.C.: American Psychiatric Press, Inc).

First, M.B., Gibbon, M., Spitzer, R.L., Williams, J.B.W., and Benjamin, L.S. (1997). Structured Clinical Interview for DSM-IV Axis II Personality Disorders, (SCID-II) (Washington, D.C.: American Psychiatric Press, Inc.).

Firth, H. V, Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am. J. Hum. Genet. *84*, 524–533.

Freeman, J.L., Perry, G.H., Feuk, L., Redon, R., McCarroll, S. a., Altshuler, D.M., Aburatani, H., Jones, K.W., Tyler-Smith, C., Hurles, M.E., *et al*. (2006). Copy number variation: new insights in genome diversity. Genome Res. *16*, 949–961.

Friedman, J.I., Vrijenhoek, T., Markx, S., Janssen, I.M., van der Vliet, W.A., Faas, B.H., Knoers, N. V, Cahn, W., Kahn, R.S., Edelmann, L., *et al*. (2008). CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. Mol. Psychiatry *13*, 261–266.

Gladkevich, A., Kauffman, H.F., and Korf, J. (2004). Lymphocytes as a neural probe: potential for studying psychiatric disorders. Prog. Neuropsychopharmacol. Biol. Psychiatry *28*, 559–576.

Glessner, J.T., and Hakonarson, H. (2009). Common variants in polygenic schizophrenia. Genome Biol. *10*, 236.

Glessner, J.T., Wang, K., Cai, G., Korvatska, O., Kim, C.E., Wood, S., Zhang, H., Estes, A., Brune, C.W., Bradfield, J.P., *et al*. (2009). Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature *459*, 569–573.

Glessner, J.T., Reilly, M.P., Kim, C.E., Takahashi, N., Albano, A., Hou, C., Bradfield, J.P., Zhang, H., Sleiman, P.M., Flory, J.H., *et al*. (2010). Strong synaptic transmission impact by copy number variations in schizophrenia. Proc. Natl. Acad. Sci. U. S. A. *107*, 10584–10589.

Grayson, B.L., and Aune, T.M. (2011). A comparison of genomic copy number calls by Partek Genomics Suite, Genotyping Console and Birdsuite algorithms to quantitative PCR. BioData Min. *4*, 8.

Halder, A., Jain, M., Chaudhary, I., and Varma, B. (2012). Chromosome 22q11.2 microdeletion in monozygotic twins with discordant phenotype and deletion size. Mol. Cytogenet. *5*, 13.

Hamilton, S.P. (2008). Schizophrenia candidate genes: are we really coming up blank? Am. J. Psychiatry *165*, 420–423.

Henrichsen, C.N., Vinckenbosch, N., Zollner, S., Chaignat, E., Pradervand, S., Schutz, F., Ruedi, M., Kaessmann, H., and Reymond, A. (2009). Segmental copy number variation shapes tissue transcriptomes. Nat. Genet. *41*, 424–429.

Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. Nat. Genet. *36*, 949–951.

Jacquemont, S., Reymond, A., Zufferey, F., Harewood, L., Walters, R.G., Kutalik, Z., Martinet, D., Shen, Y., Valsesia, A., Beckmann, N.D., *et al*. (2011). Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. Nature *478*, 97–102.

Karayiorgou, M., and Gogos, J.A. (2006). Schizophrenia genetics: uncovering positional candidate genes. Eur. J. Hum. Genet. *14*, 512–519.

Kim, S.-Y.Y., Kim, J.-H.H., and Chung, Y.-J.J. (2012). Effect of combining multiple CNV defining algorithms on the reliability of CNV calls from SNP genotyping data. Genomics Inf. *10*, 194–199.

Kirov, G. (2010). The role of copy number variation in schizophrenia. Expert Rev. Neurother. *10*, 25–32.

Kirov, G., Gumus, D., Chen, W., Norton, N., Georgieva, L., Sari, M., O'Donovan, M.C., Erdogan, F., Owen, M.J., Ropers, H.H., *et al*. (2008). Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. Hum. Mol. Genet. *17*, 458–465.

Kiss, T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. EMBO J. *20*, 3617–3622.

Korn, J.M., Kuruvilla, F.G., McCarroll, S. a., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P.J., Darvishi, K., *et al*. (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet *40*, 1253–1260.

De Krom, M., Staal, W.G., Ophoff, R.A., Hendriks, J., Buitelaar, J., Franke, B., de Jonge, M. V, Bolton, P., Collier, D., Curran, S., *et al*. (2009). A common variant in

DRD3 receptor is associated with autism spectrum disorder. Biol. Psychiatry *65*, 625–630.

Lee, C., and Scherer, S.W. (2010). The clinical context of copy number variation in the human genome. Expert Rev. Mol. Med. *12*, e8.

Lee, J.A., and Lupski, J.R. (2006). Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. Neuron *52*, 103–121.

Lin, P., Hartz, S.M., Wang, J.-C., Krueger, R.F., Foroud, T.M., Edenberg, H.J., Nurnberger, J.I., Brooks, A.I., Tischfield, J.A., Almasy, L., *et al*. (2011). Copy number variation accuracy in genome-wide association studies. Hum Hered *71*, 141–147.

Liu, H., Abecasis, G.R., Heath, S.C., Knowles, A., Demars, S., Chen, Y.J., Roos, J.L., Rapoport, J.L., Gogos, J.A., and Karayiorgou, M. (2002). Genetic variation in the 22q11 locus and susceptibility to schizophrenia. Proc. Natl. Acad. Sci. U. S. A. *99*, 16859–16864.

Liu, J., Juo, S.H., Dewan, A., Grunn, A., Tong, X., Brito, M., Park, N., Loth, J.E., Kanyas, K., Lerer, B., *et al*. (2003). Evidence for a putative bipolar disorder locus on 2p13-16 and other potential loci on 4q31, 7q34, 8q13, 9q31, 10q21-24, 13q32, 14q21 and 17q11-12. Mol. Psychiatry *8*, 333–342.

Lupski, J.R. (2010). New mutations and intellectual function. Nat. Genet. *42*, 1036–1038.

Macdonald, J.R., Ziman, R., Yuen, R.K., Feuk, L., and Scherer, S.W. (2013). The database of genomic variants: a curated collection of structural variation in the human genome. Nucleic Acids Res.

Maiti, S., Kumar, K.H.B.G., Castellani, C.A., O'Reilly, R., and Singh, S.M. (2011). Ontogenetic *de novo* copy number variations (CNVs) as a source of genetic individuality: studies on two families with MZD twins for schizophrenia. PLoS One *6*, e17125.

Malnic, B., Godfrey, P.A., and Buck, L.B. (2004). The human olfactory receptor gene family. Proc. Natl. Acad. Sci. U. S. A. *101*, 2584–2589.

Martinez, A., and Soriano, E. (2005). Functions of ephrin/Eph interactions in the development of the nervous system: emphasis on the hippocampal system. Brain Res. Res. Rev. *49*, 211–226.

Maynard, T.M., Sikich, L., Lieberman, J.A., and LaMantia, A.S. (2001). Neural development, cell-cell signaling, and the “two-hit” hypothesis of schizophrenia. Schizophr. Bull. *27*, 457–476.

McCarroll, S.A., and Altshuler, D.M. (2007). Copy-number variation and association studies of human disease. Nat. Genet. *39*, S37–S42.

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

McInnes, L.A., and Lauriat, T.L. (2006). RNA metabolism and dysmyelination in schizophrenia. Neurosci. Biobehav. Rev. *30*, 551–561.

Moberg, P.J., Kamath, V., Marchetto, D.M., Calkins, M.E., Doty, R.L., Hahn, C.G., Borgmann-Winter, K.E., Kohler, C.G., Gur, R.E., and Turetsky, B.I. (2013). Meta-Analysis of Olfactory Function in Schizophrenia, First-Degree Family Members, and Youths At-Risk for Psychosis. Schizophr. Bull.

Moessner, R., Marshall, C.R., Sutcliffe, J.S., Skaug, J., Pinto, D., Vincent, J., Zwaigenbaum, L., Fernandez, B., Roberts, W., Szatmari, P., *et al*. (2007). Contribution of SHANK3 mutations to autism spectrum disorder. Am. J. Hum. Genet. *81*, 1289–1297.

Mulle, J.G., Pulver, A.E., McGrath, J.A., Wolyniec, P.S., Dodd, A.F., Cutler, D.J., Sebat, J., Malhotra, D., Nestadt, G., Conrad, D.F., *et al.* (2014). Reciprocal duplication of the Williams-Beuren syndrome deletion on chromosome 7q11.23 is associated with schizophrenia. Biol. Psychiatry *75*, 371–377.

Need, A.C., Ge, D., Weale, M.E., Maia, J., Feng, S., Heinzen, E.L., Shianna, K. V, Yoon, W., Kasperaviciute, D., Gennarelli, M., *et al*. (2009). A genome-wide investigation of SNPs and CNVs in schizophrenia. PLoS Genet. *5*, e1000373.

O’Reilly, R.L., and Singh, S.M. (1996). Retroviruses and schizophrenia revisited. Am. J. Med. Genet. *67*, 19–24.

Pang, A.W., MacDonald, J.R., Pinto, D., Wei, J., Rafiq, M. a., Conrad, D.F., Park, H., Hurles, M.E., Lee, C., Venter, J.C., *et al*. (2010). Towards a comprehensive structural variation map of an individual human genome. Genome Biol *11*, R52–R52.

Pawitan, Y., Michiels, S., Koscielny, S., Gusnanto, A., and Ploner, A. (2005). False discovery rate, sensitivity and sample size for microarray studies. Bioinformatics *21*, 3017–3024.

Pinto, D., Marshall, C., Feuk, L., and Scherer, S.W. (2007). Copy-number variation in control population cohorts. Hum. Mol. Genet. *16 Spec No*, R168–R173.

Pinto, D., Darvishi, K., Shi, X., Rajan, D., Rigler, D., Fitzgerald, T., Lionel, A.C., Thiruvahindrapuram, B., Macdonald, J.R., Mills, R., *et al.* (2011). Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nat. Biotechnol. *29*, 512–520.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., *et al.* (2006). Global variation in copy number in the human genome. Nature *444*, 444–454.

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kahler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., *et al.* (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat. Genet. *45*, 1150–1159.

Ruderfer, D.M., Chambert, K., Moran, J., Talkowski, M., Chen, E.S., Gigek, C., Gusella, J.F., Blackwood, D.H., Corvin, A., Gurling, H.M., *et al.* (2013). Mosaic copy number variation in schizophrenia. Eur. J. Hum. Genet.

Scott, L.J., Muglia, P., Kong, X.Q., Guan, W., Flickinger, M., Upmanyu, R., Tozzi, F., Li, J.Z., Burmeister, M., Absher, D., *et al.* (2009). Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. Proc. Natl. Acad. Sci. U. S. A. *106*, 7501–7506.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., MÃ¥nÃ©r, S., Massa, H., Walker, M., Chi, M., Navin, N., *et al.* (2004). Large-scale copy number polymorphism in the human genome. Science (80-.). *305*, 525–528.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., *et al.* (2007). Strong association of *de novo* copy number mutations with autism. Science (80-.). *316*, 445–449.

Singh, S.M., and O'Reilly, R. (2009). (Epi)genomics and neurodevelopment in schizophrenia: monozygotic twins discordant for schizophrenia augment the search for disease-related (epi)genomic alterations. Genome *52*, 8–19.

Singh, S.M., Murphy, B., and O'Reilly, R. (2002). Epigenetic contributors to the discordance of monozygotic twins. Clin. Genet. *62*, 97–103.

Singh, S.M., McDonald, P., Murphy, B., and O'Reilly, R. (2004). Incidental neurodevelopmental episodes in the etiology of schizophrenia: an expanded model involving epigenetics and development. Clin. Genet. *65*, 435–440.

Singh, S.M., Castellani, C.A., and O'Reilly, R.L. (2009). Copy number variation showers in schizophrenia: an emerging hypothesis. Mol. Psychiatry *14*, 356–358.

St Clair, D. (2013). Structural and copy number variants in the human genome: implications for psychiatry. Br. J. Psychiatry *202*, 5–6.

Stankiewicz, P., and Lupski, J.R. (2010). Structural variation in the human genome and its role in disease. Annu. Rev. Med. *61*, 437–455.

Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O.P., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E., *et al*. (2008). Large recurrent microdeletions associated with schizophrenia. Nature.

Steinberg, S., de Jong, S., Mattheisen, M., Costas, J., Demontis, D., Jamain, S., Pietilainen, O.P., Lin, K., Papiol, S., Huttenlocher, J., *et al*. (2012). Common variant at 16p11.2 conferring risk of psychosis. Mol. Psychiatry.

Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C., *et al*. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science *315*, 848–853.

Sullivan, P.F. (2005). The genetics of schizophrenia. PLoS Med. *2*, e212.

Sullivan, P.F., Kendler, K.S., and Neale, M.C. (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch. Gen. Psychiatry *60*, 1187–1192.

Teo, S.M., Pawitan, Y., Ku, C.S., Chia, K.S., and Salim, A. (2012). Statistical challenges associated with detecting copy number variations with next-generation sequencing. Bioinformatics *28*, 2711–2718.

Torkamani, A., Dean, B., Schork, N.J., and Thomas, E.A. (2010). Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. Genome Res. *20*, 403–412.

Torrey, E.F., Miller, J., Rawlings, R., and Yolken, R.H. (1997). Seasonality of births in schizophrenia and bipolar disorder: a review of the literature. Schizophr. Res. *28*, 1–38.

Tsuang, M.T.M., Nossova, N., Yager, T., Tsuang, M.T.M., Guo, S.C., Shyu, K.G., Glatt, S.J., and Liew, C.C. (2005). Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. Am. J. Med. Genet. B, Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet. *133B*, 1–5.

Valsesia, A., Stevenson, B.J., Waterworth, D., Mooser, V., Vollenweider, P., Waeber, G., Jongeneel, C. V, Beckmann, J.S., Kutalik, Z., and Bergmann, S. (2012). Identification and validation of copy number variants using SNP genotyping arrays from a large clinical cohort. BMC Genomics *13*, 241.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. Science *270*, 484–487.

Verveer, J., Huizer, K., Fekkes, D., and van Beveren, N.J. (2007). Gene-expression profiling in schizophrenia: an overview. Tijdschr. Psychiatr. *49*, 7–9.

Vissers, L.E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., van Lier, B., Arts, P., Wieskamp, N., del Rosario, M., *et al*. (2010). A *de novo* paradigm for mental retardation. Nat. Genet. *42*, 1109–1112.

Wain, L. V, Pedroso, I., Landers, J.E., Breen, G., Shaw, C.E., Leigh, P.N., Brown, R.H., Tobin, M.D., and Al-Chalabi, A. (2009). The role of copy number variation in susceptibility to amyotrophic lateral sclerosis: genome-wide association study and comparison with published loci. PLoS One *4*, e8175.

Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., Malhotra, D., Bhandari, A., *et al*. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science *320*, 539–543.

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

Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J.T., Abrahams, B.S., Salyakina, D., Imielinski, M., Bradfield, J.P., Sleiman, P.M., *et al*. (2009). Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature *459*, 528–533.

Widerlov, E., Lindstrom, L.H., Wahlestedt, C., and Ekman, R. (1988). Neuropeptide Y and peptide YY as possible cerebrospinal fluid markers for major depression and schizophrenia, respectively. J. Psychiatr. Res. *22*, 69–79.

Winchester, L., Yau, C., and Ragoussis, J. (2009). Comparing CNV detection methods for SNP arrays. Br. Funct Genomic Proteomic *8*, 353–366.

Wineinger, N.E., and Tiwari, H.K. (2012). The impact of errors in copy number variation detection algorithms on association results. PLoS One *7*, e32396–e32396.

Xu, B., Roos, J.L., Levy, S., van Rensburg, E.J., Gogos, J.A., and Karayiorgou, M. (2008). Strong association of *de novo* copy number mutations with sporadic schizophrenia. Nat. Genet. *40*, 880–885.

Yavas, G., Koyutürk, M., Özsoyo, M., Gould, M.P., Koyuturk, M., Ozsoyoglu, M., Gould, M.P., and LaFramboise, T. (2009). An optimization framework for unsupervised identification of rare copy number variation from SNP array data. Genome Biol. *10*, R119–R2009 – 10–10 – r119. Epub 2009 Oct 23.

Zhang, D., Qian, Y., Akula, N., Alliey-Rodriguez, N., Tang, J., Study, B.G., Gershon, E.S., and Liu, C. (2011). Accuracy of CNV Detection from GWAS Data. PLoS One *6*, e14511.

Zogopoulos, G., Ha, K.C., Naqib, F., Moore, S., Kim, H., Montpetit, A., Robidoux, F., Laflamme, P., Cotterchio, M., Greenwood, C., *et al.* (2007). Germ-line DNA copy number variation frequencies in a large North American population. Hum. Genet. *122*, 345–353.

A version of this chapter has been submitted for publication in *Nature Scientific Reports*

Chapter 3 - Whole Genome Sequencing of Monozygotic Twins Discordant for Schizophrenia Defines *de novo* Features of Twin Differences

3.0 Overview

I analyzed the complete genomes of two pairs of monozygotic twins discordant for schizophrenia, including one representing a family tetrad (Family 1 - Mother, Father and their discordant monozygotic twins). I found multiple sequence differences between co-twins. They were caused by a variety of genetic mechanisms and included 4,275 (Family 1) and 10,725 (Family 2) small sequence changes (single nucleotide variants, small indels and block substitutions of 2+ bps), 5 (Family 1), and 5 (Family 2) copy number variations (CNVs) as well as 41 (Family 1), and 141 (Family 2) structural variants (SVs), that were not shared between co-twins. Based on their absence in the two parents of family 1, I was also able to assign some of the events as *de novo*. Interestingly, the genes affected by these changes belonged to a number of canonical pathways that have the potential to provide genetic predisposition (liability) for the development of schizophrenia in the two families. The results may account for the extensive heterogeneity observed in schizophrenia. Two specific pathway defects, Glutamate Receptor Signaling and Dopamine Feedback in cAMP Signaling Pathways, were uniquely found in the two patients. The results show that errors in multiple genes via multiple mutational mechanisms affecting different biochemical pathways may contribute to the threshold of liability in the manifestation of this complex disease. While requiring further independent confirmation, the results support the proposition that *de novo* events are important contributors to the development of schizophrenia including its discordance in monozygotic twins.

3.1 Introduction

Monozygotic (MZ) twins that originate from a single fertilized zygote have been used to study the relative contribution of nature and nurture on a variety of phenotypes and disorders for well over a century. Schizophrenia, which is among the most devastating of the major mental health disorders, is thought to have both genetic and environmental causes (Gottesman and Shields, 1967). Results show that although the frequency of schizophrenia is only 1% in the general population, its concordance in MZ twins approximates only ~50% and not 100%, as might be expected (Cardno and Gottesman, 2000). Further, recent and rare genomic results have shown that MZ twins may differ for *de novo* CNVs (Bruder and Piotrowski, 2008; Maiti *et al*., 2011), as well as epigenetic features (Dempster *et al*., 2011, 2014; Ehli *et al*., 2012; Fraga *et al*., 2005). However, the timing, rate, extent and impact of such *de novo* events have been difficult to confirm. It is assumed that post-zygotic (somatic) *de novo* mutations may occur anytime during development. Thus, *de novo* mutations appear to be a normal aspect of development of some organs (Muotri and Gage, 2006) and cell types (Muotri *et al*., 2010). Depending on the developmental stage at which they arise, these changes may be present in all or almost all cells of an individual or represent mosaicism. The random nature of these *de novo* events is expected to differ between MZ twins that started independent life from a single zygote and thus may contribute to twin discordance. Although logical and attractive as a potential hypothesis for the cause of discordance in MZ twins, an experimental approach based on this hypothesis faces two challenges. The first is the complete identification of all or almost all changes that exist between the genomes of discordant MZ twins and the second is establishing the role of the observed differences in disease discordance. Although the former is gradually becoming possible through increased resolution of genomic technologies, the latter remains challenging and will demand diligent efforts (Li *et al*., 2014; Rall *et al*., 2015).

The challenge of analysis and interpretation of complete genome sequences is attributed to a variety of factors (Baranzini *et al*., 2010). First, the sequence coverage is not always 100% (due to difficult to sequence regions and assembly errors) and second, the conclusions are not always easy to confirm due in part to expected mosaicism (Petersen *et al*., 2014). This research identifies differences between two pairs of MZ twins discordant for schizophrenia. Also, it includes the parental sequences of one twin pair. Such direct observations not only identify the differences between twins, they also offer direct evidence for their pre-zygotic (germline) or post-zygotic (somatic) origin. The next question is the involvement of these differences in the development of schizophrenia; this of course is a much more complicated question. Schizophrenia is a highly heterogeneous and severe psychiatric disorder that affects ~1% of the world's population, and has a heritability estimate of ~80% (Gottesman, 1991; Gottesman and Shields, 1967). Extensive research on the genetics of this disorder has generated a large number of candidate genes, but no causal genes. This includes a recent report involving 150,000 individuals with over 35,000 schizophrenic patients that has implicated over 108 genetic loci (Ripke *et al*., 2014). The results allow for assessment of previously identified loci in genome-wide studies of *de novo* changes between MZ twins that are discordant for schizophrenia. I note that some previously identified loci include *DRD2*, a common anti-psychotic target, as well as a number of glutamate receptors (*GRIA1, GRIN2A, GRM3*), members of the voltage gated calcium channels (*CACNA1C, CACNA1I* and *CACNB2*) and genes involved in synaptic plasticity (Ripke *et al*., 2014). As it stands, there are no common variants that account for a substantial portion of the liability to develop this disease (Kavanagh *et al*., 2014). These findings demonstrate that the disease is highly heterogeneous and that most patients are genetically distinct.

3.2 Methods

3.2.1 Subjects

This research started following ethics approval from the Committee on Research Involving Human Subjects at The University of Western Ontario, London, Ontario, Canada (Appendix B). The families were recruited and assessed in-person by Richard O'Reilly (Psychiatrist), using the Structured Clinical Interview for DSM-IV (SCID I and SCID II) and a review of available medical records (First *et al.*, 1996, 1997). All participants provided written informed consent. A second senior psychiatrist independently reviewed videotapes of the structured interviews of the twins and confirmed the diagnoses. All 6 subjects (Figure 3.1) provided blood and cheek swab samples that were used to isolate genomic DNA using the PerfectPure DNA blood kit (Blood samples), and the QIAGEN DNA Micro Kit (Buccal samples), following the manufacturer's protocols. The experimental workflow of the analysis is summarized in Figure 3.2.

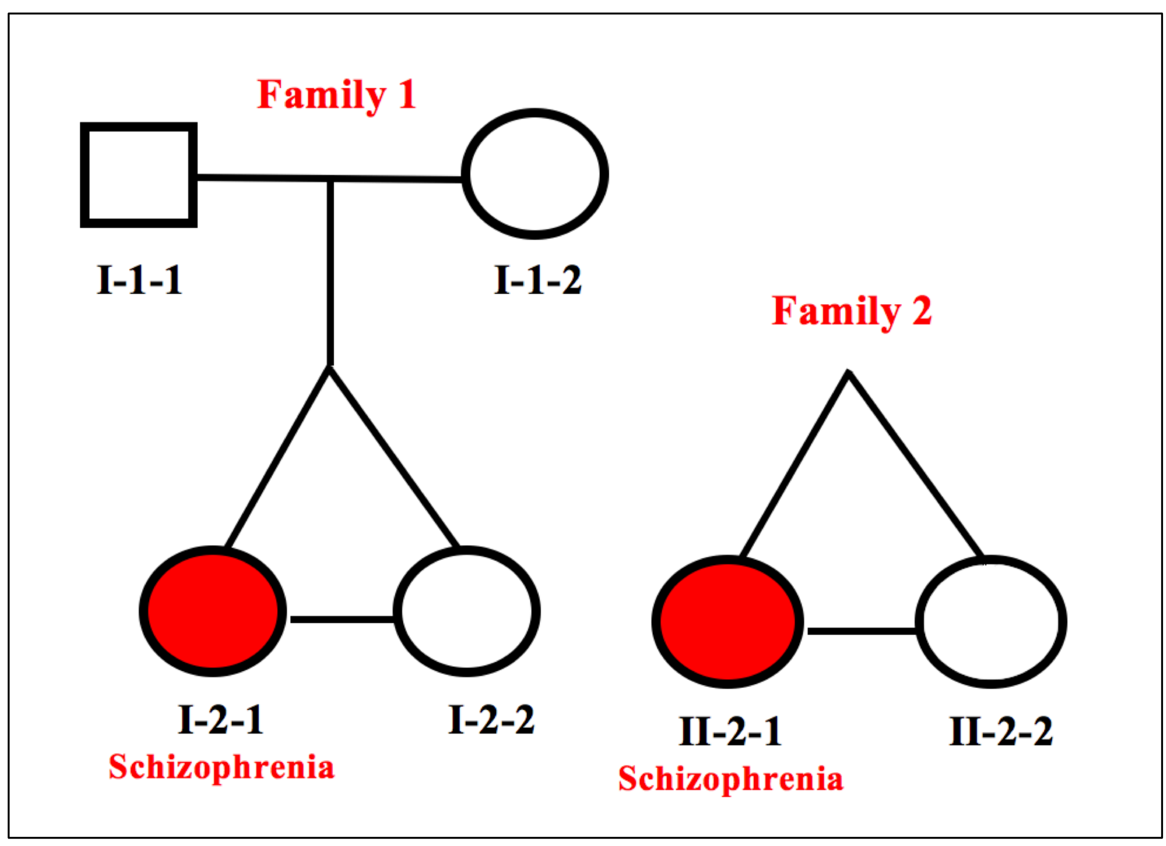


Figure 3.1. Members of the two discordant monozygotic twin pairs and one set of parents included in the study.

DNA Was Extracted from Blood Samples and Buccal Swabs (N=6)

Two pairs of Monozygotic Twins Discordant for Schizophrenia and One Set of Parents

Complete Genomics Sequencing (Mountain View, California)

Average Coverage 48x, Reads and Alignment to NCBI Build 37

Variant Calling - CGATools

SNVs, Substitutions, Indels, CNVs, Structural Variants

Variant Filtering and Annotation - Ingenuity Variant Analysis

Filtering based on Call Quality (≥100), Read Depth (≥50), Region Identity, Freq in Ref Genomes

Confirmation of Variants Using Real Time PCR and Sanger Sequencing

Some confirmed variants were further assessed in Buccal DNA from the same individual

Figure 3.2. Flowchart of Methods Employed.

The twins from Family 1 are Afro-American females and contributed DNA samples at age 53. The affected twin (I-2-1) was diagnosed with schizophrenia at age 22. She was admitted to a psychiatric unit on at least 12 occasions and experienced paranoid delusions, auditory hallucinations and occasionally euphoria during her acute episodes of illness. This twin had significant functional decline and lived with her parents throughout her adult life. The diagnosis of schizophrenia was confirmed by the study psychiatrist. I-2-2 did not experience any significant symptoms of mental illness until she developed a brief episode of depression at age 48, which was treated with an antidepressant. Four years later this twin developed an acute episode of mania which required hospitalization and she was placed on the mood stabilizer divalproex sodium. She was diagnosed as having a bipolar disorder by the study psychiatrist. This twin pair has been previously described in detail (O'Reilly *et al*., 2013). At the time of sample collection, the twins were discordant for schizophrenia for 31 years. The father (I-1-1) of the twins was aged 82 at assessment and had a mild obsessive-compulsive personality disorder but no other psychiatric problems, while the mother (I-1-2) was aged 74 and has never had any psychiatric problems.

At the time of sample collection, the twins from Family 2 were 43 years of age. The twins are Caucasian females. Twin II-2-1 became ill at age 27. She experienced paranoid delusions and hallucinations, which usually occurred in the context of a euphoric or depressed mood. She lived with her twin sister and had worked intermittently. She was diagnosed with a schizoaffective disorder. The twin sister (II-2-2) sustained skeletal injuries in a motor vehicle accident at age 18. A boyfriend died close to the time of this accident and II-2-2 became depressed for a short period around this time. She has had no further episodes of depression or other emotional or psychiatric problems. The twins were considered discordant for 16 years at the time of sample collection.

3.2.2 Complete Genomics Sequencing and Calling of Variants

The genome sequence of the six subjects (Figure 3.1) was generated at Complete Genomics Inc. (Mountain View, CA) in the form of ~2 billion overlapping 70-base nucleotide sequences. They allowed reconstruction of six individual genomes. The sequences met the criteria of high accuracy (99.999%) and were considered suitable for identification of rare variants including somatic mutations as described by Drmanac *et al* (Drmanac *et al*., 2010). These variants included single nucleotide variants (SNVs), indels and block substitutions (2+ bp), as well as larger variants classified as CNVs and SVs that were called in comparison to reference sequence (NCBI Build 37/hg19).

The Complete Genomics (CG) sequencing approach has been fully described in Drmanac *et al* (2010) and Carnevali *et al*. (2012). Briefly, Complete Genomics sequencing generates reads on self-assembling DNA nanoballs resulting in short mate-pair reads. The Complete Genomics sequencing workflow uses Combinatorial Probe Anchor Ligation (cPAL) chemistry to independently assay each base. Reads are initially mapped using a fast algorithm to find the most likely location of each read in the reference genome. Next, these initial mappings are refined by Complete Genomics local *de novo* assembly at any region of the genome that appears to differ from reference. It should be noted that the coverage at segmental duplication regions is known to be limited when using this platform for sequencing.

In the analysis of family 1, variants that were present in the affected twin and not present in the unaffected twin or their mother or father, were assessed as *de novo.* Also, in the twin analyses of family 2, without parents, variants that were present in the affected twin and not present in the unaffected twin were labeled as *provisional de novo*.

CNVs were called based on a read depth or depth-of-coverage algorithm provided by Complete Genomics (Drmanac *et al*., 2010). Sequence coverage was averaged and then GC bias was corrected for over a fixed window (2 kb) and normalized relative to a set of standard (CG 45 genome reference) genomes

sequenced by Complete Genomics. A Hidden Markov Model (HMM) was used to determine the integer copy number state (0-10).

SVs were detected by identifying discordant mate pair mappings found during the assembly process and were considered their own category apart from CNVs due to the different procedure used to call them. Although, it is recognized that CNVs are a type of structural variant and that some redundancy would be expected from the CNV read depth and SV discordant mate pair analyses. Mate pair mappings where each arm maps to the reference genome but with either an unexpected length between them or an anomalous orientation are subjected to local *de novo* assembly to refine junction breakpoints at single base pair resolution. Complete genomics then generated a HighConfidenceJunctions file that reports junctions that are most likely to be accurate. CGAtools junctions2events was then used to convert high confidence junctions to possible SV events using repeat masker 37 data, NCBI Build 37/hg19 reference file and a refseq gene information file. The junctions2events command in CGAtools identifies likely deletion, inversion and translocation events that are at least 500 bp in size, from the list of high confidence junctions delivered by Complete Genomics. SVs were further refined into 5 categories: deletions, tandem duplications, distal duplications, inter-chromosomal events and inversions. To call small variants (small sequence changes) in the dataset, Master Variation Files (MasterVar) were generated using CGAtools and NCBI Build 37/hg19 as reference. VCF files were then generated from each MasterVar file using CGAtools command mkvcf. VCF files generated this way were directly inputted into Ingenuity Variant Analysis (Ingenuity Systems, Redwood City, CA). Given the potential for false positives identified in genome sequencing, a stringent read depth of 50 and a call quality of 100 (calculated by complete genomics and based on a phred scale) were chosen as parameters for variant filtering. Variants were annotated with overlapping genes, cytoband, gene region (Exonic, Intronic, Promoter, 5'UTR, 3'UTR, Splice Site), translational impact (if applicable), SIFT function prediction (if applicable), SIFT score (if applicable), dbSNP ID (if

applicable) and frequency in the 1000 Genomes as well as frequency in the Complete Genomics Public Genome dataset. The biological context filter in Ingenuity Variant Analysis was applied to downstream data with “Schizophrenia, Neurological Disease and diseases consistent with those two phenotypes” as the filtering criteria. Genes had to have a direct connection to these phenotypes (no hops allowed).

In the case of CNVs and SVs a 50% reciprocal overlap rule (Pang *et al*., 2010; Wain *et al*., 2009; Yavas *et al*., 2009) was applied to determine if two variants were the same or different between twins/parents. All CNVs/SVs had to meet the ≥ 50% rule when compared to the other CNV/SV, otherwise, the two were considered unique. Pairwise comparisons were performed for all CNVs and SVs. To increase the efficiency of this, HD-CNV (Hotspot Detector for Copy Number Variants) was used (Butler *et al*., 2013). Files were prepared using a custom python script and the 50% reciprocal overlap rule was applied to determine the CNVs and SVs that were unique to an individual. Interchromosomal and inversion comparisons were analyzed manually due to the limitations of HD-CNV. Inter-chromosomal events were considered the same if the origin and the destination chromosome numbers matched and the junction positions were less than 500 bp apart. Inversions had to share the same direction and 50% or more identity to be classified as shared.

Circos (Krzywinski *et al*., 2009) was used for visualization of post-filtering variation data. Unique variants were separated into files to generate relevant tracks in the genome diagram. Structural variants that involved more than one chromosome were visualized using “links” in Circos.

3.2.3 Confirmation of Findings

Quantitative PCR using TaqMan® Assays in an Applied Biosystem StepOne were performed on selected CNVs and SVs on blood samples. CNVs and SVs that were confirmed via Real Time PCR in blood were then assessed in the Buccal DNA from the same individuals to identify if the variant arose before

or after the differentiation of the germ layers. RNAseP was used as the reference gene for normalization as it is known to be present in two copies in every human diploid genome. RNAseP is a type of ribonuclease that cleaves RNA. Predesigned TaqMan assays were ordered from Life Technologies for regions of interest. CopyCaller 2.0 was used to generate the predicted copy numbers. Each Real Time experiment was repeated four times.

Sanger Sequencing was used to assess small variants of interest in the dataset that were unique to the affected twin. Variants were chosen based on three criteria, 1) They were exonic; 2) They had high call qualities and, 3) They had high read depths. Regions were sequenced at the London Regional Genomics Centre (LRGC). The selected variants (24) were sent for Sanger Sequencing using traditional forward and reverse primers. Variants that appeared mosaic in nature were further assessed using PASA (PCR Amplification of Specific Alleles). PASA uses an allele specific primer to preferentially amplify DNA with a specific allele of interest.

3.2.4 Pathway Analysis

Any unique variant that overlapped a gene was identified and these gene lists were used in Core Pathway Analysis (Ingenuity Systems, Redwood City, CA) to identify pathways and networks that were overrepresented in the filtered dataset. Only the genes that were identified to overlap high confidence *de novo* (Family 1) or *provisional de novo* (Family 2) variations were included in the pathway analysis. Variants that were found to be unique to unaffected twins were also assessed in a separate pathway analysis. The top 20 canonical pathways found to be overrepresented in each individual were then compared between all members in the study; pathways that were found in affected and unaffected twins were labeled as gp (genetic predisposition). Pathways that were unique to affected twins only were labeled as gpd (genetic predisposition leading to disease).

3.3 Results

3.3.1 Defining genetic variation including *de novo* events

3.3.1.1 Sequence characteristics of six complete genomes

The complete genome sequences on two MZD twin pairs and one set of parents (Figure 3.1) were generated with a high average call quality above 100 (Figure 3.3) and an average read depth of 47-50 fold coverage (Figure 3.4) following my workflow (Figure 3.2).

They represented over 99% of the reference sequence for each genome (Table 3.1). I compared individual genome sequences with the Genome Reference Consortium Human Genome Build 37 (hg19) using Complete Genomics Analysis Tools (CGATools) for identification of individual variation events and their genomic location at single base pair resolution. The results show that each genome harbors 3.3 to 3.9 million single nucleotide variants (SNVs), 370-430 thousand indels (small insertions and deletions), 71-80 thousand block substitutions (variation in 2+ adjacent nucleotides), 1 thousand structural variations, and 150 copy number variations as well as a transition/transversion ratio of 2.1:1 (Table 3.1). The exception to this pattern was the father in Family 1 who carried 592 copy number variations and 1110 structural variations, that may be associated with a diagnosis of chronic leukemia unrelated to this study (Kumar *et al.*, 2013). I further grouped individual variations as Small Sequence Changes, (SSCs - SNVs, Indels, Block Substitutions), CNVs or SVs (Table 3.1).

Table 3.1. Identity of Sequenced Genomes.

ID	Family	Gender	Age*	Age of Onset	Disease Status	% of Genome Called	Normalized Average Coverage	Small sequence Variations (SNVs, Indels, Block Subs)	Copy Number Variants	Structural Variants	Ti:Tv
I-1-1	Family 1	Male	80	N/A	Unaffected	99.037	49.55	4,392,314	592	1110	2.1:1
I-1-2	Family 1	Female	76	N/A	Unaffected	99.042	47.27	4,093,396	163	904	2.1:1
I-2-1	Family 1	Female	43	27	Affected	99.042	48.07	4,295,920	152	919	2.1:1
I-2-2	Family 1	Female	43	N/A	Unaffected	99.042	47.41	4,265,089	154	893	2.1:1
II-2-1	Family 2	Female	53	22	Affected	99.039	49.71	3,780,127	156	996	2.1:1
II-2-2	Family 2	Female	53	N/A	Unaffected	99.037	50.25	3,789,298	157	977	2.1:1

Note: Ti/Tv = Transition/Transversion Ratio (of variants), Age* = Age at assessment

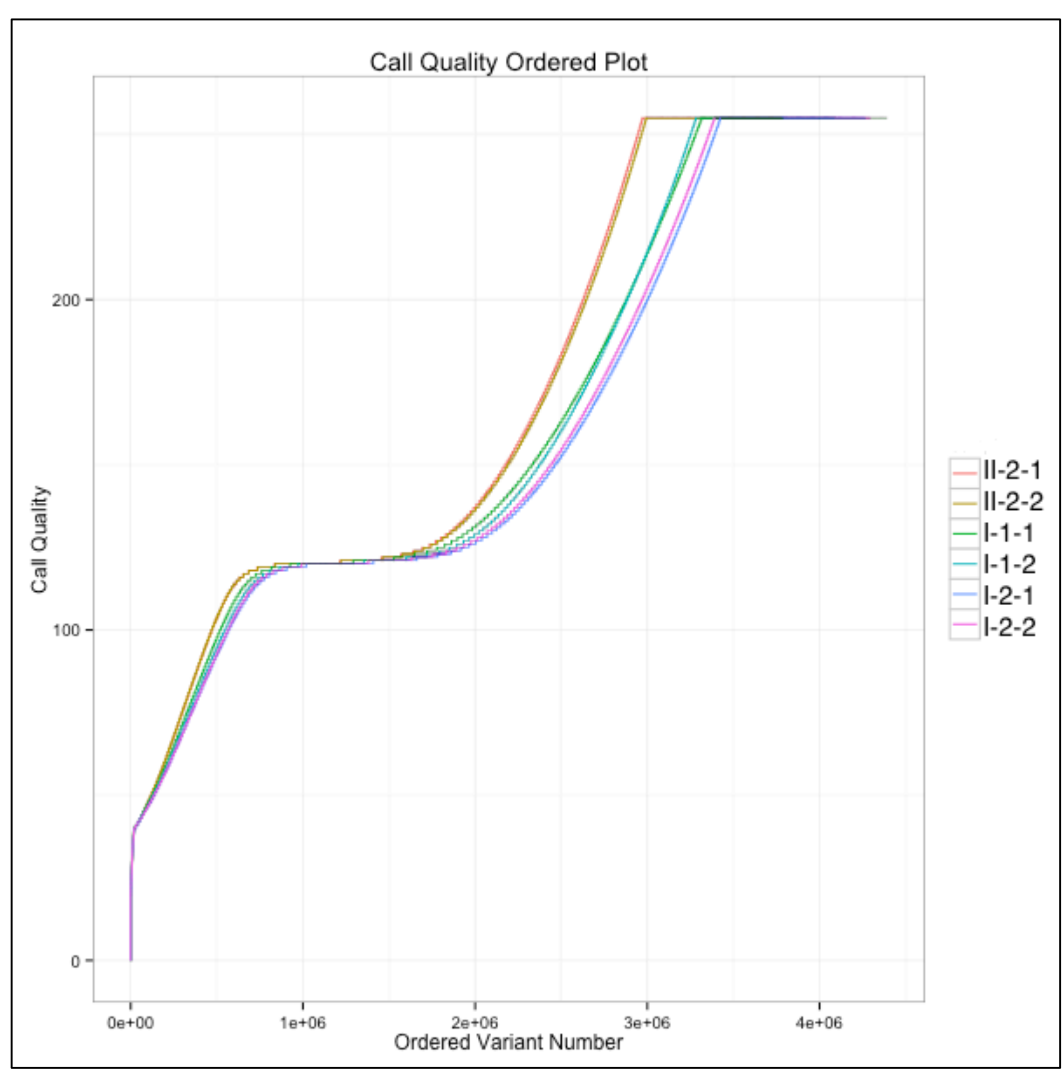


Figure 3.3. Call Quality of raw variant calls in each sample.

Each line represents the plotted call quality of each called variant in that individual sample. Call qualities were first ordered by number and then plotted in order. The Call Quality is provided by Complete Genomics and based on a Phred scale. I am thankful to Marjorie E Locke (PhD Candidate) who created this figure for use in this thesis.

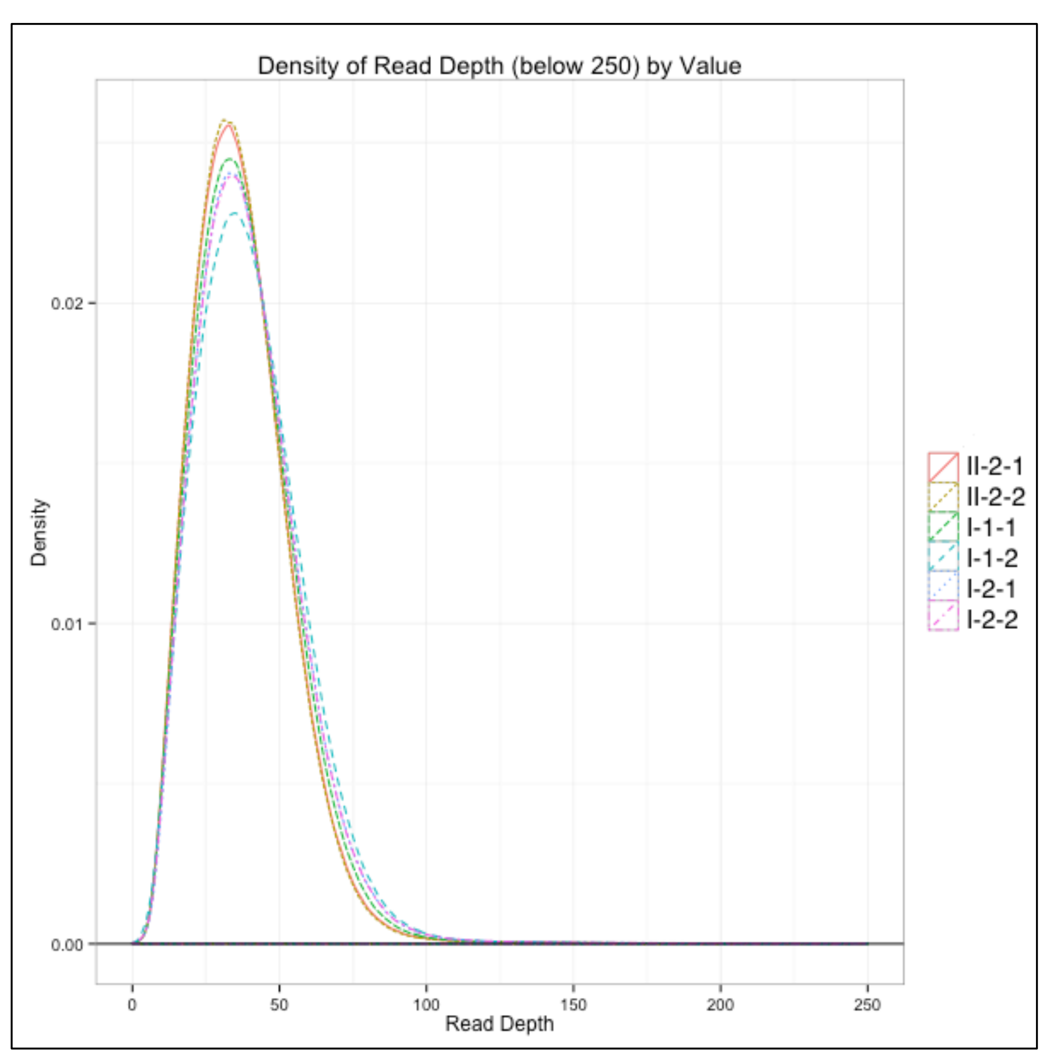


Figure 3.4. Read Depth Density of raw variant calls in each sample.

Each line represents the plotted read depth of each variant called in that individual sample. A density for each read depth was first assigned and then plotted. The read depth represents the number of times that base pair location was sequenced. Outliers (above 250) were not included. I am thankful to Marjorie E Locke (PhD Candidate) who created this figure for use in this thesis.

For each type of mutational event, the highest number of differences was found between unrelated individuals, the least number between MZ twins and an intermediate number between a parent and child. For example, two unrelated individuals differed for approximately 1.5-1.8 million SNVs, a parent and child differed for 0.9-1.0 million SNVs and a pair of MZ twins differed for approximately 100,000 SNVs. These estimates represent raw calls by the algorithm used by the Complete Genomics software (Carnevali *et al.*, 2012) that has been suggested to reliably detect a substantial portion of genetic variation distributed across the genome (Pang *et al.*, 2014). A comparison of raw sequence variations between twins in Family 1, and 2 generated approximately 250,000 and 200,000 differences respectively. Further application of strict confidence filters (see methods) resulted in approximately 11,500 (1.21%) unique variations in the affected twin of Family 1 and approximately 10,500 (1.17%) unique variations in the affected twin of Family 2 (Figure 3.5a and 3.5b). Given that these differences occurred between monozygotic twins, they were interpreted as post zygotic *de novo* mutations in the affected twin as compared to their own unaffected co-twin.

Figure 3.5. Circos plots of high confidence variant calls.

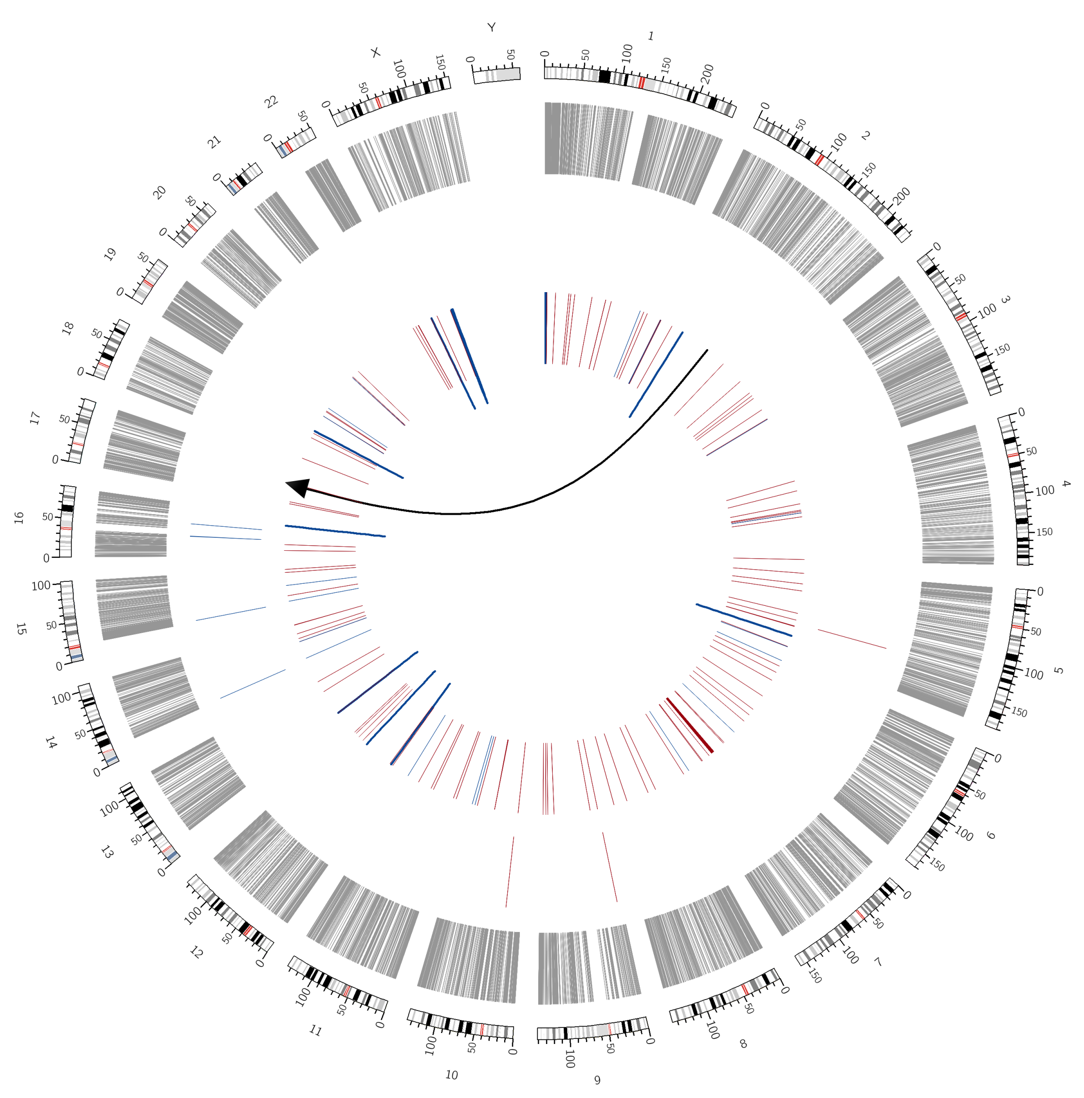


Figure 3.5a. Circos plot of high confidence mutations in the affected member of Family 1 that were not found in the co-twin or either parent. Blue represents a gain and red represents a loss. The outside track displays SSCs, the middle track displays CNVs and the inside track displays SVs. The arrow displays an inter-chromosomal event.

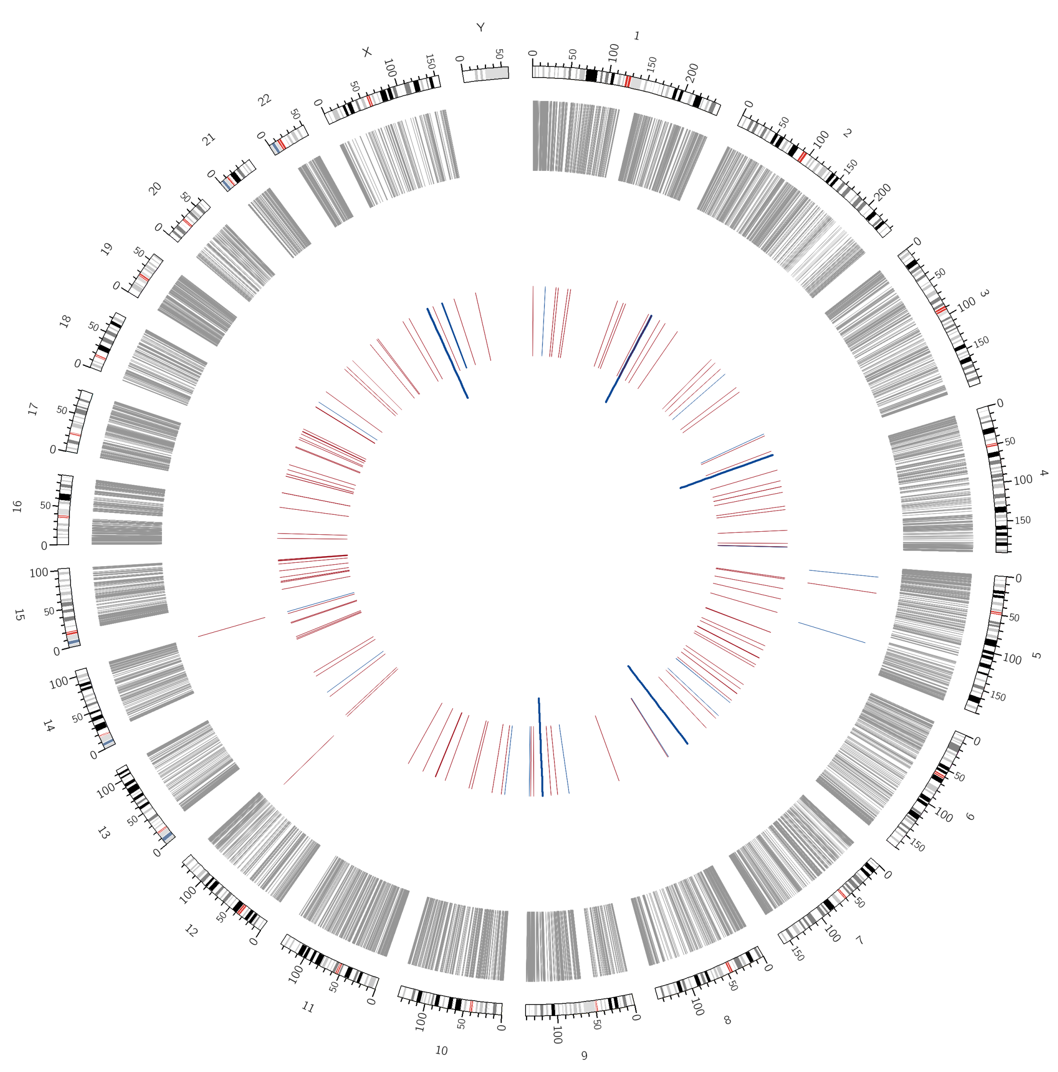


Figure 3.5b. Circos plot of high confidence mutations in the affected member of Family 2 that were not found in the co-twin. Blue represents a gain and red represents a loss. The outside track displays SSCs, the middle track displays CNVs and the inside track displays SVs.

The results presented in the preceding section show that every genome is unique and the difference between any two individuals include both familial as well as *de novo* events. Further, the majority of differences among unrelated individuals appear to be due to population and family history, while the rare differences (0.05%) between monozygotic twins are attributed to novel ontogenetic *de novo* events. In Family 1, 63% of unique small sequence changes were identified as inherited and 37% were identified as *de novo* (Table 3.2).

The *de novo* variants could further be characterized by their presence in exons in each family (Table 3.3a and 3.3b). Conversely, in family 1 where the complete sequences of both parents were available, 2 unique CNVs were identified as inherited and 5 were identified as *de novo* (Table 3.4). The CNVs unique to the affected member of two twin pairs can be seen in Tables 3.5a and 3.5b. SVs followed a pattern similar to small sequence variations with 70% of SVs identified as inherited and 30% identified as *de novo* (Table 3.6). There are a small number of unique CNVs in the affected member of family 1 (unique CNVs=7), 5 of which are not found in either parent *(de novo)* (Table 3.4). In the absence of parental data in Family 2, variations that were unique to the affected twin were classified as "*provisional de novo*". I conclude that most human variations are historical and transmitted across generations, yet a proportion are acquired through *de novo* events during genetic transmission via gametogenesis as well as during mitosis throughout development and aging.

Table 3.2. Shared and *de novo* Small Sequence Changes identified in blood (SNVs, Substitutions, Indels) as compared to NCBI Build 37/hg19.

Sample	Total Sequence Variations Before Confidence Filters Applied	Sequence Variations After Confidence Filters Applied	High Confidence Variants Unshared with Co-twin	Unshared with Co-twin	
				Inherited (Present in parent)	*de novo* (Not present in parent)
Family 1 Affected (I-2-1)	4,295,920	956,470	11,577 (1.2%)	7,302 (63.1%)	4,275 (36.9%)
Family 1 Unaffected (I-2-2)	4,265,089	833,254	9,345 (1.1%)	5,776 (61.8%)	3,569 (38.2%)
Family 2 Affected (II-2-1)	3,780,127	917,899	10,725 (1.2%)	N/A	N/A
Family 2 Unaffected (II-2-2)	3,789,298	926,691	10,351 (1.1%)	N/A	N/A

Table 3.3. Exonic *de novo* variants in Affected Twins.

Table 3.3a. Exonic *de novo* variants in the Affected Twin of Family 1 (I-2-1). Gene variants in blue are not known to be polymorphic and considered novel and those with potential pathogenic impact are marked in red.

Chr	Position	Ref	Sample	Type	Cytoband	Gene	Translation Impact	dbSNP ID	1000G	CG Public
1	117142868	CA	TG	Sub	1p13.1	*IGSF3*	In-frame			
4	85996		C	Ins	4p16.3	*ZNF595*	Frameshift			
4	86004	A	G	SNV	4p16.3	*ZNF595*	Missense	112290390		
4	86022	CAG	GAC	Sub	4p16.3	*ZNF595*	In-frame	386670355		
4	86043	T	C	SNV	4p16.3	*ZNF595*	Synonymous	377364445		
7	100549935	T	G	SNV	7q22.1	*MUC3A*	Missense	73714229		
7	100549942	T	C	SNV	7q22.1	*MUC3A*	Missense	73714230		
7	100552452	T	C	SNV	7q22.1	*MUC3A*	Synonymous	112050489		
11	48367133	C	A	SNV	11p11.2	*OR4C45*	Missense	79453749		
12	40875389	CCATCAGCTGGAGTGAC AGTGACATCCGGA		Del	12q12	*MUC19*	In-frame			
12	40880542	C	T	SNV	12q12	*MUC19*	Missense	370502411		14.81
12	40880545	C	G	SNV	12q12	*MUC19*	Missense	199768257		16.66
12	122359397		GAGGAGGA GGAGAAA	Ins	12q24.31	*WDR66*	In-frame	142042908		67.59
13	25671607	GA	AG	Sub	13q12.13	*PABPC3*	In-frame	386769143		
13	46170723	GATACTCTTCCTCCTCCA		Del	13q14.13	*ERICH6B*	In-frame			
17	74288565	TGA		Del	17q25.1	*QRICH2*	In-frame	35035566	46.96	20.37
22	29885594	A	T	SNV	22q12.2	*NEFH*	Synonymous	79235463		
22	29885599	AGGAAG		Del	22q12.2	*NEFH*	In-frame	149571560		

Note: 1000G=1000 Genomes, Sub: Substitution, Ins: Insertion, Del: Deletion, SNV: Single Nucleotide Variation.

Table 3.3b. Exonic *de novo* variants in the Affected Twin of Family 2 (II-2-1). Gene variants in blue are not known to be polymorphic and considered *de novo* and those with potential pathogenic impact are marked in red.

Chr	Position	Ref	Sample	Type	Cytoband	Gene	Translation Impact	dbSNP ID	1000G	CG Public
1	1650787	T	C	SNV	1p36.33	*CDK11B; CDK11A*	Missense	1137003		28.7
1	1650797	A	G	SNV	1p36.33	*CDK11B; CDK11A*	Missense	1059830		31.48
1	1650801	T	C	SNV	1p36.33	*CDK11B; CDK11A*	Synonymous	1137004		31.48
1	13183225	T	C	SNV	1p36.21	*HNRNPCL1/ HNRNPCL2*	Synonymous	28634306		
1	13183228	C	T	SNV	1p36.21	*HNRNPCL1/ HNRNPCL2*	Synonymous	144054379		
1	13183237	CT	TC	Sub	1p36.21	*HNRNPCL1/ HNRNPCL2*	In-frame			
1	13183243	T	C	SNV	1p36.21	*HNRNPCL1/ HNRNPCL2*	Synonymous	138897759		
1	13183248	TC	AA	Sub	1p36.21	*HNRNPCL1/ HNRNPCL2*	In-frame			
1	17266536	G	C	SNV	1p36.13	*CROCC*	Missense	9435714		10.18
1	62675619	C	T	SNV	1p31.3	*L1TD1*	Synonymous	4625314	29.85	33.33
1	109007867	G	A	SNV	1p13.3	*NBPF4/ NBPF6*	Missense			26.85
1	109007877	T	G	SNV	1p13.3	*NBPF4/ NBPF6*	Synonymous			26.85
1	109737063	C	T	SNV	1p13.3	*KIAA1324*	Synonymous	386565601	71.37	68.51
1	109737090	G	A	SNV	1p13.3	*KIAA1324*	Synonymous	386565600	71.53	67.59
1	111957583	A	G	SNV	1p13.2	*OVGP1*	Missense	1126656		3.7
1	111957592	A	G	SNV	1p13.2	*OVGP1*	Missense	56294468	22.32	3.7
1	120539742	G	A	SNV	1p12	*NOTCH2*	Missense	2258139		
1	120548025	G	A	SNV	1p12	*NOTCH2*	Synonymous	140551270		
1	120548055	T	C	SNV	1p12	*NOTCH2*	Synonymous	199592384		

1	144873963	T		Del	1q21.1	*PDE4DIP*	Frameshift		28.91	32.4
1	144922523	C	T	SNV	1q21.1	*PDE4DIP*	Missense	2455994		28.7
1	145103928	G	A	SNV	1q21.1	*SEC22B*	Synonymous	2596251	34.01	35.18
1	201178819	G	A	SNV	1q32.1	*IGFN1*	Missense	72468019		11.11
1	206566904	T	C	SNV	1q32.1	*SRGAP2B; SRGAP2D; SRGAP2; SRGAP2C*	Synonymous	2919105	49.44	
1	248813827	T	C	SNV	1q44	*OR2T27*	Missense	1782241		
2	132238043	A	C	SNV	2q21.1	*TUBA3C/ TUBA3D*	Synonymous	74625243		14.81
3	75786499	CA	TG	Sub	3p12.3	*ZNF717*	In-frame			0.92
3	75786515	GG	AT	Sub	3p12.3	*ZNF717*	In-frame		0.92	
3	75787265	C	G	SNV	3p12.3	*ZNF717*	Synonymous	144538707		
3	75787269	G	T	SNV	3p12.3	*ZNF717*	Missense			
3	75788023	C	T	SNV	3p12.3	*ZNF717*	Missense	76889571		50
3	100170634	T	C	SNV	3q12.2	*LNP1*	Synonymous	9848109		24.07
3	195510827	C	T	SNV	3q29	*MUC4*	Missense	413807	3.53	24.07
3	195511780	G	A	SNV	3q29	*MUC4*	Missense	391928		
5	140725160		A	Ins	5q31.3	*PCDHGA3; PCDHGA2; PCDHGA1*	Frameshift	372306793		
6	26406255	G	A	SNV	6p22.2	*BTN3A1*	Synonymous	3857550		41.66
6	33037579	AT	TG	Sub	6p21.32	*HLA-DPA1*	In-frame	386699858		37.96
6	33052986	T	C	SNV	6p21.32	*HLA-DPB1*	Synonymous		46.37	37.03
7	100550039		CTC	Ins	7q22.1	*MUC3A*	In-frame			
7	142460313	T	C	SNV	7q34	*PRSS1*	Synonymous	6666	39.66	24.07
7	142460335	A	G	SNV	7q34	*PRSS1*	Missense	201550522	0.16	
8	10467652	GC	CT	Sub	8p23.1	*RP1L1*	In-frame	386722180		
8	103573011	TGCAACCCCTGCAGCCCCTGCAACCCG		Del	8q22.3	*ODF1*	In-frame		31.29	23.14

11	1016776	C	T	SNV	11p15.5	*MUC6*	Missense	33988517		38.88
11	1017110	CGGT	TGCC	Sub	11p15.5	*MUC6*	In-frame			32.4
11	1018069	G	A	SNV	11p15.5	*MUC6*	Missense	10736904		
11	1018090	G		Del	11p15.5	*MUC6*	Frameshift			
11	1018095	G	TA	Sub	11p15.5	*MUC6*	Frameshift			
11	1092684	C	T	SNV	11p15.5	*MUC2*	Synonymous	201269049		
11	5270686	G	A	SNV	11p15.4	*HBG1*	Missense	1061234		75.92
11	18290866	T	C	SNV	11p15.1	*SAA1*	Synonymous	1136745		
11	48346916	G	C	SNV	11p11.2	*OR4C3*	Missense	77069283		
11	48346924	T	C	SNV	11p11.2	*OR4C3*	Synonymous	72911454		
11	48346932	G	A	SNV	11p11.2	*OR4C3*	Missense	80285195	0.28	
11	48347306	G	T	SNV	11p11.2	*OR4C3*	Missense	73465911		33.33
11	48367052		AG	Ins	11p11.2	*OR4C45*	Frameshift			
11	48367073	A	C	SNV	11p11.2	*OR4C45*	Missense	7941588	79.55	75
11	56143357	G	A	SNV	11q12.1	*OR8U1; OR8U8*	Synonymous	76949582		36.11
11	56143963	AATCTA TAGC	GATTT ACAGT	Sub	11q12.1	*OR8U1; OR8U8*	Synonymous			
11	56468020	T	C	SNV	11q12.1	*OR9G1; OR9G9*	Missense	532637	41.21	56.48
11	57982726	G	T	SNV	11q12.1	*OR1S1*	Synonymous	1993089	52.32	52.77
11	57982763	A	G	SNV	11q12.1	*OR1S1*	Missense	61763008		50.92
11	89703619	G	A	SNV	11q14.3	*TRIM64/ TRIM64B*	Missense	79824618		77.77
11	89819403	G	T	SNV	11q14.3	*UBTFL1*	Missense			
12	11214857	T	C	SNV	12p13.2	*TAS2R46; PRH1*	Missense	201891491		
12	11214870	A	T	SNV	12p13.2	*TAS2R46; PRH1*	Synonymous	200226376		
12	11244721	GAA	CAC	Sub	12p13.2	*TAS2R43; PRH1*	In-frame			19.44
12	11244730	AA	GG	Sub	12p13.2	*TAS2R43; PRH1*	In-frame			18.51

12	49522578	T	C	SNV	12q13.12	*TUBA1B*	Synonymous	1057725	54.49	
12	49522605	C	T	SNV	12q13.12	*TUBA1B*	Synonymous	1057548	45.33	
12	52843610	C	A	SNV	12q13.13	*KRT6B*	Missense			
13	46170724	ATACTCTTCCTCC TCCAG		Del	13q14.13	*ERICH6B*	In-frame			
13	53216666	G	A	SNV	13q14.3	*HNRNPA1L2*	Synonymous	113869751	1.08	1.85
13	103399222	G	T	SNV	13q33.1	*CCDC168*	Stop gain			25
18	9887394	C	T	SNV	18p11.22	*TXNDC2*	Synonymous	2240910	51.84	
19	4511197	G	A	SNV	19p13.3	*PLIN4*	Synonymous	199625614		
19	4511200	C	T	SNV	19p13.3	*PLIN4*	Synonymous	200718202	0.14	
19	55286854	A	G	SNV	19q13.42	*KIR2DL1/ KIR2DL3*	Missense	666590	0.08	
19	55286864	A	C	SNV	19q13.42	*KIR2DL1/ KIR2DL3*	Synonymous	77397437	16.63	
21	10942923	G	A	SNV	21p11.1	*TPTE*	Missense	76723236		7.4
22	38120180	CTC		Del	22q13.1	*TRIOBP*	In-frame			
22	39387558	C	T	SNV	22q13.1	*APOBEC3AB/ APOBEC3B*	Synonymous	1065184		

Note: 1000G=1000 Genomes, Sub: Substitution, Ins: Insertion, Del: Deletion, SNV: Single Nucleotide Variation.

Table 3.4. Shared and unshared Copy Number Variants (CNVs) identified in blood as compared to NCBI Build 37/hg19.

Sample	Total CNVs	Shared with Co-twin	Unshared with Co-twin	Shared with Co-twin		Unshared with Co-twin	
				Inherited	*De novo*	Inherited	*De novo*
Family 1 Affected (I-2-1)	152	145 (95.4%)	7 (4.6%)	131/145 (90.3%)	14/145 (9.7%)	2/7 (28.6%)	5/7 (71.4%)
Family 1 Unaffected (I-2-2)	154	145 (94.2%)	9 (5.8%)	131/145 (90.3%)	14/145 (9.7%)	5/9 (55.6%)	4/9 (44.4%)
Family 2 Affected (II-2-1)	156	151 (96.8%)	5 (3.2%)	N/A	N/A	N/A	N/A
Family 2 Unaffected (II-2-2)	157	151 (96.2%)	6 (3.8%)	N/A	N/A	N/A	N/A

3.3.1.2 Mutational characteristics in twins affected with Schizophrenia

(i) Small Sequence Changes (SSCs)

I classified SNVs, small indels and block substitutions as small sequence changes (SSCs). The chromosomal distribution of such variants can be seen in Figure 3.6. The only exception to the expected chromosomal size pattern was chromosome 9, which had on average, less small sequence variations from reference. The identity of *de novo* (family 1) and *provisional de novo* (family 2) high confidence variations in affected twins shows that majority of variants are intergenic SNVs and that each individual harbors approximately the same number of block substitutions, insertions and deletions (Figure 3.7).

Overall, the proportions of variants in each category are similar in the affected twin in family 1 as well as family 2; however, the actual *de novo* variants in the twins of family 2 that represent twin differences and lack parental data are family specific. As expected, the exonic variants comprised a very small number. The exonic *de novo* variations identified were annotated with gene information and a summary of these variations can be seen in Table 3.3a and 3.3b for the two patients. Of the exonic variants, 13 exonic variants in Family 1 and 22 exonic variants in Family 2 were related to brain function or previously implicated in schizophrenia. Specifically, a number of *de novo* exonic variants in Family 1 are noteworthy (Table 3.3a). They include two chromosome 22 variants in *NEFH* that are not present in either parent or the unaffected co-twin. These changes however represent synonymous and in-frame changes and may not be pathogenic. In addition, this patient carries four exonic *de novo* variants in *ZNF595*, a zinc finger protein, one of which is a frameshift and the other which is a missense mutation at 4p16.3. Further, two missense mutations in *MUC3A* (that encode for a transmembrane mucin) and a missense in *OR4C45* (that encodes for olfactory receptors) may contribute to pathology. Interestingly, these variations have not been reported in 1000 genomes as well as CG public genomes and may not be polymorphic. In comparison, the patient from Family 2 appears to carry a much larger number of variants. However, in the absence of

parental sequences this only represents the differences between the discordant twins. Not surprisingly, most (51/82) of these variants in this patient have been reported in 1000 genomes and/or CG public genomes and are considered polymorphic. It leaves 31 variants that may be viewed as *de novo*. Further assessment of the impact of these changes has revealed that three of these represent frameshifts (140725160 on 5q31.3; 1018090 and 1018095 on 11q15.5) and nine represent missense mutations. The genes affected by missense changes include some that are compatible with the pathophysiology of schizophrenia. These include *NOTCH2, OR2T2, ZNF717, MUC4, MUC6, OR4C3, UBTFL1, TAS2R46* and *KRT6B*. Additionally, this patient also carried frameshift mutations that affected *PCDHGA, MUC6* and *OR4C45* (Table 3.3b).

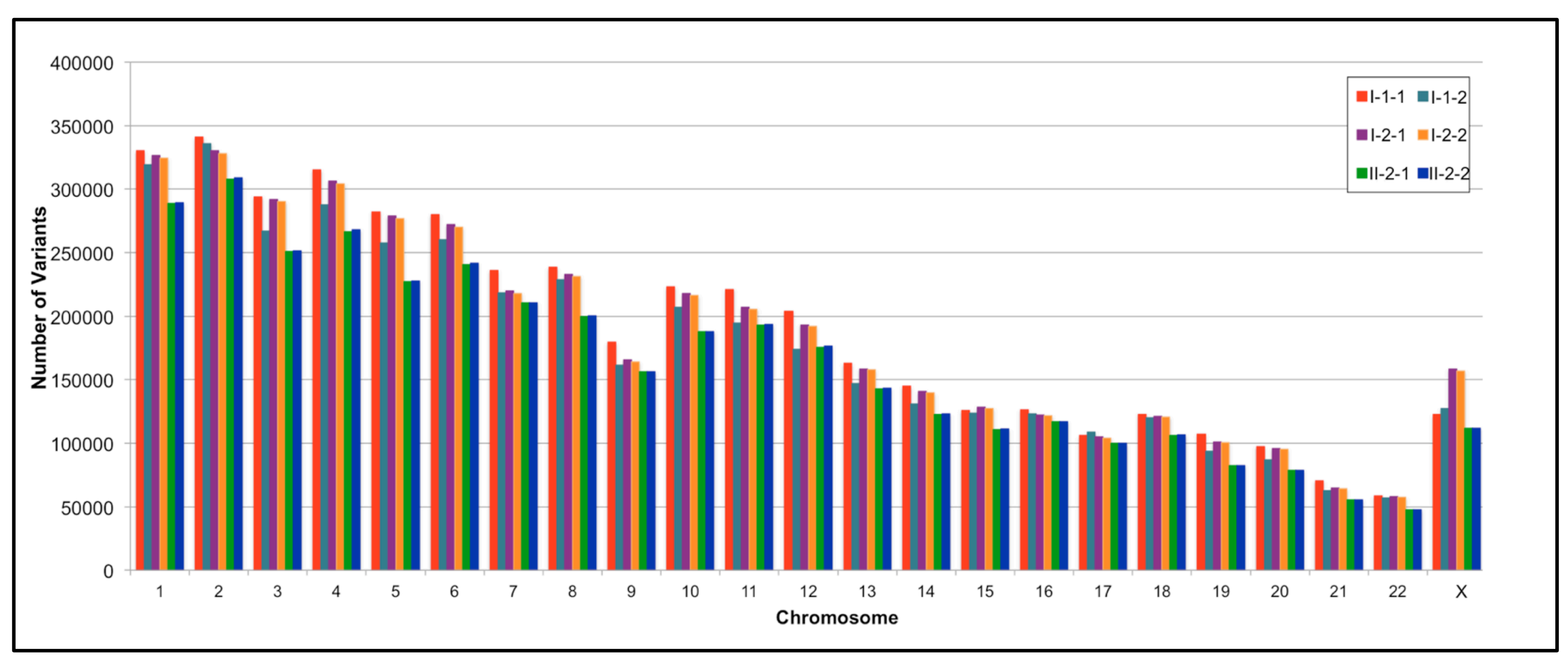


Figure 3.6. Chromosomal distribution of small sequence changes (SNVs, Indels, Block Substitutions) across 6 samples.

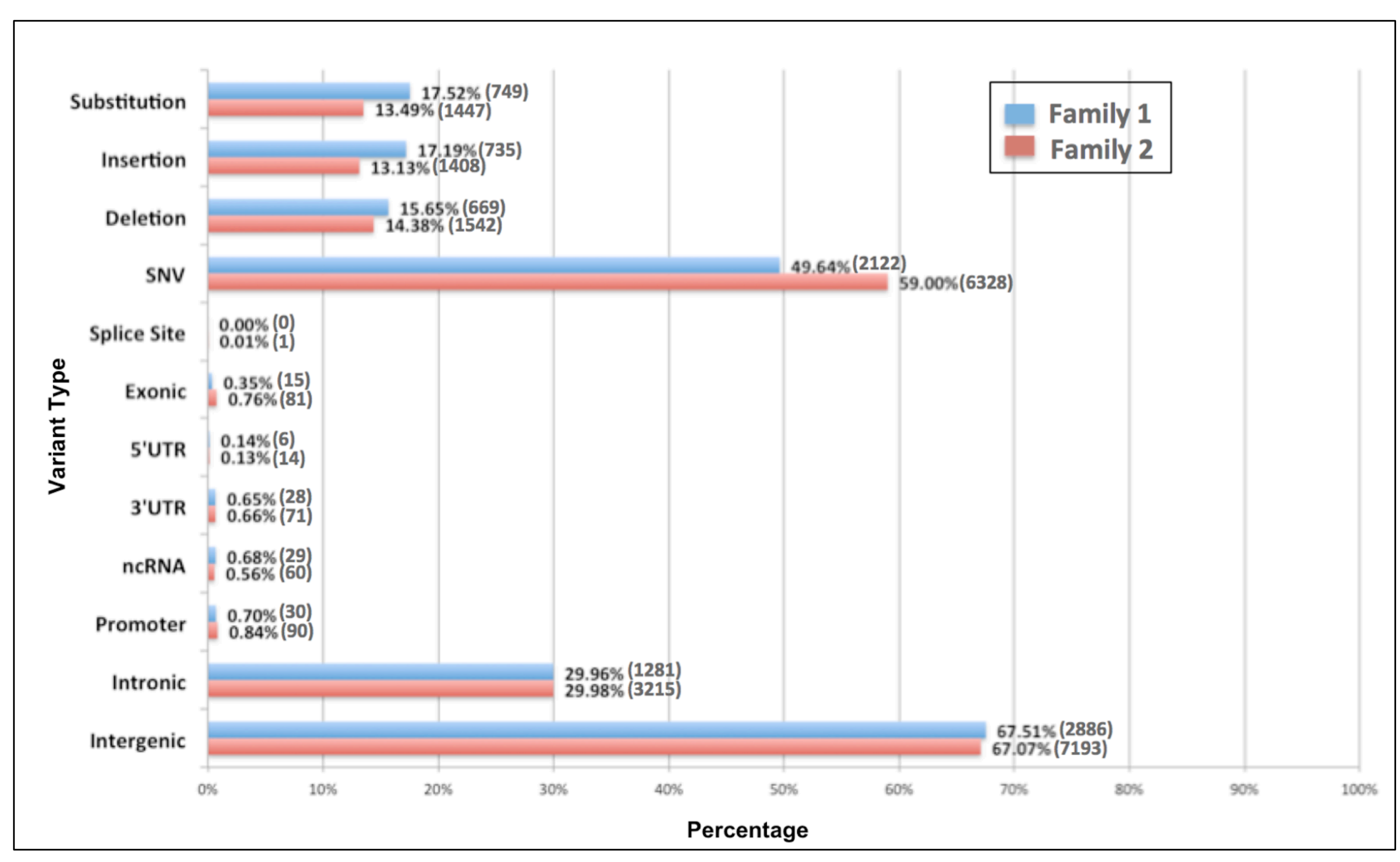


Figure 3.7. Identity of *de novo* (Family 1) and *provisional de novo* (Family 2) Variants Identified in the Affected Twin of Family 1 (I-2-1) and the Affected Twin of Family 2 (II-2-1) (not found in co-twin).

(ii) Copy Number Variants (CNVs)

The complete genome sequences of the six individuals were used to identify individual specific CNVs. Table 3.4 shows the distribution of observed CNVs in MZ twins representing the two families. The analysis allowed classification of CNVs representing inherited (present in at least one parent) and *de novo* (not present in either parent and also not present in co-twin) for family 1 and unshared between twins in family 2. It shows that monozygotic twins shared ~95% of called CNVs. Of these, ~90% of them were inherited from the parent and ~10% were not seen in either parent. Further, of the rare CNVs that were not shared between twins, some represented inherited events and others represented non-inherited and *de novo* events that must have occurred during independent mitosis (Table 3.5a and 3.5b). In Family 1, five of the seven unique CNVs were also not found in either parent (*de novo).* Discordant CNVs between twin pairs were annotated and are presented in Table 3.5a and 3.5b, for the affected twin from family 1 and 2, respectively. Three genes overlapped unique CNVs in Family 1. The *PDSS1*, and *LOC642131* genes represented *de novo* events, as they were not present in either parent while the CNV that overlapped *CES1* was present in one parent and was viewed as inherited. Although the affected twin of family 2 also carried 5 CNVs that were not present in her unaffected twin, these regions did not directly overlap any genes.

Table 3.5. Copy Number Variants Unique to Affected Twins.

Table 3.5a. Copy Number Variants unique to the affected member of Family 1 (I-2-1). The parental sequences have been used to establish the *de novo* (absent in parent(s)) or *inherited* (present in parent(s)) nature of each CNV. Shared identity is based on a 50% reciprocal overlap rule. (NCBI Build 37/hg19)

Chr	Start	End	Cytoband	Size (bp)	CNV Type	Identity	Overlapping genes
5	119380128	119382128	5q23.1	2000	Del	*de novo*	None
9	6700000	6710000	9p24.1	10000	Del	*de novo*	None
10	26998675	27002675	10p12.1	4000	Del	*de novo*	PDSS1
14	20314000	20328000	14q11.2	14000	Amp	Inherited	None
15	22422114	22492114	15q11.2	70000	Amp	*de novo*	LOC642131
16	34467150	34515150	16p11.2	48000	Amp	*de novo*	None
16	55841801	55855801	16q12.2	14000	Amp	Inherited	CES1

Table 3.5b. Copy Number Variants unique to the affected member of Family 2 (II-2-1) as compared to unaffected MZ twin. Shared identity is based on a 50% reciprocal overlap rule. (NCBI Build 37/hg19)

Chr	Start	End	Cytoband	Size (bp)	CNV Type	Overlapping genes
5	17612657	17620657	5p15.1	8000	Del	None
5	46244657	46246657	5p11	2000	Amp	None
5	135114128	135120128	5q31.1	6000	Del	None
12	114521470	114529470	12q24.21	8000	Amp	None
14	106926000	106930000	14q32.33	4000	Amp	None

(iii) Structural Variants (SVs)

The complete genome sequences of the six individuals were also used to identify individual specific deletions, tandem duplications, distal duplications, inter-chromosomal variations and inversions in comparison to NCBI Build 37/hg19. They were grouped as structural variants. In this analysis, the SVs ranged from 50 bp to 218 Mb in length. Table 3.6 shows that ~85% of the SVs were shared between co-twins in family 1 as well as in family 2. Also, most (>96%) of these were present in parent(s) and a small fraction (<4%) were not present in any parent. They were considered familial and *de novo,* respectively. Interestingly, 10-15% of SVs were not shared between twins, making a significant proportion *de novo* variations that originated after the separation of the twins during their independent development. Most of the SVs in family 1 (total 138) and 2 (total 141) represented deletions (78% and 86%), followed by tandem duplications (6% and 10%) and distal duplications (6.5% and 3.5%). Also, I found one inter-chromosomal move involving a translocation of *CDC27* from chromosome 2 to 17 to be unique to the patient in family 1.

Table 3.6. Shared and unshared Structural Variation (SV) identified in blood as compared to NCBI Build 37/hg19.

Sample	Total SVs	Shared with Co-twin	Unshared with Co-twin	Shared with Co-twin		Unshared with Co-twin	
				Inherited	*De novo*	Inherited	*De novo*
Family 1 Affected (I-2-1)	919	781 (85.0%)	138 (15.0%)	750 (96.0%)	31 (4.0%)	97 (70.3%)	41 (29.7%)
Family 1 Unaffected (I-2-2)	893	781 (87.5%)	112 (12.5%)	750 (96.0%)	31 (4.0%)	66 (58.9%)	46 (41.1%)
Family 2 Affected (II-2-1)	996	855 (85.8%)	141 (14.2%)	N/A	N/A	N/A	N/A
Family 2 Unaffected (II-2-2)	977	855 (87.5%)	122 (12.5%)	N/A	N/A	N/A	N/A

Note: Structural variants fell under five categories - deletions, tandem duplications, distal duplications, inter-chromosomal variations and inversions. There were no inversion differences found between twins in either family and therefore were not included in subsequent analyses.

Further, I characterized the genomic details of individual SVs that were unique to the affected twins of the two families (Appendix C and D). For family 1 where parental sequences are available I was able to eliminate familial events and concentrate on *de novo* events. I note that these changes have affected two genes (*ANKS1B* and *CLCN5*) via deletion, four genes (*LOC285768, NTM, SNORD115-29* and *GZMM*) due to tandem repeats and a 5 Mb tandem duplication on chromosome 1 (825765) affecting 101 genes. Similarly, I identified SV differences between twins in family 2, but in the absence of parental sequences I could not characterize them further. Consequently, the SV differences between twins in family 2 are viewed as *provisional de novo*. Similarly to family 1, they affected 47 genes by deletion, 7 genes at 7 sites by tandem duplication along with a 3 Mb tandem duplication on X (52886720) overlapping 39 genes and 2 genes by distal duplication.

Also, the genomes carried rare (two or three) inversions, both twin pairs shared them with their co-twin and all inversions were inherited in Family 1. Finally, the sharing of CNVs and SVs between any two individuals is directly correlated with their genetic relatedness; high but not 100% between monozygotic twins, low across unrelated individuals and intermediate between a parent and an offspring, as expected.

3.3.1.5 Independent confirmation of NGS results

I sought to confirm a subset of CNVs, SVs and SSCs by independent experiments involving Sanger Sequencing (Appendix E), PCR Amplification of Specific Alleles (PASA) (Table 3.7) (Figure 3.8a and 3.8b) and Real Time PCR (Table 3.8). They represented 4 CNVs, 6 SVs and 24 SSCs. Further, the RealTime PCR for CNVs and SVs using blood DNA confirmed 5/10 variants, three of which were also confirmed on the DNA from buccal swabs from the same individuals. The use of buccal swabs allows for assessment of the timing of *de novo* mutation. Because blood and buccal cells arise from two different germ layers, namely, the mesoderm and ectoderm, variants that are found in both tissues would be expected to have arisen very early in development, before the separation of the germ layers. Similarly, the results of PASA followed by Sanger Sequencing confirmed 2 SSCs identified by NGS (Appendix E).

Table 3.7. Regions analyzed using PCR Amplification of Specific Alleles (PASA).

Gene	Original Forward	Original Reverse	Allele Specific Forward	PASA Result
ZNF595	5'TCGTTTTACATGT CACACCTAACT3'	5'TTATGTTCGTTC AGAACTGTGG3'	5'CAGCAGGAATAAC AGGTACAAATA*3'	Confirmed, see Figure 3.8a
PDE4DIP	5'CCTGAGGAGTAT GGGGTAATCA3'	5'TGCCTCCACTT CTTTGTTCC3'	5'CTACCTGTCAAAA CCTCCAGTA*3'	Confirmed, see Figure 3.8b
PLIN4	5'ATTTACGGCACC AGTGACTC3'	5'GACCCAAAATA TCGCAACAG3'	5'ACGCCGGTCTGGA CAGTCCCTG*3'	Was not able to amplify correct product
QRICH2	5'TACCTTGCTGAC CTATTCCAG3'	5'AGATTGATGTG GTGCAACCT3'	5'CACGCTGATCCAC TCCC*3'	Heterozygote in both twins
MUC19	5'TAGAACATCGGT TGAAGAATCA3'	5'AGCTGATGGCC GAATTGT3'	5'CAGCAGGAATAAC AGGTACAAATA*3'	Heterozygote in both twins

Note: The asterisk indicates the allele specific nucleotide.

Figure 3.8. Presence of Mosaicism in Sanger Sequencing Results (PASA).

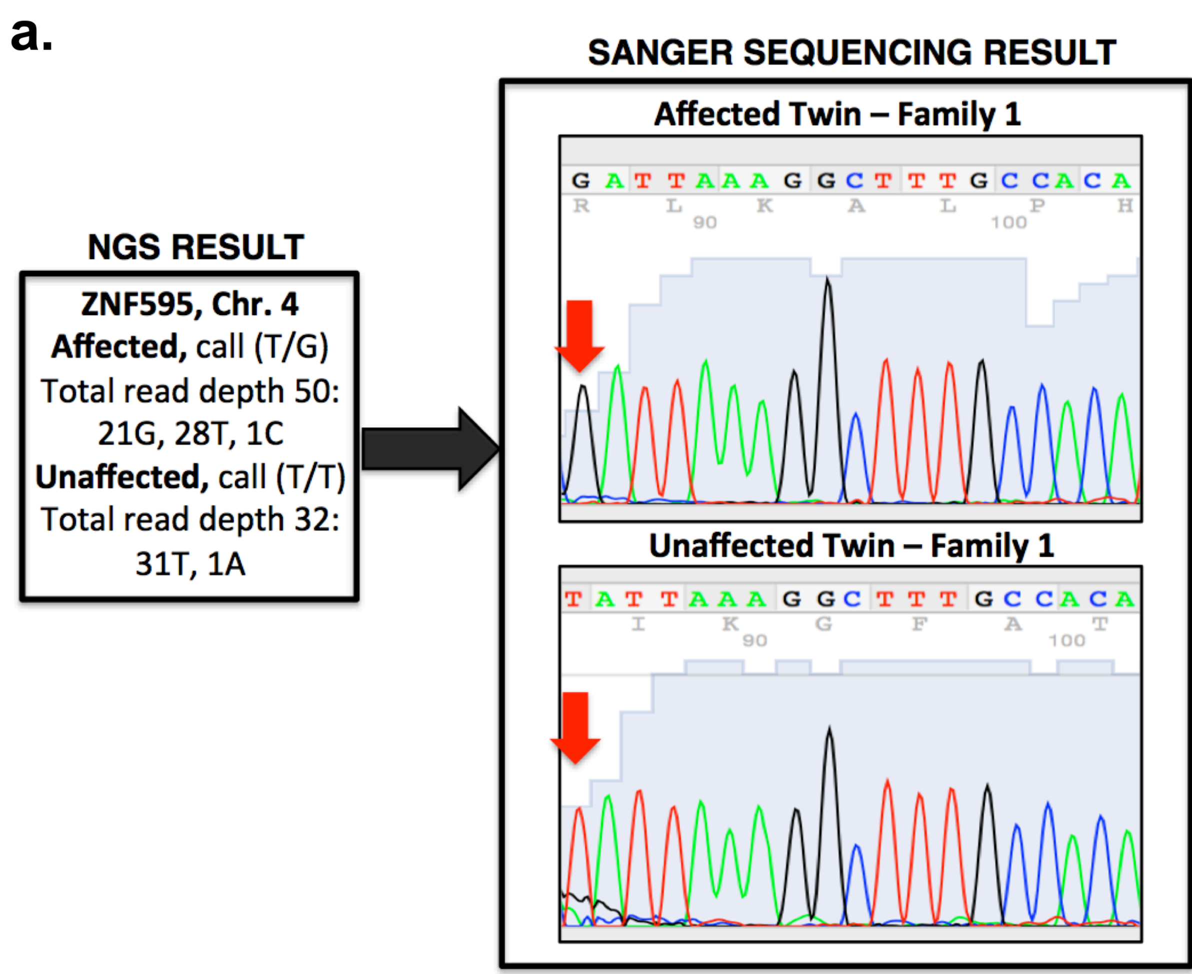


Figure 3.8a. Confirmed mosaic variant between monozygotic twins in Family 1 in the 4th exon of the ZNF595 gene. This variant changes isoleucine to serine. The mother and father both have a T at this location (ATT - Isoleucine, I) in their complete genomics results.

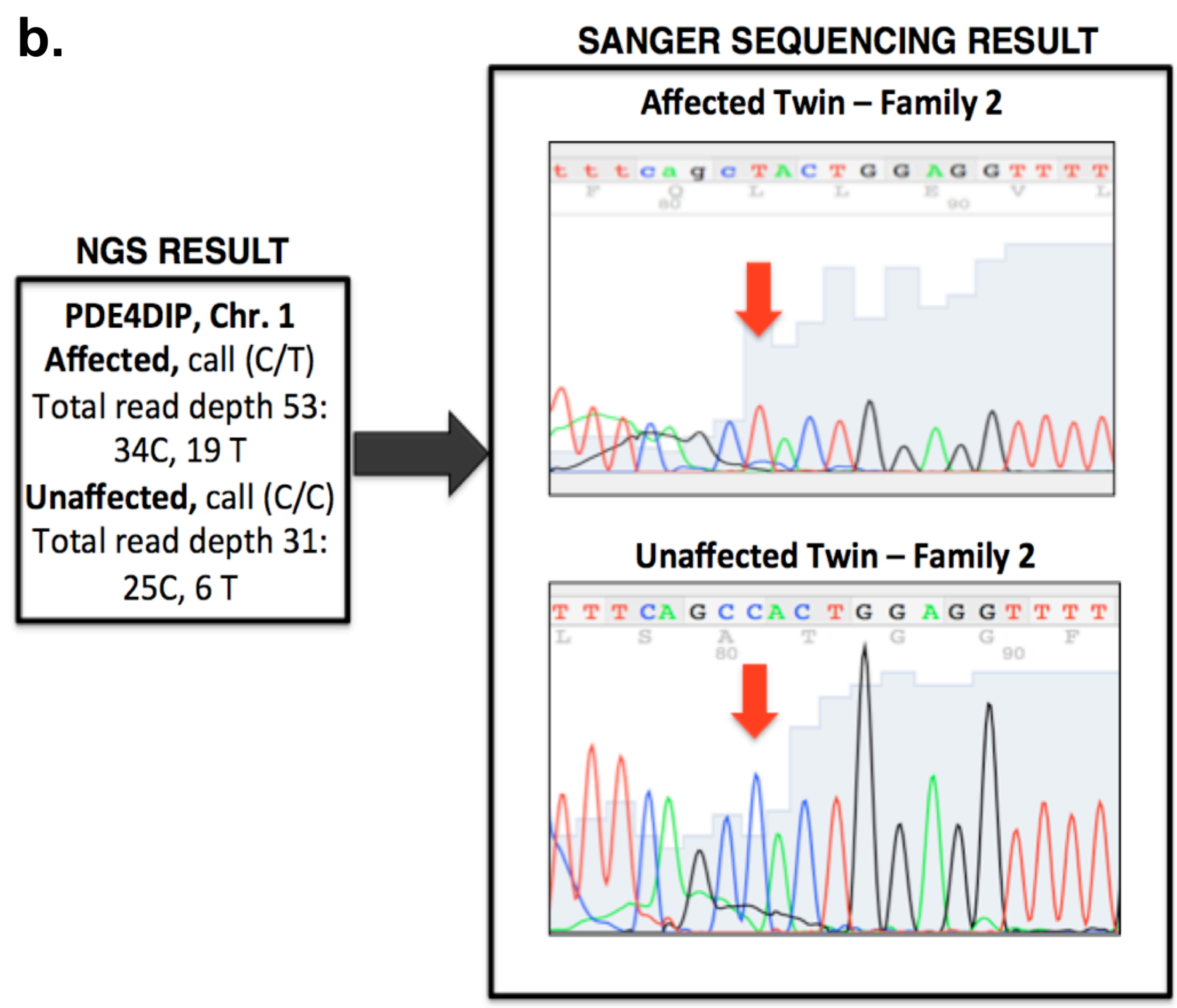


Figure 3.8b. Confirmed mosaic variant between monozygotic twins in Family 2 in the 13th exon of the PDE4DIP gene. This variant changes Tryptophan to a premature stop codon. There is no information available for the mother and father at this location.

Table 3.8. Regions predicted by NGS to be unshared between twins and selected for assessment by Real Time Quantitative PCR Analysis. (NCBI Build 37/hg19)

Gene/ Region	LT Product Number	Chr	Start	End	Size (bp)	Family	Type	Prediction	Result in Blood	Result in Buccal
GABRD	Hs01969851	1	825765	5726936	4901171	1	SV	Increase	2.62*	2.65*
CLSTN2	Hs04743760	3	139670597	139674454	3857	2	SV	Increase	1.16*	1.64
OPRM1	Hs06783090	6	154446930	154447253	323	2	SV	Decrease	2.15	N/A
STX1A	Hs01848137	7	66193913	73382120	7188207	1	SV	Decrease	2.75*	1.98
12q24.21	Hs06365560	12	114521470	114529470	8000	2	CNV	Increase	1.33*	0.82*
ANKS1B	Hs03835432	12	99793948	99802771	8823	1	SV	Decrease	1.9^	N/A
15q11.2	Hs03896607	15	22422114	22492114	70000	1	CNV	Increase	2.69*	2.58*
CES1	Hs00139541	16	55841801	55855801	14000	1	CNV	Increase	1.56	N/A
16p11.2	Hs03313918	16	34467150	34515150	48000	1	CNV	Increase	1.8	N/A
PLTP	Hs07203816	20	44535505	44535941	436	1	SV	Decrease	2.1	N/A

Note: LT=Life Technologies, ^=Correct direction. Prediction is based on sequencing result. Regions were first assessed on blood DNA; regions that confirmed in blood DNA were then assessed in buccal DNA. The asterisk refers to a significant difference. Significance was defined as a Z-Score <1.75 and 95% confidence level.

3.3.2 Genetic differences between MZ twins and their discordance for schizophrenia

The results also allowed for assessment of any involvement of genetic differences between MZ twins in the development of schizophrenia using a threshold model (McGue *et al*., 1983). It assumes that both members of the monozygotic discordant twins will carry family specific genetic predisposition for the disease and additional changes in the affected twin will swing the predisposition to a threshold that will be sufficient for the onset of the disease. This model is testable given near complete coverage of genetic changes in the affected as well as unaffected members of the two MZD twin pairs. It begins with development of a list of all genes that are affected by a variety of mutational mechanisms in four members of the two MZD twin pairs. This list is rather comprehensive and individual specific. I have used this list to assess the effect of these differences using Ingenuity Pathway Analysis (IPA). It moves the analysis from a focus on affected genes to affected pathways.

Table 3.9 shows the top 20 individual specific canonical pathways in two affected individuals belonging to the two families. Figure 3.9 outlines my strategy in assessing the nature of the threshold model in the two patients. Of the top 20 pathways in patient 1 and patient 2, 10 are shared between them (Figure 3.10). Also, 8 of the 10 pathways shared by the two patients are present in either of the two unaffected members of the two twin pairs. This results in 2 pathways that may be viewed as highly specific to the two patients (Figure 3.9). They represent Dopamine-DARPP32 Feedback in cAMP Signaling (Figure 3.11a, Family 1, p=2.28E-04; Family 2, p=5.84E-03) and Glutamate Receptor Signaling (Figure 3.11b, Family 1, p=1.46E-03; Family 2, p=3.88E-05). I note that 13 and 29 genes of the two pathways respectively are affected in the patient of family 1 (Table 3.10). The corresponding numbers for the patient in family 2 are 15 and 23 genes for the two pathways, respectively (Table 3.10). Although few of the mutated genes are patient specific, the two patients share a large number of affected molecules in Glutamate Receptor Signaling as well as Dopamine-DARPP32

Feedback in cAMP Signaling pathways. Further, I note that four of the top five pathways in the patient of family 1 and three of the top five canonical pathways in the patient of family 2 are relevant to neural functions and have the potential to contribute to the disease in a threshold model discussed in the next section (Table 3.9).

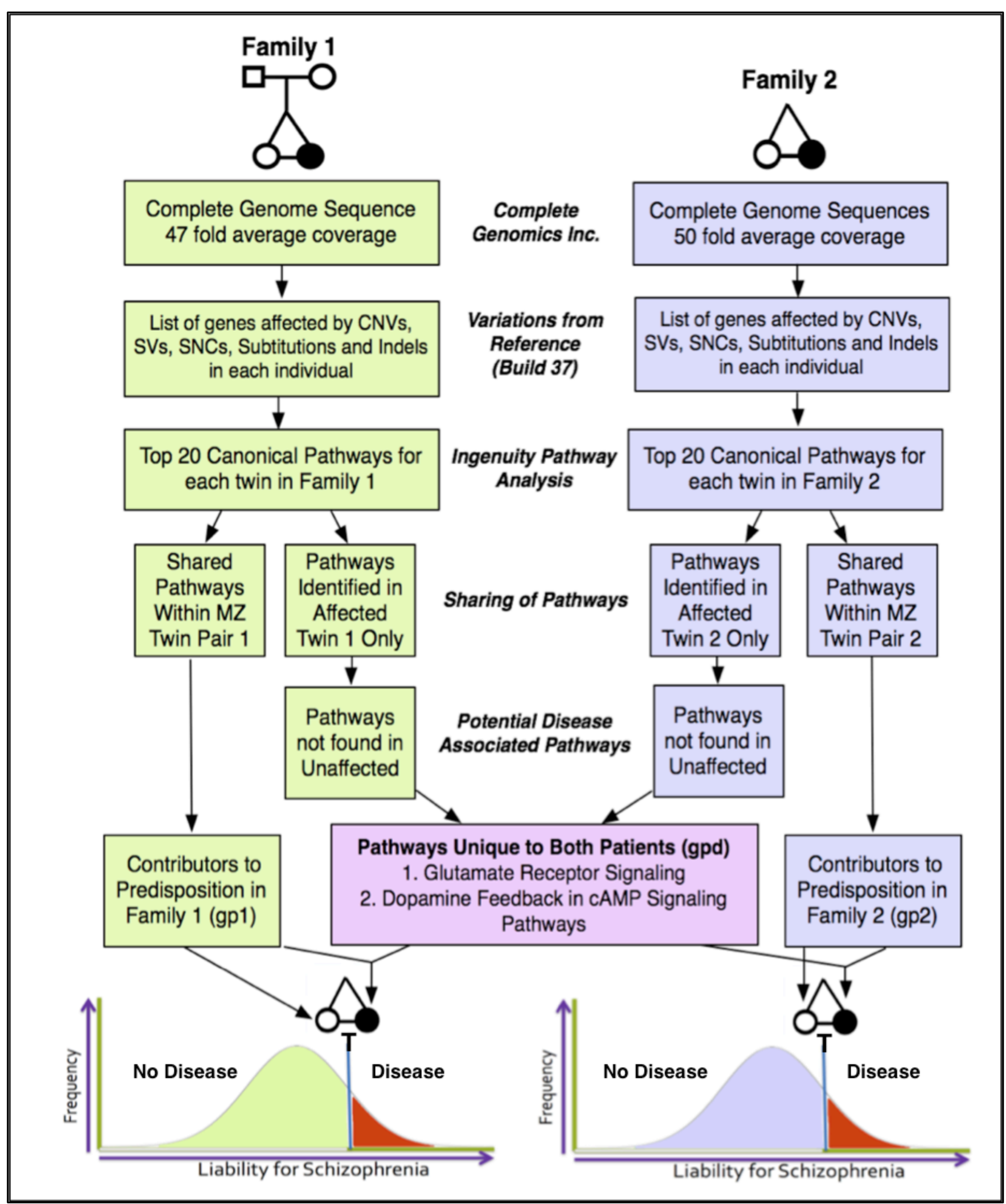


Figure 3.9. Approach for analysis of Ingenuity Pathway Results.

Blue line represents disease threshold (T).

Table 3.9. Top 20 Canonical Pathways identified by Ingenuity Pathway Analysis (IPA) in the affected twin of each family.

Canonical Pathways - Family 1 Affected (1A)	Shared With	*p*-value
1. CREB Signaling in Neurons	2A, 2U	0.0000041687
2. Neuropathic Pain Signaling In Dorsal Horn Neurons	2A, 1U, 2U	0.0000083176
3. Axonal Guidance Signaling	2A, 1U	0.0000630957
4. Cellular Effects of Sildenafil (Viagra)	2A, 1U	0.0001479108
5. Role of NFAT in Cardiac Hypertrophy	-	0.0001513561
6. Dopamine-DARPP32 Feedback in cAMP Signaling	2A	0.0002290868
7. Synaptic Long Term Depression	2A, 1U, 2U	0.0002398833
8. Wnt/Ca+ pathway	-	0.0003630781
9. Synaptic Long Term Potentiation	2A, 2U	0.0008511380
10. PPARÎ±/RXRÎ± Activation	2U	0.0011748976
11. Gap Junction Signaling	-	0.0013803843
12. Glutamate Receptor Signaling	2A	0.0014791084
13. 14-3-3-mediated Signaling	-	0.0020417379
14. Netrin Signaling	2A, 1U, 2U	0.0022908677
15. Leptin Signaling in Obesity	-	0.0024547089
16. Nitric Oxide Signaling in the Cardiovascular System	2A, 1U	0.0024547089
17. Hepatic Cholestasis	2U	0.0026915348
18. Uracil Degradation II (Reductive)	2U	0.0029512092
19. Thymine Degradation	2U	0.0029512092
20. Melatonin Signaling	1U	0.0033113112
Canonical Pathways - Family 2 Affected (2A)	**Shared With**	***p*-value**
1. Sperm Motility	1U	0.0000053703
2. Glutamate Receptor Signaling	1A	0.0000389045
3. Neuropathic Pain Signaling In Dorsal Horn Neurons	1A, 2U, 1U	0.0000575440
4. Cellular Effects of Sildenafil (Viagra)	1A, 1U	0.0002238721
5. Nitric Oxide Signaling in the Cardiovascular System	1A, 1U	0.0002884032
6. Synaptic Long Term Depression	1A, 2U, 1U	0.0003235937
7. CREB Signaling in Neurons	1A, 2U	0.0003630781
8. Synaptic Long Term Potentiation	1A, 2U	0.0005888437
9. Phospholipase C Signaling	-	0.0008709636
10. Netrin Signaling	1A, 2U, 1U	0.0010964782
11. G-Protein Coupled Receptor Signaling	2U	0.0012589254
12. Î±-Adrenergic Signaling	-	0.0016595869
13. Protein Kinase A Signaling	2U, 1U	0.0023442288
14. nNOS Signaling in Neurons	-	0.0040738028

15. Huntington's Disease Signaling	-	0.0042657952
16. Dopamine-DARPP32 Feedback in cAMP Signaling	1A	0.0058884366
17. Cardiac Î²-adrenergic Signaling	2U	0.0067608298
18. Calcium Signaling	2U	0.0077624712
19. Breast Cancer Regulation by Stathmin1	-	0.0087096359
20. Axonal Guidance Signaling	1A, 1U	0.0089125094

Note: 1A=Family 1- Affected, 1U=Family 1- Unaffected, 2A=Family 2- Affected, 2U=Family 2- Unaffected.

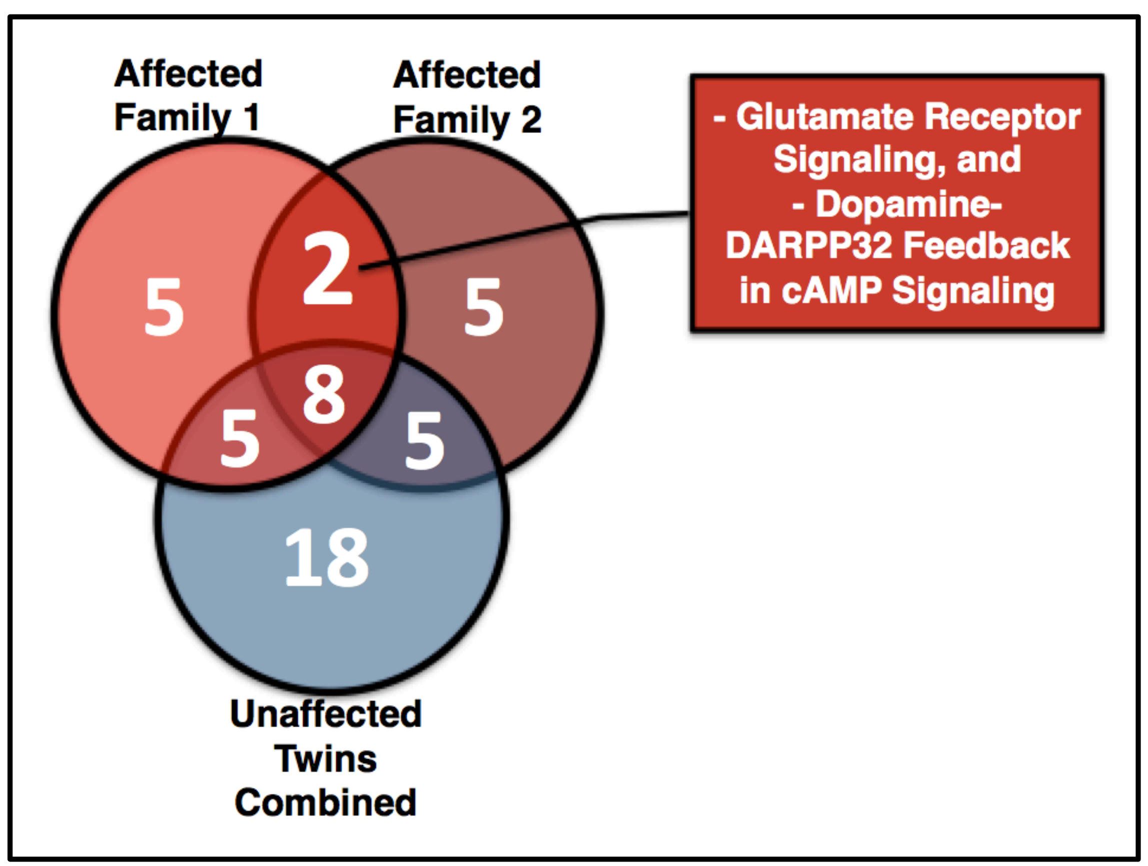
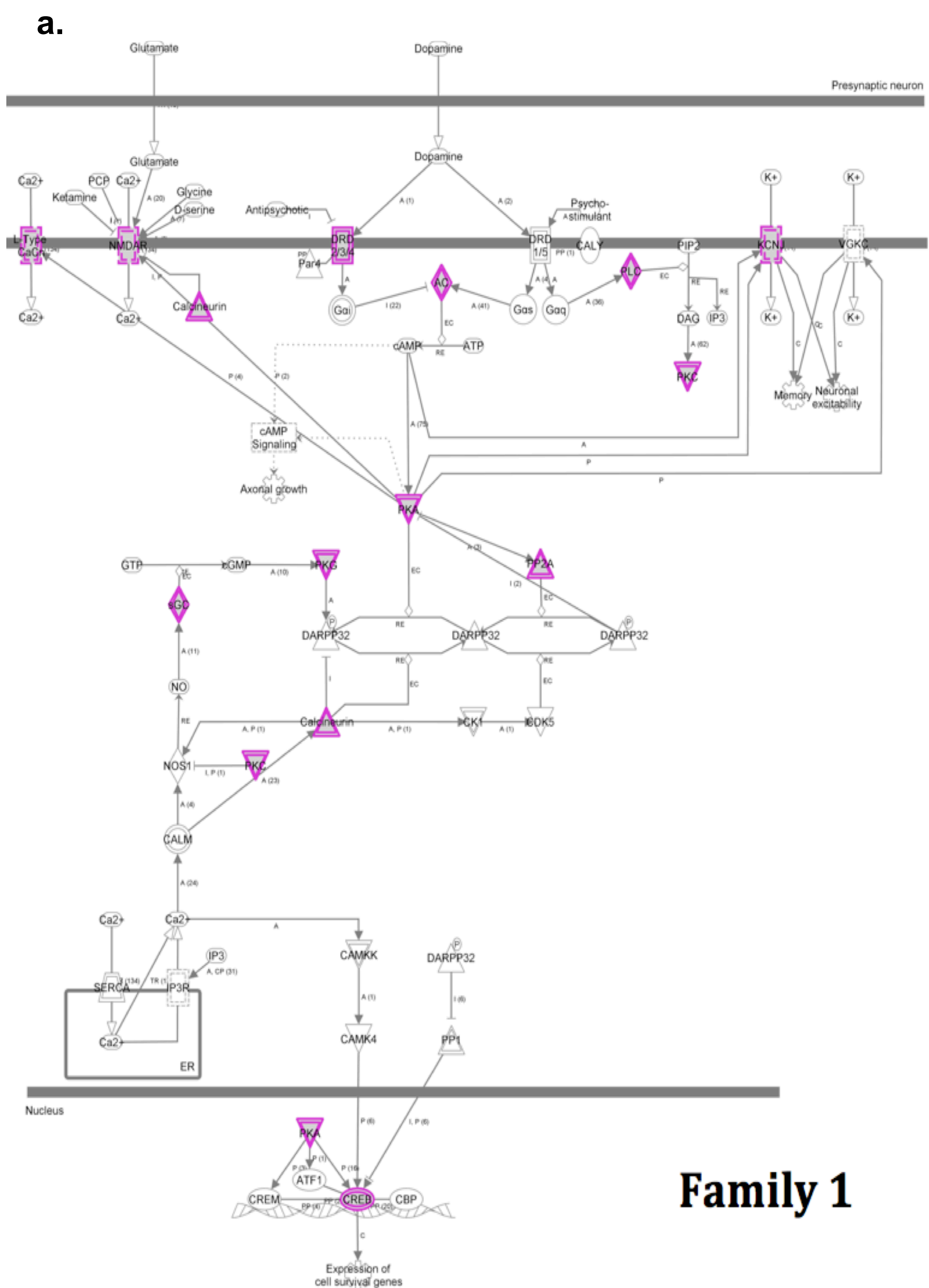


Figure 3.10. Venn diagram of the top 20 canonical pathways found to be overrepresented by variants in the sequencing dataset of affected and unaffected twins.

Note: The pathways from the unaffected twin in Family 1 and Family 2 have been merged to represent 36 unique pathways (4 shared). The 18 pathways unique to unaffected twins are broken down into 11 from Family 1 and 7 from Family 2.

Figure 3.11. Ingenuity Pathways Analysis Identified Pathways in Affected Twins Only. Legend for symbols available in Appendix L.

a.



a. continued

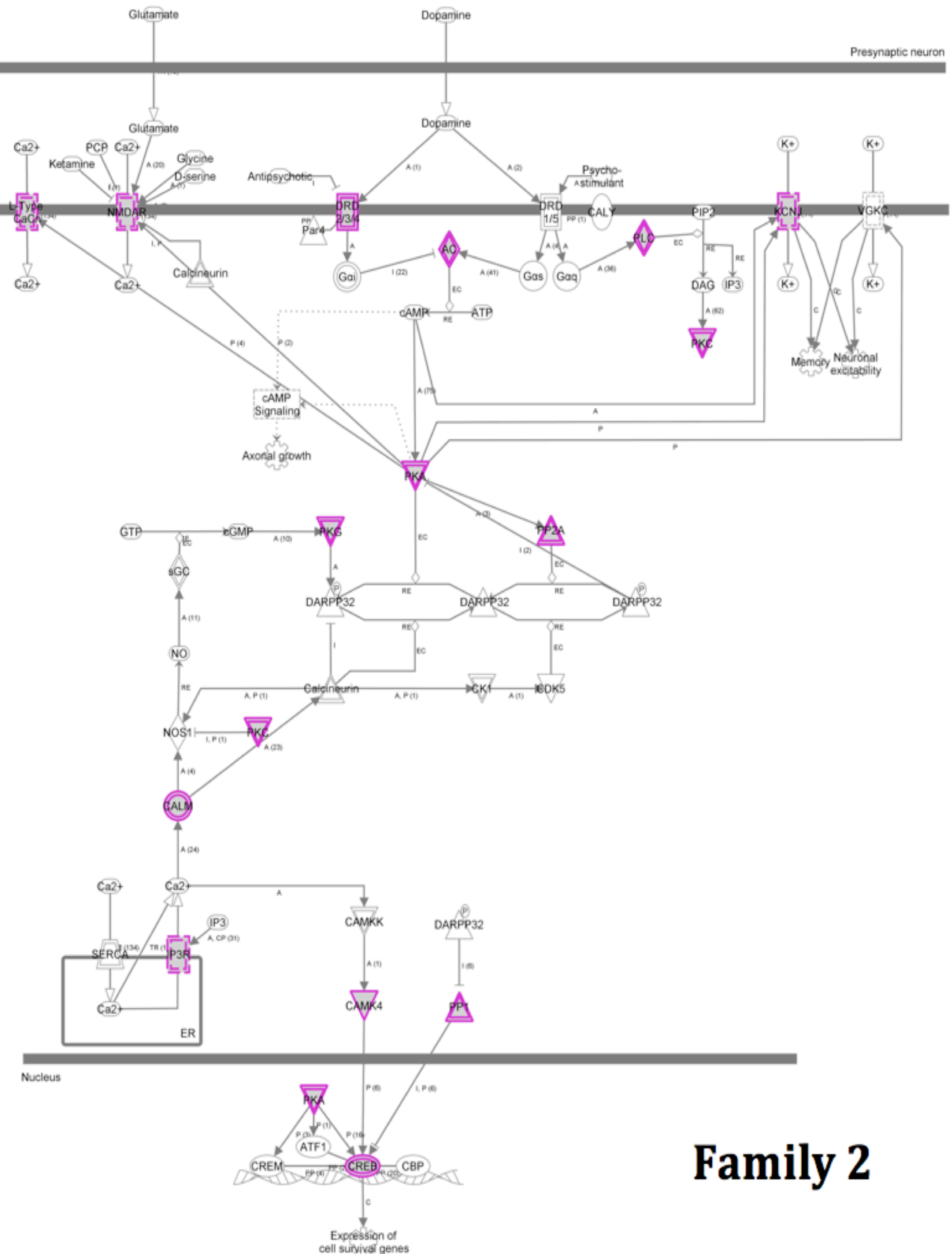
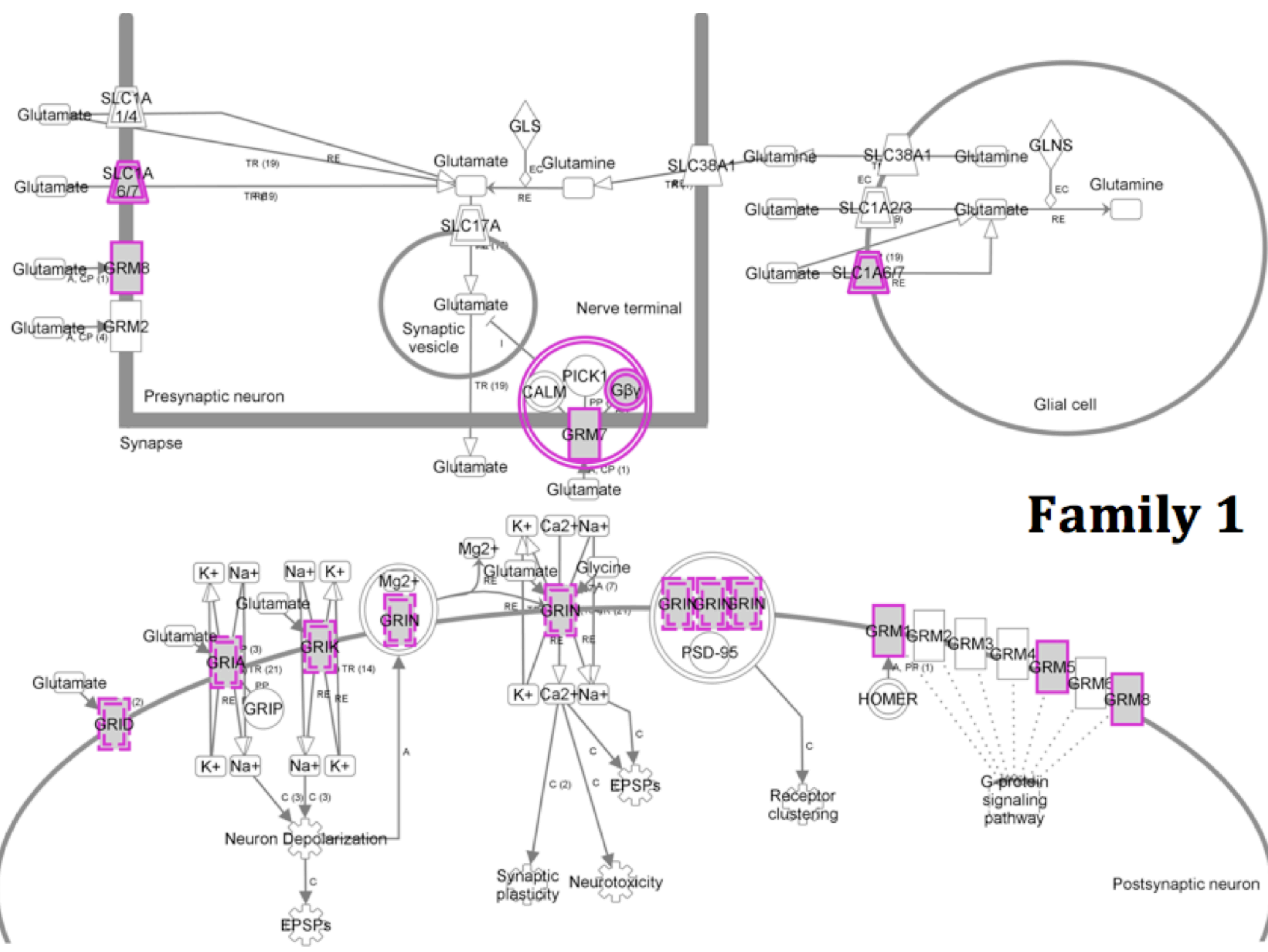


Figure 3.11a. Dopamine Feedback in cAMP Signaling pathway identified using Ingenuity Pathway Analysis (IPA). This pathway emerged independently in both affected twins. This pathway was also not found to be enriched in unaffected twins in the study. Purple represents genes harboring a unique high confidence variant (Family 1, p=2.28E-04; Family 2, p=5.84E-03).

b.



b. continued

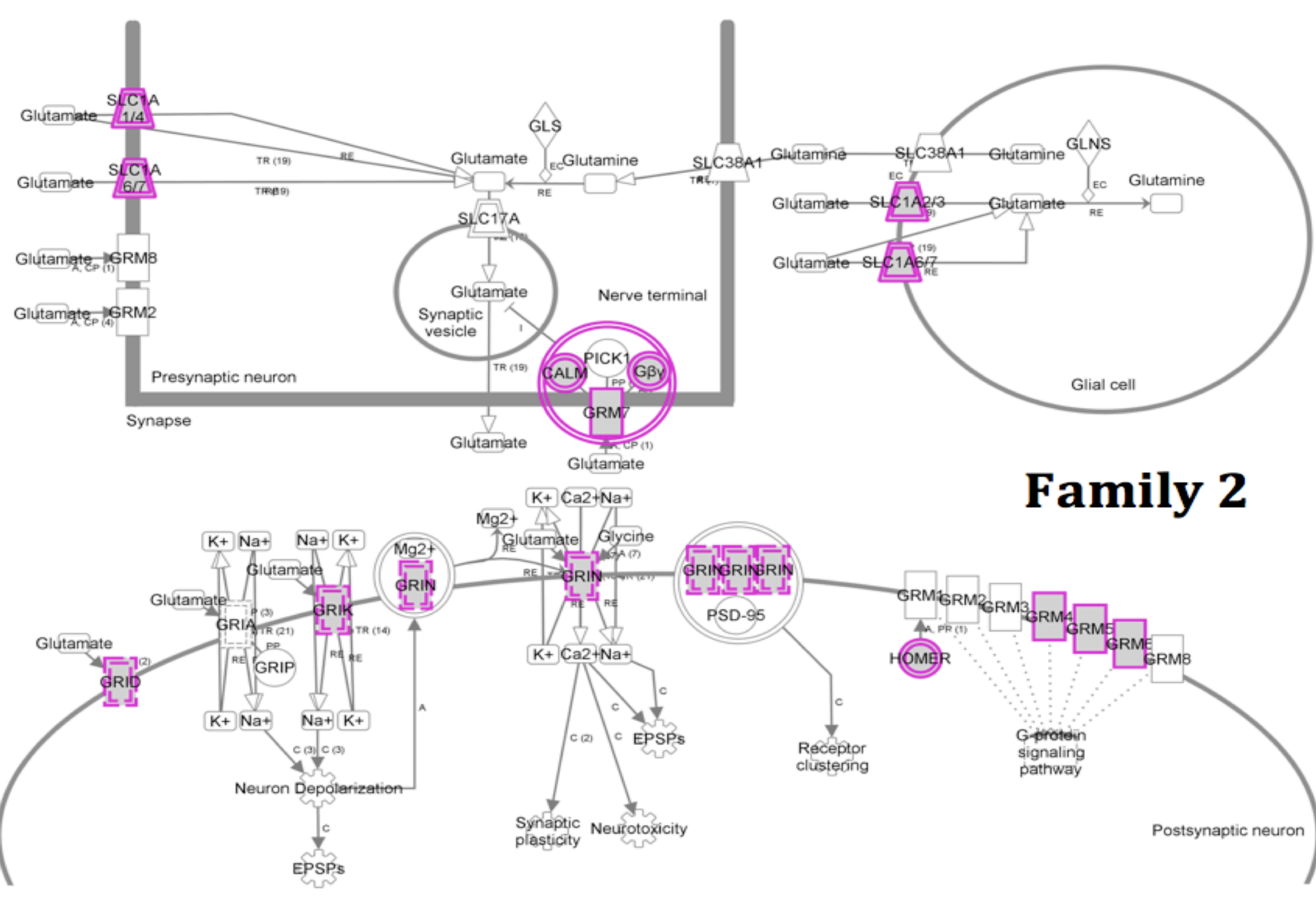


Figure 3.11b. Glutamate Receptor Signaling pathway identified using Ingenuity Pathway Analysis (IPA). This pathway emerged independently in both affected twins. This pathway was also not found to be enriched in unaffected twins in the study. Purple represents genes harboring a unique high confidence variant (Family 1, p=1.46E-03; Family 2, p=3.88E-05).

Table 3.10. Genes affected in two pathways (Glutamate Receptor Signaling and Dopamine Feedback in cAMP Signaling) unique to the two unrelated schizophrenia patients representing family 1 and family 2.

Pathway	Family 1: Genes in Pathway		Family 2: Genes in Pathway		Common to both Affected Twins
	Number of genes affected	Identity	Number of genes affected	Identity	
Dopamine Feedback in cAMP Signaling	29	PRKG2, PPP2R1U, GUCY1A3, PRKAG2, ADCY3, PRKCB, PRKAR1B, KCNJ12, ADCY2, PRKCZ, ADCY8, PLCL1, CACNA1D, DRD3, PLCB1, GUCY1A2, EP300, PRKG1, PPP3CA, CACNA1A, GRIN2U, PLCG2, KCNJ3, PRKCA, PLCH2, CACNA1C, PLCL2, PLCZ1, KCNJ10	23	PLCB1, PRKG2, PRKG1, PPM1L, CAMK4, PLCE1, PRKAG2, PPP2R2C, ADCY10, PPP1R14C, GRIN2A, CREB5, PRKCA, CACNA1C, PRKCB, KCNJ12, ITPR1, PRKAR1B, PRKCI, PRKCZ, ADCY8, PRKCG, DRD3	**12** (PRKG2, PRKAG2, PRKCB, PRKAR1B, KCNJ12, PRKCZ, ADCY8, DRD3, PLCB1, PRKG1, PRKCA, CACNA1C)
Glutamate Receptor Signaling	13	GRIA1, GRIK2, GRID1, GNB1, GRIN2U, SLC1A7, GRIK1, GRIK4, GRM7, GRM1, GRM5, GRM8, GRIK3	15	GRID1, GNG7, CAMK4, SLC1A2, HOMER1, SLC1A7, GRIK1, GRIN2A, GRM4, GRIK4, SLC1A1, GRM7, GNG2, GRM5, GRM6	**6** (GRID1, SLC1A7, GRIK1, GRIK4, GRM7, GRM5)

3.4 Discussion

The search for genetic causes of disease, that has accelerated with our ability to genotype individuals with and without the disease, has concluded that not all genetic diseases are caused by a single mutation in a single gene (Kong *et al.*, 2014). In fact, most complex diseases have heterogeneous causation and every individual is genetically unique. It makes experimental matching of patients and controls a difficult proposition and studies on ever increasing numbers of patients and controls do not always meet the objectives. With this in mind, I focused this research on the assessment of genome-wide mutations for two unrelated pairs of MZD twins. Using complete genome sequences, I identified individual specific genome-wide gene mutations (SSCs, CNVs and SVs) that are exceptionally comprehensive and reflect the highest resolution currently available for individual genomes. Caution must be taken in interpretation of the results due to the persistence of false positives found in large-scale sequencing studies. However, the results allow for the appraisal of the nature and source of genetic variation in individual genomes. Also, given the discordance of MZ twins for schizophrenia, the results are valuable in assessing the threshold model for the development of this disease, that has been previously suggested (Gottesman and Shields, 1967; McGue *et al.*, 1983) but remained untested. I will present a model to show how my data on mutations observed in individual genomes and their relationship involving MZ twins offers novel insight into the development of schizophrenia. First, I will discuss the observations on the individual genome sequences.

3.4.1 Individual human genomes contain extensive variability

In general, the identity and distribution of variations in this study follow the results on 1000 genomes data and expected distributions based on chromosome size (Figure 3.6 and 3.7). The data is accumulating in the literature that every individual harbors hundreds of *rare* variants (Abecasis *et al.*, 2012). Even in healthy individuals, approximately 300 to 500 missense mutations are identified

per genome, and it is estimated that each individual carries approximately 60 missense mutations that damage protein structure as well as approximately 100 loss of function variants (Xue *et al*., 2012). My results are consistent with 1000 genomes pilot data which found that individual genomes have on average, 3.6 million SNVs, 344 thousand indels and 717 large deletions (Abecasis *et al*., 2012). My results are also consistent with findings of *de novo* variation within families; current results in the field show that both somatic and germline *de novo* variation can be identified in offspring. However, the rate of *de novo* mutation appears to be variable and may be family specific (Project *et al*., 2011). Unfortunately, this variability is not always easy to discern. Studies of twin sequencing data in the literature have noted validation rates of variants as low as 1/15 and groups continue to confirm the high error rates of current next generation sequencing technologies (Dal *et al*., 2014). In addition, validation rates for data from the 1000 genomes project averaged a validation rate of 1.8%, and similar to my study, large SVs were easier to validate (2.1%) than other types of variation. My analysis also focused on the use of higher quality variant calls as was done in the 1000 genomes pilot study to evaluate the likelihood of a candidate variant call being a real event (Abecasis *et al*., 2012). The results show that individual genomes harbor extensive variability and that this variability can be measured both within and between generations.

3.4.2 MZ twins are genetically different

Results included in this report also show that the genome sequences of pairs of MZ twins although similar, are not identical. It follows a number of recent reports on identical twins (Kondo *et al*., 2002; Reumers *et al*., 2012; Vadlamudi *et al*., 2010; Weber-Lehmann *et al*., 2014; Ye *et al*., 2013). Although likely to be a conservative estimate, Krawczak *et al* has predicted that there exists at minimum an average of >1.3 SNVs discriminating MZ twins in each tissue type (Krawczak *et al*., 2012). The results of one forensic study found DNA sequence differences between identical twins that appear to reflect mosaicism in that the newly arisen

allele was generally only found in a small fraction (approximately 20% of the cells assayed) as estimated by Sanger Sequencing results (Weber-Lehmann *et al*., 2014). Many reports have now identified genetic and epigenetic differences between identical twins (Dempster *et al*., 2011, 2014; Ehli *et al*., 2012). For example, Bruder *et al* (2008) reported that all of 19 MZ twins studied differed in CNVs (Bruder *et al*., 2008) and Forsberg *et al* confirmed 10 post-twinning CNV mutations in 159 MZ twin pairs (Forsberg *et al*., 2012). Many post-twinning single nucleotide mutations have also been reported (Kondo *et al*., 2002; Reumers *et al*., 2012; Vadlamudi *et al*., 2010; Ye *et al*., 2013), however these are expected to be rarer than post-twinning CNVs. It is likely that older twin pairs will have accumulated more *de novo* mutations over time (Ye *et al*., 2013) and this will be expected to affect the degree of somatic mosaicism across tissues (Piotrowski *et al*., 2008). Additionally, my previous research on MZ twin pairs proposes that *de novo* copy number variants can be identified by comparison of the genomes of MZ twins with their parents (Maiti *et al*., 2011). Although the exact contribution of *de novo* differences to disease has not been identified for most diseases, one disease, Mayer-Rokitansky-Kuster-Hauser syndrome, has been found to arise from a patient-specific duplication in the affected monozygotic twin in a tissue specific manner leading to disease manifestation (Rall *et al*., 2015). In addition, a recent report that identified somatic mutations at the base-pair level in monozygotic twins found two *de novo* somatic mutations that appear to have occurred early in embryonic development, suggesting that early development may be enriched for *de novo* change (Li *et al*., 2014). Although one group concluded that there are no differences between the genomes of co-twins discordant for multiple sclerosis (Baranzini *et al*., 2010), it has subsequently been suggested that the analytical methods used, particularly the low average sequence coverage (22-fold), was not sufficiently powerful to reveal rare somatic mutations (Handunnetthi *et al*., 2010). It has been suggested that a fold coverage of 30x or greater is necessary to identify true somatic mutations (Bentley *et al*., 2008) in the genome, which this current study reflects.

This study, with complete genome sequences of pairs of MZD twins offers a number of advantages. First, it focuses on individual patients and matches the patient with her unaffected monozygotic twin; nature's near perfect genetic match. The consequence is that it drastically reduces genetic heterogeneity between a case and the matched control. Also, it focuses on the genetic factors of schizophrenia in an individual patient rather than grouping together heterogeneous cases of schizophrenia. Second, the availability of complete genome sequences allows for identification of almost all forms of genetic mutations per individual. Third, the identification of all affected genes in a patient allows for further identification of canonical pathways that may be affected by individual specific gene mutations. Finally, it assesses the sum combination of affected pathways that are specific to the patient(s) versus pathways that are also overrepresented in the unaffected twin.

The analysis of affected pathways began with identification of genes that were affected by any mutational mechanisms in each patient and not found in their co-twin. The genes that met this criteria were used in pathway analysis. This analysis yielded a set of pathways for four individuals representing two twin pairs. The top 20 pathways identified in four members of the two twin pairs are given in Table 3.9. Interestingly, a number of pathways were shared across all four individuals. Next, I focused on affected pathways that were shared or not shared in the two MZD twins. I then assessed the data to evaluate the threshold model of schizophrenia (Figure 3.9). Within twin pair sharing of pathways was viewed to represent "genetic predisposition" and were considered to not be sufficient to cause the disease; I labeled these as genetic-predisposition pathways (gp1 and gp2 for each family, respectively). The composition of the genetic-predisposing pathways was expected to differ between unrelated pairs.

Further, I selected pathways that were unique to the two patients in this study. In my model, these additional defects are needed to cross the liability threshold and manifest the disease. Any potential disease-associated pathways were labeled gpd. The composition of disease-associated pathways may vary

across families. In this analysis I identified two potential pathways in both families that could be contributing to disease causation (due to their exclusive presence in affected twins); Dopamine Feedback in cAMP Signaling Pathway (Figure 3.11a) and Glutamate Receptor Signaling (Figure 3.11b).

Interestingly, a number of pathways labeled as genetic predisposing (gp1 and gp2) have been previously implicated in this heterogeneous disease including CREB Signalling in Neurons as blocking of this pathway is thought to be associated with a decrease in *BDNF* transcription that may contribute to disease (Katanuma *et al*., 2014). In addition, Axonal Guidance Signalling and Netrin Signaling, have been labelled as potential predisposing pathways in this study and have also been found in the literature to be targeted pathways of schizophrenia associated genes such as PCDH12 (Gregório *et al*., 2009). Another interesting link to the literature is nNOS Signalling in Neurons, which was labeled as potentially predisposing in this analysis and has been previously found to have increased activity in affective disorders (Oliveira *et al*., 2008) . Novel and informative additional complete genome level studies using a larger number of samples will be needed to support this theoretical model.

A. Dopamine Feedback in cAMP signaling in Schizophrenia

The dopamine feedback in cAMP signalling pathway is one of the pathways that has emerged independently in both schizophrenia-affected twins in this study (Figure 3.11a). This pathway has been shown to be associated with psychiatric disorders due to the fact that it has critical function in integrating dopaminergic and glutamatergic signalling, and in turn affecting striatal function and plasticity (Kunii *et al*., 2014; Meyer-Lindenberg *et al*., 2007). The leading theory to account for the pathophysiology of schizophrenia involves an excess of the neurotransmitter dopamine, either through excess production or postsynaptic dopamine over-activity possibly mediated by increased receptor density (Abi-Dargham and Laruelle, 2005). Recent reports reinforce the importance of dopamine in schizophrenia, including strong associations with Dopamine

Receptor D_2 (*DRD2),* a target of many antipsychotic drugs (Ripke *et al*., 2014). The results support the contention that the defects associated with this pathway have the potential to contribute to schizophrenia in the two patients studied.

B. Glutamate Receptor Signalling in Schizophrenia

A glutamate receptor-signalling pathway is the second pathway that is found only in the two patients studied (Figure 3.11b). The significance of this pathway in this disease is backed by the underlying neurochemical basis of schizophrenia that includes a hypofunctional glutamate system (Labrie and Roder, 2010). Many genes associated with glutamatergic neurotransmission have been previously implicated in schizophrenia, including *GRM3, GRIN2A, SRR* and *GRIA1* (Ripke *et al*., 2014). Further, morphological alteration of dendrites of glutamatergic neurons in the cerebral cortex of schizophrenia-affected individuals has been reported (Hu *et al*., 2015), suggesting a role for glutamate signalling in the etiology of schizophrenia. Interestingly, schizophrenia patients show enriched *de novo* mutations in genes regulating the postsynaptic density at glutamatergic synapses (Hall *et al*., 2015). Furthermore, small *de novo* mutations were found to be overrepresented among glutamatergic postsynaptic proteins (Fromer *et al*., 2014) and genes harbouring detrimental *de novo* mutations were reported to be enriched in networks affecting protein interactions (Mostafavi *et al*., 2008). In addition, a recent study investigating rare mutations in exonic regions of genes implicated in schizophrenia and autism spectrum disorder (ASD) revealed that post-synaptic glutamate receptor complexes are key molecular mechanisms associated with schizophrenia and ASD (Kenny *et al*., 2014).

I propose that the two pathways described above have the potential to at least partially contribute to the disease in these two patients. More important, the gene defects associated with these two pathways in both patients may have allowed for crossing of the liability threshold and manifestation of the disease.

3.5 Conclusion

The genome-wide results included in this report are compatible with a threshold model for schizophrenia. They involve a number of pathways and an even larger number of genes. Most of the mutations involved may provide a family specific liability that may (e.g. Concordant MZ twins) or may not be alone sufficient (e.g. Discordant MZ twins) to develop the disease. In cases where it is not sufficient, additional aberrations (genetic, epigenetic or environmental) can shift the liability beyond the threshold for the development of disease. Interestingly, some of these changes may represent *de novo* mutations during ontogeny involving a variety of mutational mechanisms. I note that the effect of a *de novo* mutation may or may not be apparent depending on the degree of mosaicism caused by its timing and the organ system affected, but necessitate more intensive investigation. Finally, the comprehensive genomic results per individual included in this report provide a promising approach to the understanding of schizophrenia and may apply to other related disorders.

3.6 References

Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.

Abi-Dargham, A., and Laruelle, M. (2005). Mechanisms of action of second generation antipsychotic drugs in schizophrenia: insights from brain imaging studies. Eur. Psychiatry *20*, 15–27.

Baranzini, S.E., Mudge, J., van Velkinburgh, J.C., Khankhanian, P., Khrebtukova, I., Miller, N.A., Zhang, L., Farmer, A.D., Bell, C.J., Kim, R.W., *et al*. (2010). Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. Nature *464*, 1351–1356.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., *et al*. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature *456*, 53–59.

Bloom, R.J., Kahler, A.K., Collins, A.L., Chen, G., Cannon, T.D., Hultman, C., and Sullivan, P.F. (2013). Comprehensive analysis of copy number variation in monozygotic twins discordant for bipolar disorder or schizophrenia. Schizophr. Res. *146*, 289–290.

Bouchard, T.J., and McGue, M. (1981). Familial studies of intelligence: a review. Science *212*, 1055–1059.

Breckpot, J., Thienpont, B., Gewillig, M., Allegaert, K., Vermeesch, J.R., and Devriendt, K. (2012). Differences in Copy Number Variation between Discordant Monozygotic Twins as a Model for Exploring Chromosomal Mosaicism in Congenital Heart Defects. Mol. Syndromol. *2*, 81–87.

Bruder, C., and Piotrowski, A. *et al* (2008). Phenotypically Concordant and Discordant Monozygotic Twins Display Different DNA Copy Number Variation Profiles. Am. J. Hum. Genet. *82*, 1–9.

Bruder, C.E.G., Piotrowski, A., Gijsbers, A.A.C.J., Andersson, R., Erickson, S., Diaz de Ståhl, T., Menzel, U., Sandgren, J., von Tell, D., Poplawski, A., *et al.* (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am. J. Hum. Genet. *82*, 763–771.

Butler, J.L., Osborne Locke, M.E., Hill, K.A., and Daley, M. (2013). HD-CNV: hotspot detector for copy number variants. Bioinformatics *29*, 262–263.

Cardno, A.G., and Gottesman, I.I. (2000). Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. Am. J. Med. Genet. *97*, 12–17.

Carnevali, P., Baccash, J., Halpern, A.L., Nazarenko, I., Nilsen, G.B., Pant, K.P., Ebert, J.C., Brownley, A., Morenzoni, M., Karpinchyk, V., *et al.* (2012). Computational techniques for human genome resequencing using mated gapped reads. J. Comput. Biol. *19*, 279–292.

Castellani, C.A., Awamleh, Z., Melka, M.G., O'Reilly, R.L., and Singh, S.M. (2014). Copy number variation distribution in six monozygotic twin pairs discordant for schizophrenia. Twin Res. Hum. Genet. *17*, 108–120.

Dal, G.M., Ergüner, B., Sağıroğlu, M.S., Yüksel, B., Onat, O.E., Alkan, C., and Özçelik, T. (2014). Early postzygotic mutations contribute to *de novo* variation in a healthy monozygotic twin pair. J. Med. Genet. *51*, 455–459.

Dempster, E.L., Pidsley, R., Schalkwyk, L.C., Owens, S., Georgiades, A., Kane, F., Kalidindi, S., Picchioni, M., Kravariti, E., Toulopoulou, T., *et al.* (2011).

Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. Hum. Mol. Genet. *20*, 4786–4796.

Dempster, E.L., Wong, C.C.Y., Lester, K.J., Burrage, J., Gregory, A.M., Mill, J., and Eley, T.C. (2014). Genome-wide methylomic analysis of monozygotic twins discordant for adolescent depression. Biol. Psychiatry *76*, 977–983.

Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., *et al*. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science *327*, 78–81.

Ehli, E.A., Abdellaoui, A., Hu, Y., Hottenga, J.J., Kattenberg, M., van Beijsterveldt, T., Bartels, M., Althoff, R.R., Xiao, X., Scheet, P., *et al*. (2012). *De novo* and inherited CNVs in MZ twin pairs selected for discordance and concordance on Attention Problems. Eur. J. Hum. Genet. *20*, 1037–1043.

Farrell, M.S., Werge, T., Sklar, P., Owen, M.J., Ophoff, R.A., O'Donovan, M.C., Corvin, A., Cichon, S., and Sullivan, P.F. (2015). Evaluating historical candidate genes for schizophrenia. Mol. Psychiatry.

First, M.B., Spitzer, R.L., Gibbon, M., and and Williams, J.B.W. (1996). Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV) (Washington, D.C.: American Psychiatric Press, Inc).

First, M.B., Gibbon, M., Spitzer, R.L., Williams, J.B.W., and Benjamin, L.S. (1997). Structured Clinical Interview for DSM-IV Axis II Personality Disorders, (SCID-II) (Washington, D.C.: American Psychiatric Press, Inc.).

Forsberg, L.A., Rasi, C., Razzaghian, H.R., Pakalapati, G., Waite, L., Thilbeault, K.S., Ronowicz, A., Wineinger, N.E., Tiwari, H.K., Boomsma, D., *et al*. (2012). Age-related somatic structural changes in the nuclear genome of human blood cells. Am. J. Hum. Genet. *90*, 217–228.

Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suñer, D., Cigudosa, J.C., Urioste, M., Benitez, J., *et al*. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. Proc. Natl. Acad. Sci. U. S. A. *102*, 10604–10609.

Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D.M., *et al*. (2014). *De novo* mutations in schizophrenia implicate synaptic networks. Nature *506*, 179–184.

Georgieva, L., Rees, E., Moran, J.L., Chambert, K.D., Milanova, V., Craddock, N., Purcell, S., Sklar, P., McCarroll, S., Holmans, P., *et al*. (2014). *De novo* CNVs in bipolar affective disorder and schizophrenia. Hum. Mol. Genet. *23*, 6677–6683.

Girard, S.L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., Dionne-Laporte, A., Spiegelman, D., Henrion, E., Diallo, O., *et al*. (2011). Increased exonic *de novo* mutation rate in individuals with schizophrenia. Nat. Genet. *43*, 860–863.

Gottesman, I.I. (1991). Schizophrenia genesis: the origin of madness. (New York.: W.H. Freeman and Co.).

Gottesman, I.I., and Shields, J. (1967). A polygenic theory of schizophrenia. Proc. Natl. Acad. Sci. U. S. A. *58*, 199–205.

Gregório, S.P., Sallet, P.C., Do, K.-A., Lin, E., Gattaz, W.F., and Dias-Neto, E. (2009). Polymorphisms in genes involved in neurodevelopment may be associated with altered brain morphology in schizophrenia: preliminary evidence. Psychiatry Res. *165*, 1–9.

Gulsuner, S., and McClellan, J.M. (2014). *De novo* mutations in schizophrenia disrupt genes co-expressed in fetal prefrontal cortex. Neuropsychopharmacology *39*, 238–239.

Gulsuner, S., Walsh, T., Watts, A.C., Lee, M.K., Thornton, A.M., Casadei, S., Rippey, C., Shahin, H., Nimgaonkar, V.L., Go, R.C.P., *et al*. (2013). Spatial and temporal mapping of *de novo* mutations in schizophrenia to a fetal prefrontal cortical network. Cell *154*, 518–529.

Hall, J., Trent, S., Thomas, K.L., O'Donovan, M.C., and Owen, M.J. (2015). Genetic risk for schizophrenia: convergence on synaptic pathways involved in plasticity. Biol. Psychiatry *77*, 52–58.

Handunnetthi, L., Handel, A.E., and Ramagopalan, S. V (2010). Contribution of genetic, epigenetic and transcriptomic differences to twin discordance in multiple sclerosis. Expert Rev. Neurother. *10*, 1379–1381.

Hu, W., MacDonald, M.L., Elswick, D.E., and Sweet, R.A. (2015). The glutamate hypothesis of schizophrenia: evidence from human brain tissue studies. Ann. N. Y. Acad. Sci. *1338*, 38–57.

Insel, T.R. (2014). Brain somatic mutations: the dark matter of psychiatric genetics? Mol. Psychiatry *19*, 156–158.

Katanuma, Y., Numakawa, T., Adachi, N., Yamamoto, N., Ooshima, Y., Odaka, H., Inoue, T., and Kunugi, H. (2014). Phencyclidine rapidly decreases neuronal mRNA of brain-derived neurotrophic factor. Synapse *68*, 257–265.

Kavanagh, D.H., Tansey, K.E., O'Donovan, M.C., and Owen, M.J. (2014). Schizophrenia genetics: emerging themes for a complex disorder. Mol. Psychiatry *20*, 72–76.

Kendler, K.S. (2015). A joint history of the nature of genetic variation and the nature of schizophrenia. Mol. Psychiatry *20*, 77–83.

Kenny, E.M., Cormican, P., Furlong, S., Heron, E., Kenny, G., Fahey, C., Kelleher, E., Ennis, S., Tropea, D., Anney, R., *et al*. (2014). Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders. Mol. Psychiatry *19*, 872–879.

Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., *et al*. (2012). *De novo* CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol. Psychiatry *17*, 142–153.

Kondo, S., Schutte, B.C., Richardson, R.J., Bjork, B.C., Knight, A.S., Watanabe, Y., Howard, E., Ferreira de Lima, R.L.L., Daack-Hirsch, S., Sander, A., *et al*. (2002). Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat. Genet. *32*, 285–289.

Kong, S.W., Lee, I.H., Leshchiner, I., Krier, J., Kraft, P., Rehm, H.L., Green, R.C., Kohane, I.S., and MacRae, C.A. (2014). Summarizing polygenic risks for complex diseases in a clinical whole-genome report. Genet. Med.

Krawczak, M., Cooper, D.N., Fändrich, F., Engel, W., and Schmidtke, J. (2012). How to distinguish genetically between an alleged father and his monozygotic twin: a thought experiment. Forensic Sci. Int. Genet. *6*, e129–e130.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

Kumar, K., Maiti, S., Castellani, C.A., O'Reilly, R., and Singh, S.M. (2013). A novel deletion cluster at 13q14.2-q21.33 in an 80-year man with late onset leukemia: Clinical and molecular findings. Indian J. Hum. Genet. *19*, 96–100.

Kunii, Y., Hyde, T.M., Ye, T., Li, C., Kolachana, B., Dickinson, D., Weinberger, D.R., Kleinman, J.E., and Lipska, B.K. (2014). Revisiting DARPP-32 in

postmortem human brain: changes in schizophrenia and bipolar disorder and genetic associations with t-DARPP-32 expression. Mol. Psychiatry *19*, 192–199.

Labrie, V., and Roder, J.C. (2010). The involvement of the NMDA receptor D-serine/glycine site in the pathophysiology and treatment of schizophrenia. Neurosci. Biobehav. Rev. *34*, 351–372.

Li, R., Montpetit, A., Rousseau, M., Wu, S.Y.M., Greenwood, C.M.T., Spector, T.D., Pollak, M., Polychronakos, C., and Richards, J.B. (2014). Somatic point mutations occurring early in development: a monozygotic twin study. J. Med. Genet. *51*, 28–34.

Maiti, S., Kumar, K.H.B.G., Castellani, C.A., O'Reilly, R., and Singh, S.M. (2011). Ontogenetic *de novo* copy number variations (CNVs) as a source of genetic individuality: studies on two families with MZD twins for schizophrenia. PLoS One *6*, e17125.

McCarthy, S.E., Gillis, J., Kramer, M., Lihm, J., Yoon, S., Berstein, Y., Mistry, M., Pavlidis, P., Solomon, R., Ghiban, E., *et al*. (2014). *De novo* mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. Mol. Psychiatry *19*, 652–658.

McGue, M., Gottesman, I.I., and Rao, D.C. (1983). The transmission of schizophrenia under a multifactorial threshold model. Am. J. Hum. Genet. *35*, 1161–1178.

Meyer-Lindenberg, A., Straub, R.E., Lipska, B.K., Verchinski, B.A., Goldberg, T., Callicott, J.H., Egan, M.F., Huffaker, S.S., Mattay, V.S., Kolachana, B., *et al*. (2007). Genetic evidence implicating DARPP-32 in human frontostriatal structure, function, and cognition. J. Clin. Invest. *117*, 672–682.

Modinos, G., Iyegbe, C., Prata, D., Rivera, M., Kempton, M.J., Valmaggia, L.R., Sham, P.C., van Os, J., and McGuire, P. (2013). Molecular genetic gene-environment studies using candidate genes in schizophrenia: a systematic review. Schizophr. Res. *150*, 356–365.

Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C., and Morris, Q. (2008). GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome Biol. *9 Suppl 1*, S4.

Muotri, A.R., and Gage, F.H. (2006). Generation of neuronal variability and complexity. Nature *441*, 1087–1093.

Muotri, A.R., Marchetto, M.C.N., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K., and Gage, F.H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. Nature *468*, 443–446.

O'Huallachain, M., Karczewski, K.J., Weissman, S.M., Urban, A.E., and Snyder, M.P. (2012). Extensive genetic variation in somatic human tissues. Proc. Natl. Acad. Sci. U. S. A. *109*, 18018–18023.

O'Reilly, R., Torrey, E.F., Rao, J., and Singh, S. (2013). Monozygotic twins with early-onset schizophrenia and late-onset bipolar disorder: a case report. J. Med. Case Rep. *7*, 134.

Oliveira, R.M.W., Guimarães, F.S., and Deakin, J.F.W. (2008). Expression of neuronal nitric oxide synthase in the hippocampal formation in affective disorders. Braz. J. Med. Biol. Res. *41*, 333–341.

Pang, A.W., MacDonald, J.R., Pinto, D., Wei, J., Rafiq, M. a., Conrad, D.F., Park, H., Hurles, M.E., Lee, C., Venter, J.C., *et al*. (2010). Towards a comprehensive structural variation map of an individual human genome. Genome Biol *11*, R52–R52.

Pang, A.W.C., Macdonald, J.R., Yuen, R.K.C., Hayes, V.M., and Scherer, S.W. (2014). Performance of high-throughput sequencing for the discovery of genetic variation across the complete size spectrum. G3 (Bethesda). *4*, 63–65.

Petersen, B.-S., Spehlmann, M.E., Raedler, A., Stade, B., Thomsen, I., Rabionet, R., Rosenstiel, P., Schreiber, S., and Franke, A. (2014). Whole genome and exome sequencing of monozygotic twins discordant for Crohn's disease. BMC Genomics *15*, 564.

Picchioni, M.M., and Murray, R.M. (2007). Schizophrenia. BMJ *335*, 91–95.

Piotrowski, A., Bruder, C.E.G., Andersson, R., Diaz de Ståhl, T., Menzel, U., Sandgren, J., Poplawski, A., von Tell, D., Crasto, C., Bogdan, A., *et al*. (2008). Somatic mosaicism for copy number variation in differentiated human tissues. Hum. Mutat. *29*, 1118–1124.

Project, the 1000 G., Conrad, D.F., Keebler, J.E., Depristo, M.A., Lindsay, S.J., Zhang, Y., Casals, F., Idaghdour, Y., Hartl, C.L., Torroja, C., *et al*. (2011). Variation in genome-wide mutation rates within and between human families. Nat. Genet.

Rall, K., Eisenbeis, S., Barresi, G., Rückner, D., Walter, M., Poths, S., Wallwiener, D., Riess, O., Bonin, M., and Brucker, S. (2015). Mayer-Rokitansky-Küster-Hauser syndrome discordance in monozygotic twins: matrix

metalloproteinase 14, low-density lipoprotein receptor-related protein 10, extracellular matrix, and neoangiogenesis genes identified as candidate genes in a tissue-specific . Fertil. Steril. *103*, 494–502.e3.

Reumers, J., De Rijk, P., Zhao, H., Liekens, A., Smeets, D., Cleary, J., Van Loo, P., Van Den Bossche, M., Catthoor, K., Sabbe, B., *et al*. (2012). Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. Nat. Biotechnol. *30*, 61–68.

Ripke, S., Neale, B.M., Corvin, A., Walters, J.T.R., Farh, K.-H., Holmans, P.A., Lee, P., Bulik-Sullivan, B., Collier, D.A., Huang, H., *et al*. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

Schwab, S.G., and Wildenauer, D.B. (2013). Genetics of psychiatric disorders in the GWAS era: an update on schizophrenia. Eur. Arch. Psychiatry Clin. Neurosci. *263 Suppl* , S147–S154.

Singh, S.M., and O'Reilly, R. (2009). (Epi)genomics and neurodevelopment in schizophrenia: monozygotic twins discordant for schizophrenia augment the search for disease-related (epi)genomic alterations. Genome *52*, 8–19.

Singh, S.M., Castellani, C.A., and O'Reilly, R.L. (2009). Copy number variation showers in schizophrenia: an emerging hypothesis. Mol. Psychiatry *14*, 356–358.

Szatkiewicz, J.P., O'Dushlaine, C., Chen, G., Chambert, K., Moran, J.L., Neale, B.M., Fromer, M., Ruderfer, D., Akterin, S., Bergen, S.E., *et al*. (2014). Copy number variation in schizophrenia in Sweden. Mol. Psychiatry *19*, 762–773.

Vadlamudi, L., Dibbens, L.M., Lawrence, K.M., Iona, X., McMahon, J.M., Murrell, W., Mackay-Sim, A., Scheffer, I.E., and Berkovic, S.F. (2010). Timing of *de novo* mutagenesis--a twin study of sodium-channel mutations. N. Engl. J. Med. *363*, 1335–1340.

Wain, L. V, Pedroso, I., Landers, J.E., Breen, G., Shaw, C.E., Leigh, P.N., Brown, R.H., Tobin, M.D., and Al-Chalabi, A. (2009). The role of copy number variation in susceptibility to amyotrophic lateral sclerosis: genome-wide association study and comparison with published loci. PLoS One *4*, e8175.

Weber-Lehmann, J., Schilling, E., Gradl, G., Richter, D.C., Wiehler, J., and Rolf, B. (2014). Finding the needle in the haystack: differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. Forensic Sci. Int. Genet. *9*, 42–46.

Xu, B., Roos, J.L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., Gogos, J.A., and Karayiorgou, M. (2011). Exome sequencing supports a *de novo* mutational paradigm for schizophrenia. Nat. Genet. *43*, 864–868.

Xue, Y., Chen, Y., Ayub, Q., Huang, N., Ball, E. V, Mort, M., Phillips, A.D., Shaw, K., Stenson, P.D., Cooper, D.N., *et al*. (2012). Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. Am. J. Hum. Genet. *91*, 1022–1032.

Yavas, G., Koyutürk, M., Özsoyo, M., Gould, M.P., Koyuturk, M., Ozsoyoglu, M., Gould, M.P., and LaFramboise, T. (2009). An optimization framework for unsupervised identification of rare copy number variation from SNP array data. Genome Biol. *10*, R119–R2009 – 10–10 – r119. Epub 2009 Oct 23.

Ye, K., Beekman, M., Lameijer, E.-W., Zhang, Y., Moed, M.H., van den Akker, E.B., Deelen, J., Houwing-Duistermaat, J.J., Kremer, D., Anvar, S.Y., *et al*. (2013). Aging as accelerated accumulation of somatic variants: whole-genome sequencing of centenarian and middle-aged monozygotic twin pairs. Twin Res. Hum. Genet. *16*, 1026–1032.

A version of this chapter has been published in Castellani *et al.*, *BMC Medical Genomics* (2015)

Chapter 4 - DNA Methylation Analysis of Families with Monozygotic Twins Discordant for Schizophrenia Identifies Psychosis Related Genes and Networks

4.0 Overview

The involvement of DNA methylation in schizophrenia appears logical, but remains to be established. I have used blood DNA from two pairs of monozygotic twins discordant for schizophrenia and their parents in order to assess genome-wide methylation using a NimbleGen 720k Methylation Promoter Microarray.

The genome-wide results show that differentially methylated regions (DMRs) exist between members representing discordant monozygotic twins. Some DMRs are shared with parent(s) and others appear to be *de novo*. I found twenty-seven genes affected by DMR changes that were shared in the affected member of two discordant monozygotic pairs from unrelated families. Interestingly, the genes affected by pair-specific DMRs share networks. Specifically, this study has identified two networks; ‘cell death and survival’ and ‘cellular movement and immune cell trafficking’. These two networks and the genes affected have been previously implicated in the etiology of schizophrenia.

The results are compatible with the suggestion that DNA methylation may contribute to the discordance of monozygotic twins for schizophrenia. Also, this may be accomplished by the effect of gene specific methylation changes on specific biological networks rather than individual genes.

4.1 Introduction

Results during the last few years have established that differences between MZ twins exist at the genetic (Bruder *et al.*, 2008; Maiti *et al.*, 2011) as well as epigenetic (Fraga *et al.*, 2005; Ling and Groop, 2009; Petronis, 2010) levels. In fact, methylation differences between identical twins have been reported as early as in newborns (Ollikainen *et al.*, 2010). Also, DNA methylation

is reported to increase with age (Levesque *et al*., 2014; Fraga *et al*., 2005). Epigenetic differences between MZ twins include features like X-inactivation, genomic imprinting, or differential methylation of genes, and may cause MZ twin pairs to diverge, leading to disease discordance (Dempster *et al*., 2011; Singh *et al*., 2002). Studies of this kind have concluded that no two individuals are alike; not even identical twins (Maiti *et al*., 2011). However, the genetic similarity between MZ twins is comparable to no other two individuals. In addition, identical twins are matched for age, sex, maternal environment, and population cohort effects - making them the best matched control available (Bell and Spector, 2011). Indeed, MZ twins provide a unique backdrop to assess epigenetic states that are shared due to inheritance or common environments, as well as differences that may be in response to individual specific exposures or random events (Fraga *et al*., 2005; Ling and Groop, 2009; Manikkam *et al*., 2012; Petronis, 2010; Wong *et al*., 2010). These changes, affecting critical cellular processes, may allow monozygotic twins to develop discordance for almost any trait through reprogramming of gene expression via epigenetic mechanisms which may increase liability to disease (Xu *et al*., 2014). This is particularly relevant in neurodevelopmental disorders, especially schizophrenia, and reports are now accumulating from twin studies to support an epigenetic model of disease contribution. For instance, a schizophrenic twin from a pair of discordant MZ twins is epigenetically more similar to affected concordant MZ twins than to their own unaffected co-twin at the *DRD2* gene (Petronis *et al*., 2003). In addition, methylation of genes in blood samples of twins discordant for schizophrenia, including medication free patients, shows hypermethylation and hypomethylation of several genes (Bonsch *et al*., 2012; Kinoshita *et al*., 2013). Indeed, the molecular results accumulating on schizophrenia are encouraging and include many recent reports of associations between DNA methylation and schizophrenia (Abdolmaleky *et al*., 2004; Dempster *et al*., 2011; Wockner *et al*., 2014).

The research presented here identifies genes whose methylation is altered in schizophrenia patients as compared to their unaffected MZ twin. It uses blood DNA from two sets of monozygotic twin pairs discordant for schizophrenia and their parents. The results identify DNA methylation differences between MZD twins in two families discordant for schizophrenia. Also, the patients across families share affected genes, and more importantly, biological networks.

4.2 Methods

4.2.1 Subjects

This study on monozygotic twins received ethics approval by the University of Western Ontario's Committee on research involving human subjects (Appendix B). All subjects provided written informed consent to participate in this study. All of the patients were adults at the time of consent. Capacity for consent was ensured using three measures 1) Schizophrenic patients gave consent only during a "normal" phase (no psychosis present), 2) Both twins of the twin pair were present and gave consent at the same time (the normal twin and their affected sibling), 3) If R.O'Reilly felt that the capacity to consent was compromised, the patients were not included in the study. They were interviewed and clinically assessed by a single senior Psychiatrist (R. O'Reilly) using the SCID-I and SCID-II (First *et al*., 1996, 1997). Past clinical notes were available to aid in diagnosis. Both families were comprised of identical female twins. The twins from Family 1 (Figure 4.1) were Afro-American females aged 53. The affected member of twin pair 1 was diagnosed with schizophrenia at age 22. The twins were discordant for 31 years at the time of sample collection. The twins from Family 2 (Figure 4.1) were Caucasian females aged 43. The affected member of twin pair 2 was diagnosed with schizoaffective disorder at age 27. The twins were discordant for 16 years at the time of sample collection. It should be noted that the Father of Twin Pair 1 was diagnosed with Chronic Leukemia (CLL) at age 69. The affected patient from Family 1 was treated for

schizophrenia symptoms using a combination of the medications Clozapine, Divalproex and Benztropine. The affected patient from Family 2 was treated for schizophrenia symptoms using a combination of the medications Seroquel, Effexor and Topiramate.

4.2.2 DNA Collection and Array Processing

The twins and their parents (Figure 4.1) included in this study contributed whole blood samples. DNA was extracted from whole blood using the PerfectPure DNA Blood Kit (Gaithersburg, MD, USA), following the manufacturer's protocol. Zygosity was confirmed by Affymetrix 6.0 microarray and specifically using the Affymetrix Genotyping Console 4.0 concordance feature (Maiti *et al.*, 2011).

The genomic DNA was processed at ArrayStar Inc (Rockville, MD, USA); this included the methylated DNA immunoprecipitation (MeDIP), sample labeling, and hybridization to the NimbleGen Human DNA Methylation Promoter Plus CpG Island 720k Array. The NimbleGen Human DNA Methylation 3x720k CpG Island Plus RefSeq Promoter Microarray is a multiplex slide with 3 identical arrays per slide. Each Roche Nimblegen Inc (Madison, MI, USA) array covers 27,728 annotated CpG islands as well as 22,532 promoters of the RefSeq genes derived from the UCSC RefFlat files. Median-centering, quantile normalization, and linear smoothing was performed by Bioconductor packages *Ringo*, *limma,* and *MEDME* at ArrayStar. Lastly, in order to compare the two groups' differentially enriched regions the average of the log2-ratio values for each group (i.e. experimental patient [E] and healthy control [C]) was used to calculate an M' value (defined by the following equation) for each probe:

$$\mathbf{M' = Average\ (\log_2 MeDIP_E / Input_E) - Average\ (\log_2 MeDIP_C / Input_C)}$$

The differential enrichment peaks were filtered according to the following criteria:

i) At least one of the two groups has a median (log2 MeDIP/Input) > =0.3 and M'> 0.

ii) At least half of probes in a peak may have coefficient of variability (CV) < = 0.8 in both groups.

Before hybridization to the array, genomic DNA was sonicated to random fragments in size of about 200–1000 bp. Immunoprecipitation of methylated DNA was performed using Biomag™ magnetic beads coupled to a mouse monoclonal antibody against 5-methylcytidine. The immunoprecipitated DNA was eluted and purified by phenol chloroform extraction and ethanol precipitation. The total input and immunoprecipitated DNA were labeled with Cy3- and Cy5-labeled random 9-mers. Scanning was performed with the Axon GenePix 4000B microarray scanner. Raw data was extracted as pair files by NimbleScan software.

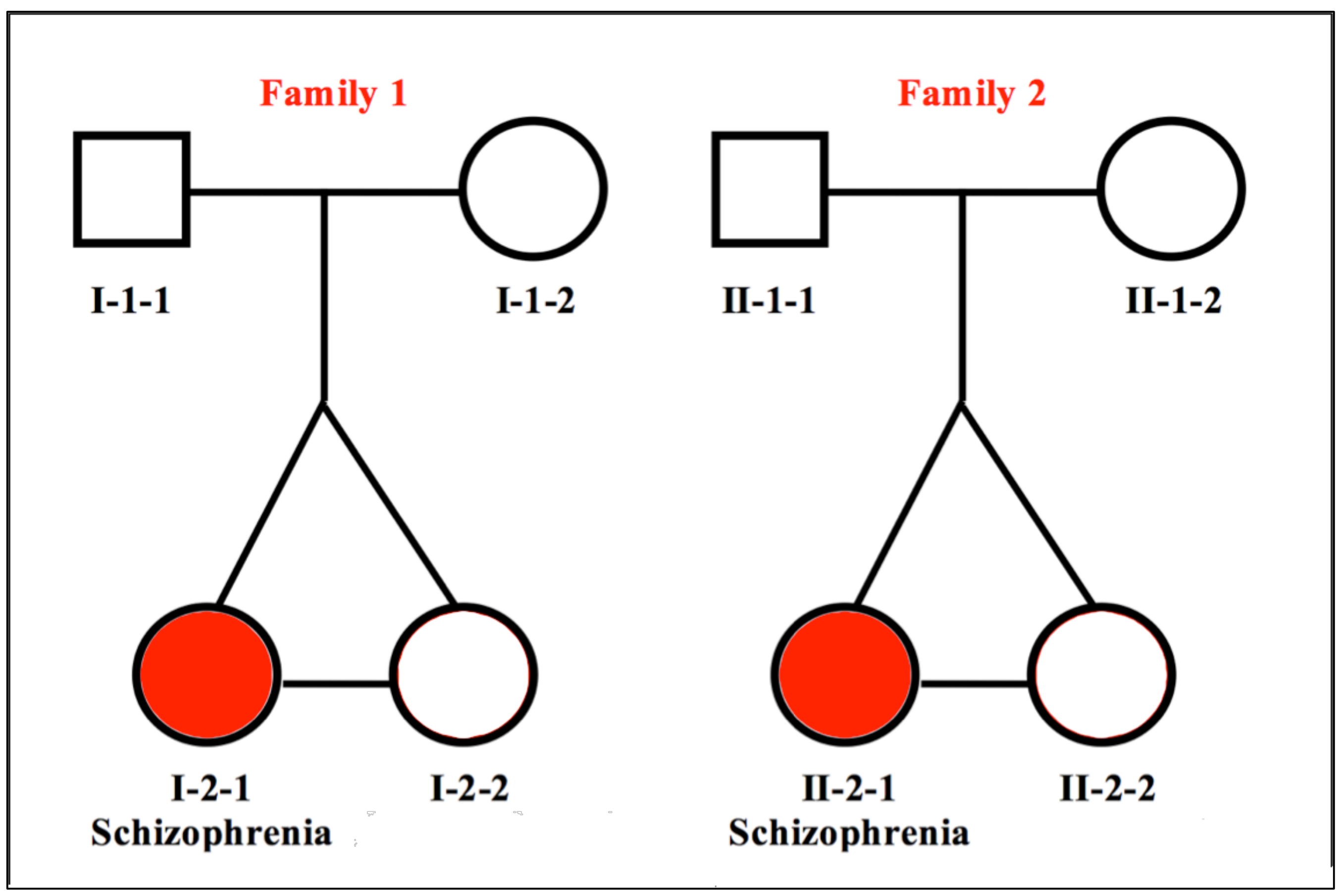


Figure 4.1. Pedigrees of families included in the study.

Shaded circles represent affected twins.

4.2.3 Microarray Analysis

The pair files were analyzed with the tiling workflow in Partek Genomics Suite® version 6.6 (St. Louis, Missouri, USA). Nimblegen scan pair files (635 nm and 532 nm) for each sample were annotated against NCBI Build 36/hg18 and enriched regions were detected using a two-way ANOVA between an affected twin and their unaffected co-twin. The enriched regions settings were set at a minimum p-value of 0.001 and the number of probes to call a region was set at a minimum of 4. MAT scores were generated for each differentially methylated region. Overlapping genes were then identified as those RefSeq (2014-04-29 version) genes that were either within the gene or 5000 bp upstream or 3000 downstream of the gene. Differentially methylated regions (DMRs) in each affected twin were identified in relation to the pattern in the well twin. Also, the presence or absence of each DMR was assessed as familial or *de novo* based on their presence or absence in Mom and/or Dad.

The identified genes with significant changes in DNA methylation between twins discordant for schizophrenia (DMRs) were then analyzed using Ingenuity Pathway Analysis (Ingenuity Systems Inc, CA, USA) towards identification of networks and canonical pathways overrepresented in the enriched genes. Also, pathway analysis and gene ontology analysis were conducted using Partek Pathways (Fishers Exact Test) and Enrichr (Chen *et al.*, 2013). Shared genes were annotated with imprinting data from GeneImprint (http://www.geneimprint.com) and The Catalogue of Parent of Origin Effects (http://igc.otago.ac.nz/home.html).

4.3 Results

I report the genome-wide analysis of methylation differences in two families with monozygotic twins discordant for schizophrenia using the NimbleGen Human DNA Methylation Promoter Plus CpG Island 720k Array. I

analyzed the data with Partek Genomics Suite and yielded three main lines of results presented below:

4.3.1 MZ twins show differences in DNA Methylation profiles

The genome-wide DNA methylation profiles have revealed differentially methylated regions (DMRs) between the MZ twin pairs in the study ($p \leq 0.001$). These differences were widespread and distributed across all chromosomes. Further, the availability of parental data has allowed for assessment of each DMR for its presence/absence in the two parents. The results show that methylation profiles in twins include both shared and *de novo* events. I note that in Family 1 as well as in Family 2, most of the DMRs appear *de novo* (are not found in Mom or Dad). Only 13% and 25% of the DMRs in each family, respectively, were present in at least one parent. The results have also allowed identification of specific genes that are differentially methylated between the affected twin and their identical unaffected twin. Specifically, I note that 330 genes are differentially methylated in the twin pair from Family 1 (Appendix F) and 138 genes are differentially methylated in Family 2 (Appendix G). A visual representation of these results is given in Figure 4.2. As might be expected, some of the DMRs were increased in methylation when compared to co-twin and others were decreased in methylation when compared to co-twin. An overlap between the DMRs present in the affected member of the two unrelated families (Figure 4.3) suggests that most (80-92%) DMRs are twin pair specific. Chromosome 1 (36 and 19 DMRs respectively) and Chromosome 15 (30 and 21 DMRs respectively) harbour a large proportion of DMRs in Family 1 as well as in Family 2. Family 1 also has a large number of DMRs on Chromosome 19 (30 DMRs).

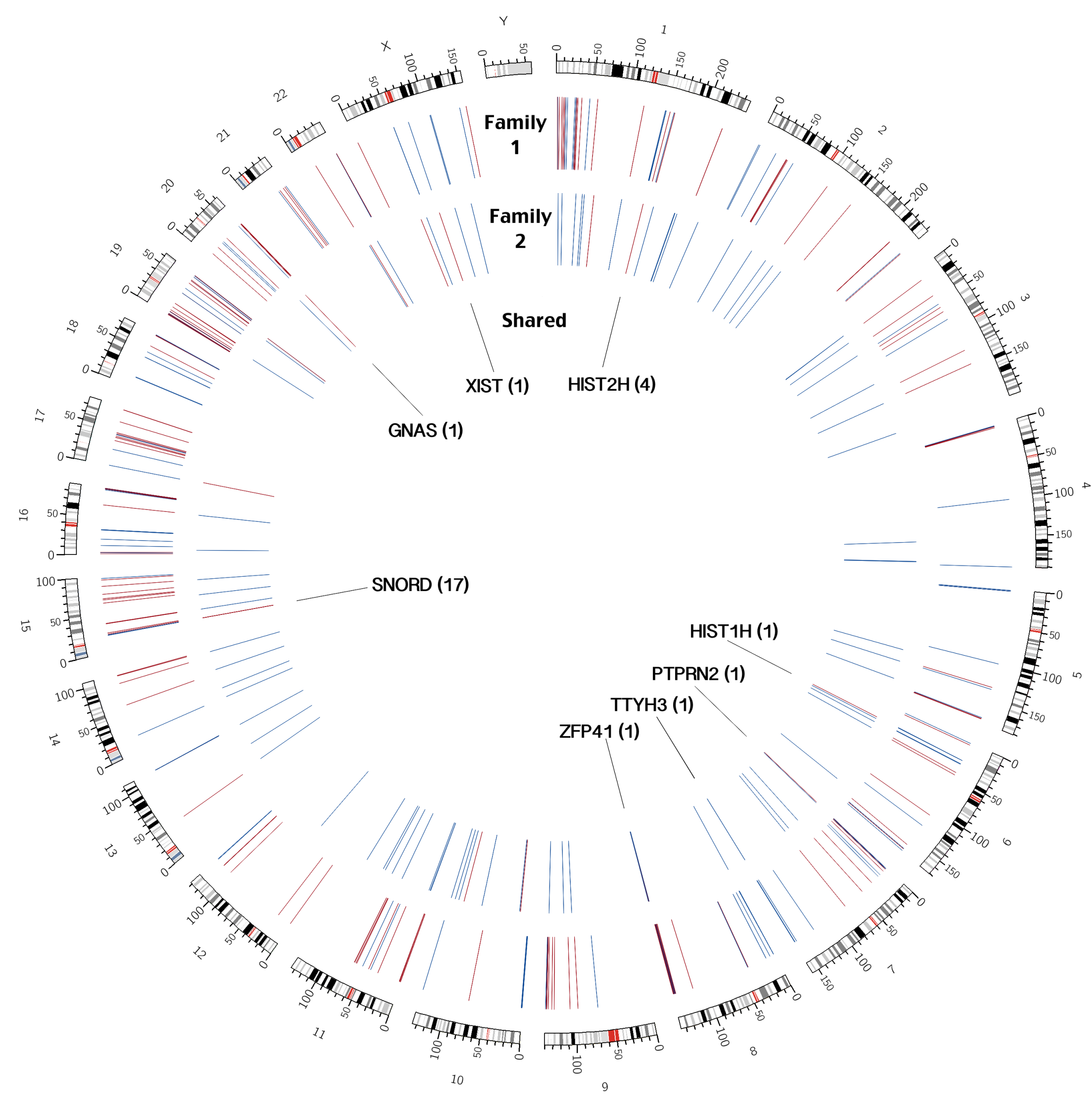


Figure 4.2. Differential methylation in two twin pairs visualized in a Circos plot covering all chromosomes.

Red represents a decrease in methylation in the affected twin and blue represents an increase in methylation in the affected twin. The outside track represents differentially methylated regions in the affected member of Family 1, the middle track represents differentially methylated regions in the affected member of Family 2 and the inside track represents the shared regions found in both unrelated affected twins. Shared genes are labeled.

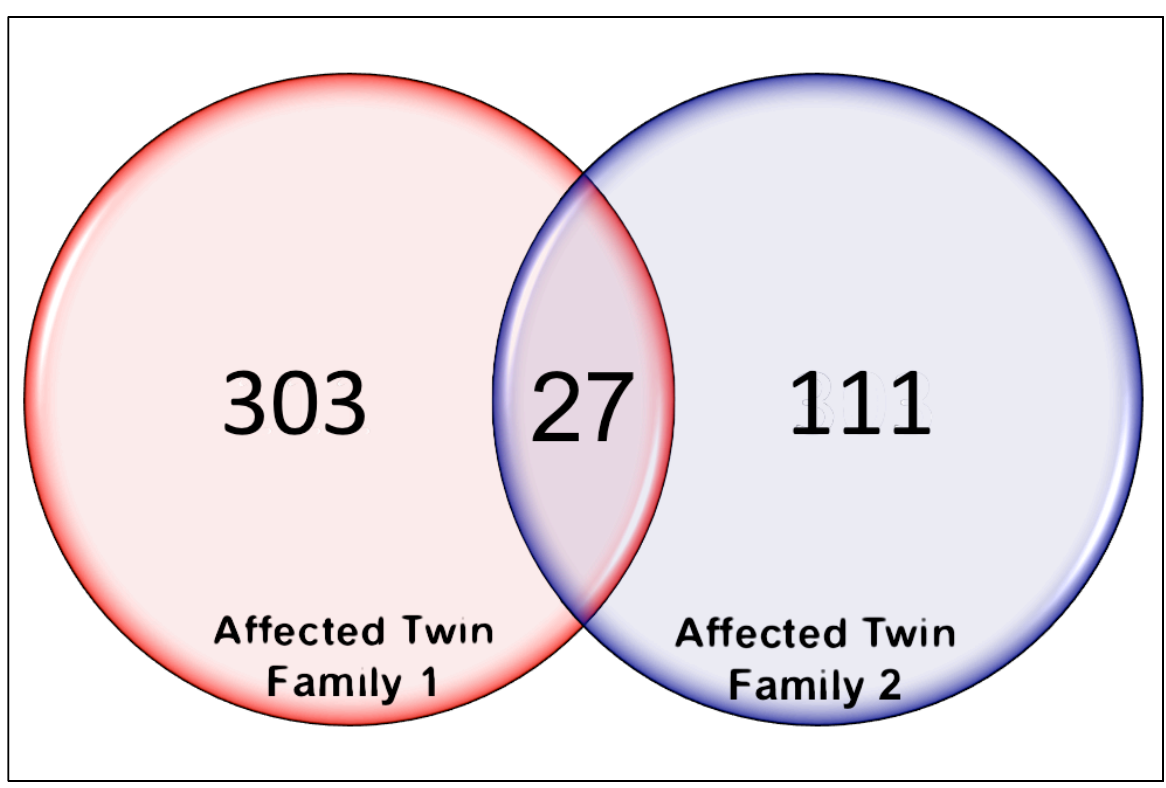


Figure 4.3. Overlap of differentially methylated regions between two families.

Venn diagram showing the number of genes differentially methylated in each patient (138 and 330, respectively) as well as the genes enriched in both affected twins in this study (27).

4.3.2 MZ twins discordant for schizophrenia share genomic regions of differential methylation

Figure 4.3 highlights 27 genes that were differentially methylated in the affected twin in both families. Consequentially, there are 303 genes that are only differentially methylated in Family 1 and 111 genes that are only differentially methylated in Family 2 (Figure 4.3). Of the 27 genes that showed methylation differences in both sets of twins, 24 were increased in methylation status in the affected twins and the remaining showed a decrease in methylation. The genes identified included five *HIST* (Chromosome 1 and 6) and 17 *SNORD115-116* genes (Chromosome 15). Also, the imprinted gene *GNAS* (Chromosome 20) as well as *XIST* (Chromosome X) were found to share DMR regions between the two patients. *PTPRN2* and *TTYH3*, both found on Chromosome 7, as well as *ZFP41* (Chromosome 8), are the remaining shared DMRs. The list of genes (Table 4.1) identified share common regions with exact DMR beginning/end locations in the two patients. The exception to this pattern was the *PTPRN2*, *TTYH3* and *ZFP41* regions where the beginning/end locations were found to be different but nearby (Table 4.1). As expected, all DMRs identified are specific to the promoter regions.

As stated, five of the common genes identified belong to either the *HIST2H* cluster on Chromosome 1 or the *HIST1H* region on Chromosome 6 (Table 4.1). Further, 17 of the 27 genes belong to either the *SNORD115* or *SNORD116* clusters on chromosome 15. All seventeen of the *SNORD* genes identified in two patients are known to be genomically imprinted and are thought to produce ncRNA transcripts that undergo extensive processing (Table 4.1). A Manhattan Plot of the region encompassing the SNORD genes is presented in detail in Figure 4.4, this region extends from the *SNRPN* gene to the *UBE3A* gene and encompasses both the *SNORD115* and *SNORD116* gene families. In addition, a Manhattan Plot of the HIST region on Chromosome 1 is presented in Figure 4.5.

Table 4.1. Differentially methylated regions identified in two affected MZD twins belonging to two unrelated families. (NCBI Build 36/hg18)

Transcript	Chr	Region Start	Region End	Methylation Status	MAT Score Family 1	MAT Score Family 2	DMR In Family 1 Parental	DMR in Family 2 Parental
HIST2H2AA3	1	148085850	148085870	Increase	12.7557	8.45215	YES (Mom)	YES (Both)
HIST2H2AA4	1	148085850	148085870	Increase	12.7557	8.45215	YES (Mom)	YES (Both)
HIST2H3A	1	148085850	148085870	Increase	12.7557	8.45215	YES (Mom)	YES (Both)
HIST2H3C	1	148085850	148085870	Increase	12.7557	8.45215	YES (Mom)	YES (Both)
HIST1H1C	6	26164302	26164322	Increase	7.54892	6.0215	NO	NO
PTPRN2	7	F1: 157352628 F2: 157141154	F1:157352648 F2:157141174	Increase	7.11519	2.41616	NO	NO
TTYH3	7	F1: 2664585 F2: 2653547	F1: 2664605 F2: 2653567	F1:Decrease F2:Increase	−2.82779	2.47725	NO	NO
ZFP41	8	F1: 144409425 F2: 144403919	F1:144409445 F2:144403939	Increase	15.1625	2.05379	NO	NO
SNORD115-10	15	22983806	22983826	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD115-11	15	22983806	22983826	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD115-12	15	22983806	22983826	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD115-29	15	22983806	22983826	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD115-33	15	23030052	23030072	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD115-34	15	23030052	23030072	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD115-35	15	23030052	23030072	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD115-36	15	22983806	22983826	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD115-37	15	23030052	23030072	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD115-43	15	22983806	22983826	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD115-5	15	22983806	22983826	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD115-9	15	22983806	22983826	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD116-10	15	22868070	22868090	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD116-11	15	22868070	22868090	Increase	7.49051	7.8983	YES (Both)	YES (Both)

SNORD116-3	15	22868070	22868090	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD116-8	15	22868070	22868090	Increase	7.49051	7.8983	YES (Both)	YES (Both)
SNORD116-9	15	22868070	22868090	Increase	7.49051	7.8983	YES (Both)	YES (Both)
GNAS	20	56879799	56879819	F1:Increase F2:Decrease	17.2187	−10.7502	YES (Both)	YES (Both)
XIST	X	72974756	72974776	F1:Increase F2:Decrease	12.1903	−7.36623	NO	YES (Both)

Note: Chr = Chromosome; F1 = Family 1; F2 = Family

Ingenuity Pathway Analysis (IPA) with the 27 shared genes has identified 'Protein Kinase A Signaling' as the most enriched canonical pathway (*p* = 3.09E-04). In addition, 'Granzyme A Signaling' (*p* = 6.83E-03), 'G Protein Signaling Mediated by Tubby' (*p* = 1.24E-02), 'Serotonin Receptor Signaling' (*p* = 1.72E-02) and 'UVB-Induced MAPK Signaling' (*p* = 2.12E-02) were identified as canonical pathways of interest (Table 4.2C). IPA also identified *DRD4*, a dopamine receptor gene, to be the top upstream regulator of these twenty-seven genes.

Similarly, IPA identified 'Developmental Disorders' (*p* = 4.03E-04-1.21E-03) as a top disease associated with this gene set. In addition, 'Cell Signaling' (*p* = 4.03E-04-3.73E-02), 'Nucleic Acid Metabolism' (*p* = 4.03E-04-3.73E-02) and 'Gene Expression' (*p* = 3.62E-03-9.63E-03) were the most significant molecular and cellular functions. Interestingly, 'Nervous System Development and Function' (*p* = 1.61E-03) and 'Immune Cell Trafficking' (*p* = 6.43E-03) were two top physiological systems related to this gene set. Further, Infectious Disease, Hereditary Disorders, Embryonic Development and Cell Death and Survival were associated network functions related to the differentially methylated gene set that overlapped in both schizophrenic twins. A full summary of the IPA results on these 27 genes can be found in Table 4.2C.

When the 27 genes were analyzed using Enrichr (Chen *et al.*, 2013), expression in whole brain was identified as the top human gene atlas finding. Enrichr also identified OMIM disease classifications related to neurodevelopment to be enriched in the gene list; these included Asperger's Syndrome (*p* = 0.039) and Mental Retardation (*p* = 0.065).

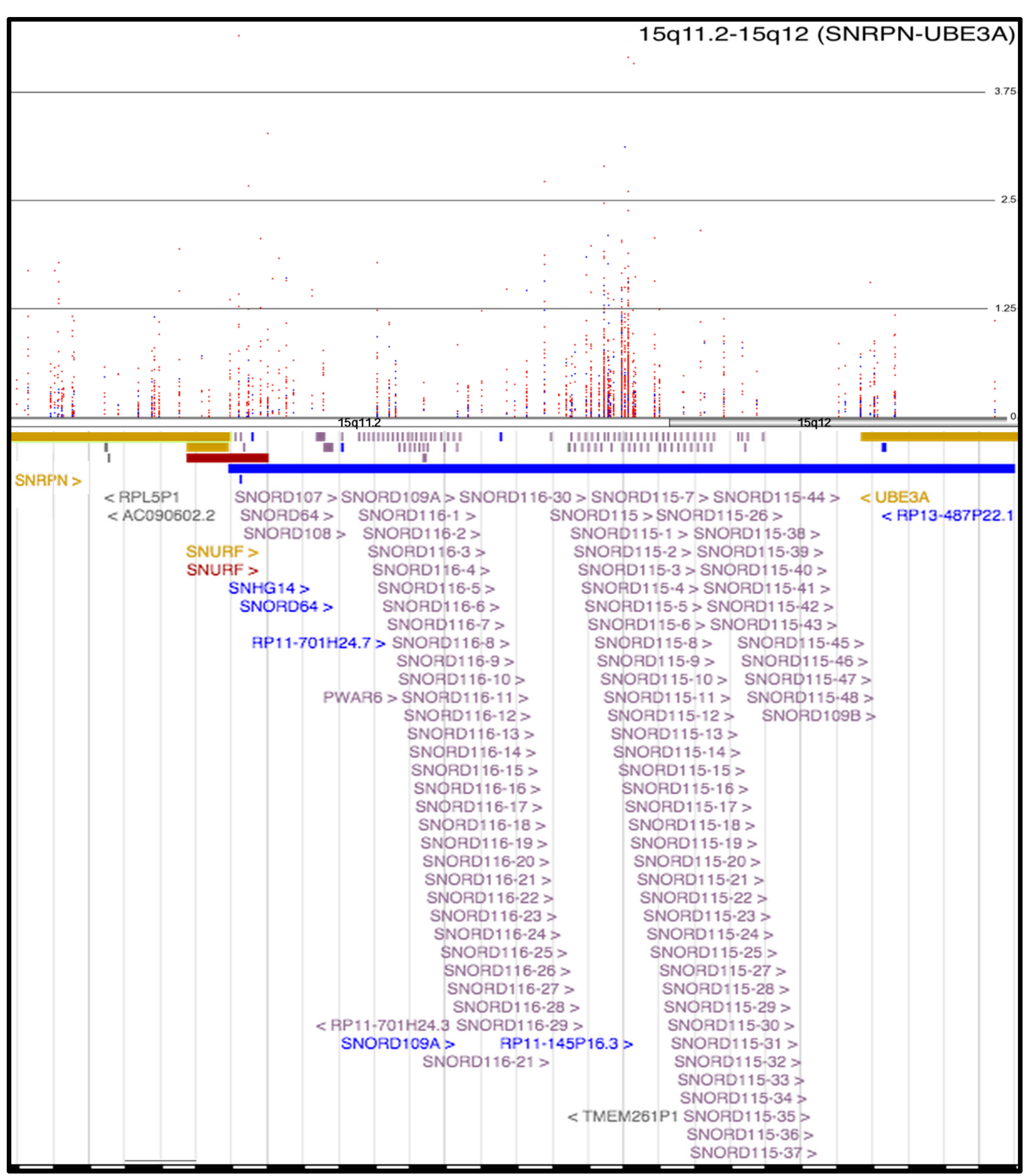


Figure 4.4. Manhattan plot representing methylation in the 15q11.2-15q12 region (Chr 15:17527952-25617776).

This region spans from *SNRPN* to *UBE3A* and encompasses members of the *SNORD115* and *SNORD116* gene families. A red dot indicates an increase in methylation in the affected twin. A blue dot indicates a decrease in methylation in the affected twin. This region contains a complex regulatory ncRNA involved in imprinting control and neurodevelopment.

Figure 4.5. Manhattan plot representing methylation in the 1q21.1-1q21.3 region (Chr 1:143518837-152631162).

This region includes many genes including the histone genes discussed in this report. A red dot indicates an increase in methylation in the affected twin. A blue dot indicates a decrease in methylation in the affected twin.

4.3.3 MZ twins discordant for schizophrenia have differentially methylated networks and pathways; some (i) Pair-specific and others (ii) Shared

i. Pair-Specific Networks and Pathways

IPA analysis on the DMRs identified in Family 1 implicates 'Skeletal and Muscular Disorders' (*p* = 9.34E-04-4.07E-02) as a top disorder. In addition, 'Cancer' (*p* = 2.09E-04-4.07E-02), 'Gastrointestinal Disease' (*p* = 8.86E-04-4.07E-02) and 'Organismal Injury and Abnormalities' (p = 9.34E-04-3.94E-02) were identified as top diseases or disorders overrepresented in Family 1. Similarly, in Family 2 the pathway analysis of DMR associated genes highlights 'Developmental Disorders' (*p* = 3.93E-03-3.21E-02), 'Hereditary Disorders' (*p* = 3.93E-03-4.58E-02), 'Skeletal and Muscular Disorders', (*p* = 3.93E-03-3.21E-02) and 'Neurological Disease' (*p* = 4.77E-03-3.54E-02) as top associated diseases and disorders (Table 4.2A and B). The analysis also identified a number of interesting canonical pathways in Family 1 including 'Hepatic Cholestasis' (*p* = 1.05E-04), 'Granzyme A Signaling' (*p* = 1.51E-03) and the 'STAT3 Pathway' (*p* = 3.31E-03). In Family 2, 'Human Embryonic Stem Cell Pluripotency' (*p* = 6.45E-05), 'Tec Kinase Signaling' (*p* = 1.51E-04) and 'IL-4 Signaling' (*p* = 6.49E-03) were the top canonical pathways identified.

Also, in Family 1, 'Embryonic Development' emerged as an overrepresented physiological system in the gene list (*p* = 1.88E-04-4.07E-02). In Family 2, the 'Immune Cell Trafficking' physiological system was identified by this analysis (*p* = 5.20E-03-4.13E-02).

A number of additional pair-specific networks also emerged from this analysis of DMRs in each family including 'Hereditary Disorder, Developmental Disorder' in Family 2 and 'Connective Tissue Development and Function' in Family 1.

These and other twin specific networks and functions can be found in Tables 4.2A and B.

Table 4.2. Ingenuity Pathway Analysis (IPA) results A) Family 1 B) Family 2 C) Subset of 27 genes found in both affected twins.

4.2A. Family 1

Associated Network Functions		**IPA Score**
Cell Death and Survival, Cellular Development, Connective Tissue Development and Function		Score=39
Cellular Movement, Immune Cell Trafficking, Hematological System Development and Function		Score=12
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease		Score=11
Cell Death and Survival, Cellular Movement, Renal Necrosis/Cell Death		Score=11
Cell Morphology, Cellular Function and Maintenance, Cell Cycle		Score=9
Biological Function	***p-value***	**Genes**
Physiological System Development and Function		
Cardiovascular System Development and Function	1.88E-04 - 4.07E-02	10
Digestive System Development and Function	1.88E-04	2
Embryonic Development	1.88E-04 - 4.07E-02	14
Organ Development	1.88E-04 - 4.07E-02	6
Molecular and Cellular Functions		
Lipid Metabolism	1.88E-04 - 4.07E-02	7
Molecular Transport	1.88E-04 - 4.07E-02	9
Small Molecule Biochemistry	1.88E-04 - 4.07E-02	19
Cellular Growth and Proliferation	3.06E-04 - 4.07E-02	21
Cell Morphology	5.59E-04 - 3.63E-02	10
Diseases and Disorders		
Cancer	2.09E-04 - 4.07E-02	204
Gastrointestinal Disease	8.86E-04 - 4.07E-02	29
Organismal Injury and Abnormalities	9.34E-04 - 3.94E-02	46
Skeletal and Muscular Disorders	9.34E-04 - 4.07E-02	36
Canonical Pathways		
Hepatic Cholestasis	1.05E-04	9/141 (0.064)
Granzyme A Signaling	1.51E-03	3/17 (0.176)
Ovarian Cancer Signaling	2.24E-03	7/138 (0.051)
STAT3 Pathway	3.31E-03	5/74 (0.068)
Colorectal Cancer Metastasis Signaling	4.9E-03	9/244 (0.037)

4.2B. Family 2

Associated Network Functions		IPA Score
Cell Death and Survival, Cellular Movement, Cellular Function and Maintenance		Score=19
Cellular Movement, Immune Cell Trafficking, Hematological System Development and Function		Score=17
Hereditary Disorder, Skeletal and Muscular Disorders, Developmental Disorder		Score=9
Connective Tissue Disorders, Dermatological Diseases and Conditions, Hematological System Development and Function		Score=2
Organ Morphology, Reproductive System Development and Function, Cellular Function and Maintenance		Score=2
Biological Function	***p-value***	**Genes**
Physiological System Development and Function		
Cardiovascular System Development and Function	5.20E-03 - 3.58E-02	1
Hair and Skin Development and Function	5.20E-03 - 4.58E-02	2
Hematological System Development and Function	5.20E-03 - 4.13E-02	7
Immune Cell Trafficking	5.20E-03 - 4.13E-02	3
Molecular and Cellular Functions		
Carbohydrate Metabolism	7.88E-04 - 2.06E-02	3
Lipid Metabolism	7.88E-04 - 2.06E-02	3
Small Molecule Biochemistry	7.88E-04 - 4.58E-02	7
Gene Expression	2.63E-03 - 4.58E-02	18
Cell Death and Survival	5.20E-03 - 3.16E-02	13
Diseases and Disorders		
Developmental Disorder	3.93E-03 - 3.21E-02	10
Hereditary Disorder	3.93E-03 - 4.58E-02	19
Skeletal and Muscular Disorders	3.93E-03 - 3.21E-02	5
Neurological Disease	4.77E-03 - 3.54E-02	13
Canonical Pathways		
Human Embryonic Stem Cell Pluripotency	6.45E-05	6/149 (0.04)
Tec Kinase Signaling	1.51E-04	6/175 (0.034)
Sphingosine-1-phosphate Signaling	2.42E-03	4/115 (0.035)
Renal Cell Carcinoma Signaling	5.55E-03	3/71 (0.042)
IL-4 Signaling	6.49E-03	3/75 (0.04)

4.2C. Shared Differentially Methylated Regions (DMRs) In Both Families

Associated Network Functions		**IPA Score**
Cell Death and Survival, Cellular Function and Maintenance, Connective Tissue Development and Function		Score=3
Infectious Disease, Cancer, Gastrointestinal Disease		Score=3
Tissue Morphology, Organismal Survival, Gene Expression		Score=3
Hereditary Disorder, Gene Expression, Embryonic Development		Score=3
DNA Replication, Recombination, and Repair, Gene Expression, Cancer		Score=3
Biological Function	***p-value***	**Genes**
Physiological System Development and Function		
Nervous System Development and Function	1.61E-03	1
Hematological System Development and Function	6.43E-03 - 2.47E-02	1
Immune Cell Trafficking	6.43E-03	1
Behavior	2.04E-02	1
Molecular and Cellular Functions		
Cell Signaling	4.03E-04 - 3.73E-02	1
Nucleic Acid Metabolism	4.03E-04 - 3.73E-02	1
Small Molecule Biochemistry	4.03E-04 - 3.73E-02	1
Gene Expression	3.62E-03 - 9.63E-03	1
Cellular Movement	6.43E-03	1
Diseases and Disorders		
Cancer	4.03E-04 - 4.50E-02	5
Connective Tissue Disorders	4.03E-04 - 8.05E-04	1
Developmental Disorder	4.03E-04 - 1.21E-03	1
Endocrine System Disorders	4.03E-04 - 4.50E-02	2
Canonical Pathways		
Protein Kinase A Signaling	3.09E-04	3/368 (0.008)
Granzyme A Signaling	6.83E-03	1/17 (0.059)
G Protein Signaling Mediated by Tubby	1.24E-02	1/31 (0.032)
Serotonin Receptor Signaling	1.72E-02	1/43 (0.023)
UVB-Induced MAPK Signaling	2.12E-02	1/53 (0.019)

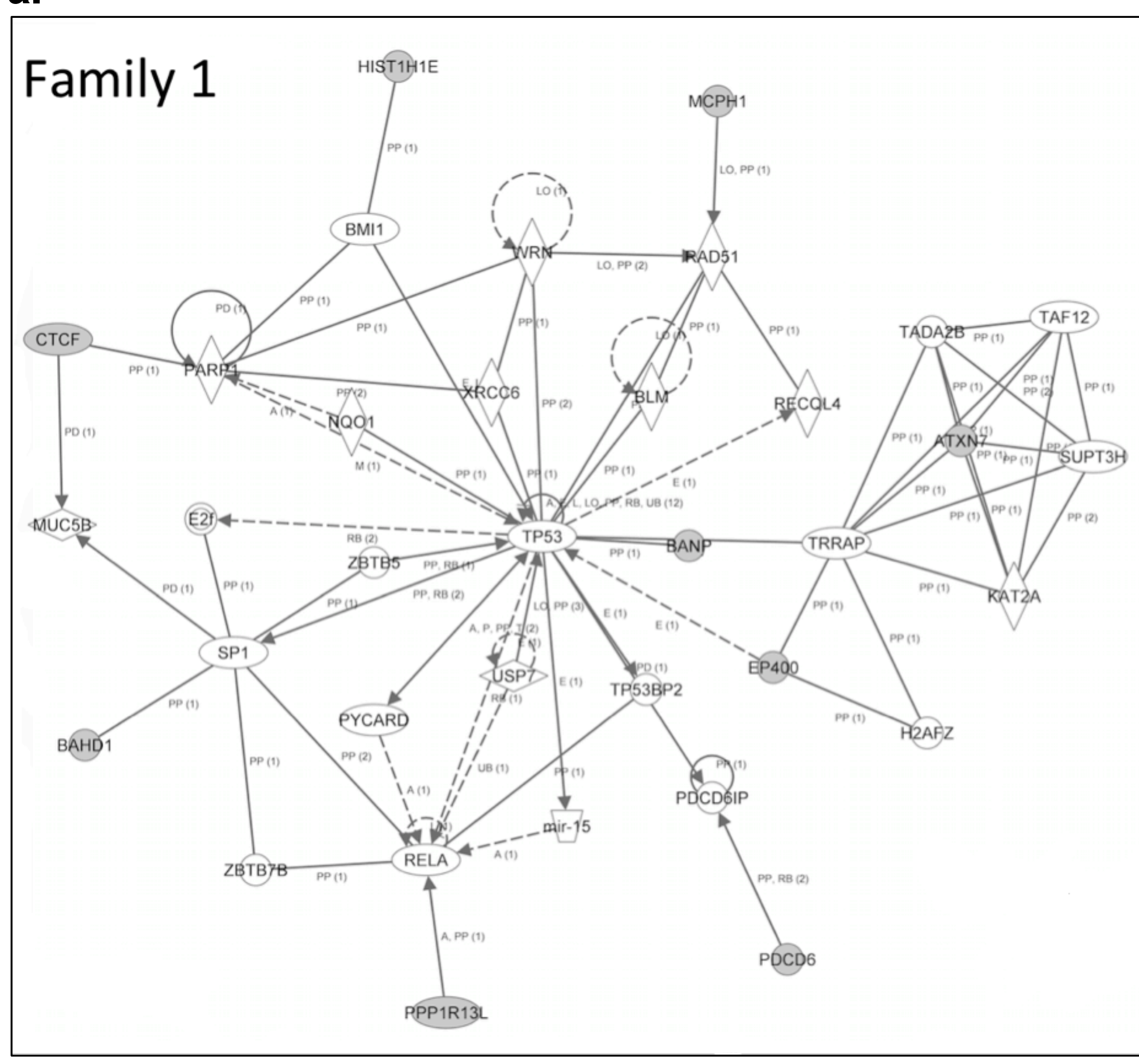
ii. Shared Networks and Pathways

When Ingenuity Pathways Analysis (IPA) was used on the independent gene sets identified in each family, some of the identified networks arose as top results in *both* families. This differs from the previous focus on 27 shared genes between families by focusing on the shared pathways that emerged from the independent gene lists (Figure 4.6). The shared pathways include a 'Cell Death and Survival' network (ratio of differentially methylated genes to total number of genes in the network was 9/35 and 12/35, in Family 1 and 2, respectively) (Figure 4.6a) and a 'Cellular Movement and Immune Cell Trafficking' network (ratio of differentially methylated genes to total number of genes in the network was 11/35 and 14/35 in Family 1 and 2, respectively) (Figure 4.6b). In Family 1, the 'Cell Death and Survival' network identified *TP53* as the primary hub gene of the network, while in Family 2 *IL1B* was identified as the primary hub gene in this network. Similarly, the 'Cellular Movement and Immune Cell Trafficking' networks revealed *IGF1R* and *EGFR* as hub genes in Family 1 and *TNF* and *IFNG* as hub genes in Family 2.

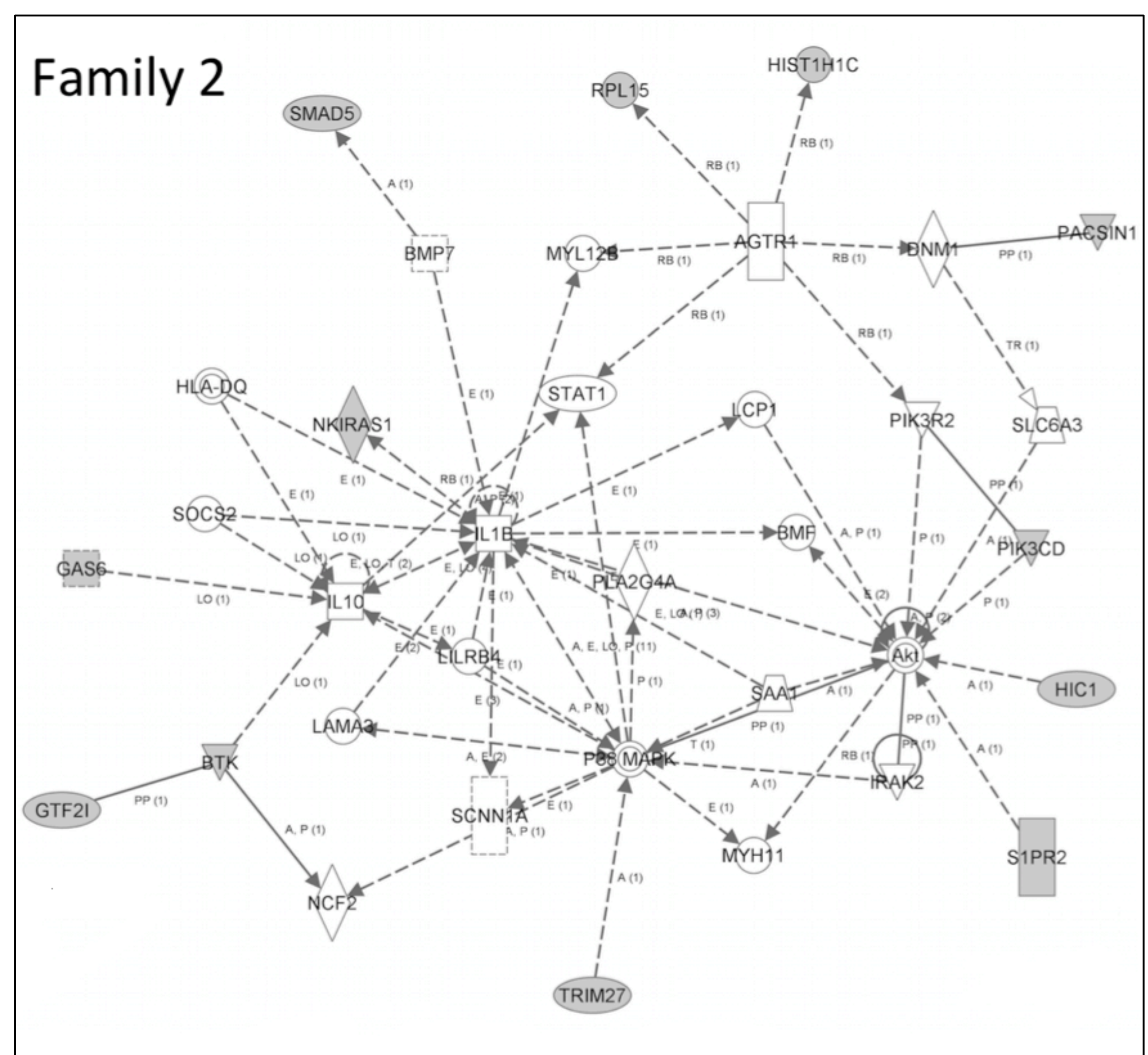
The disorder classification 'Skeletal and Muscular Disorders' was identified in both affected twins. Also, two shared molecular and cellular functions of interest were identified in both Family 1 and Family 2, namely 'Lipid Metabolism' (F1, p = 1.88E-04-4.07E-02; F2, p = 7.88E-04-2.06E-02) and Small Molecular Biochemistry (F1, p = 1.88E-04-4.07E-02; F1, p = 7.88E-04-4.58E-02).

Figure 4.6. Common networks identified in both families. Legend for symbols can be found in Appendix L.

a.

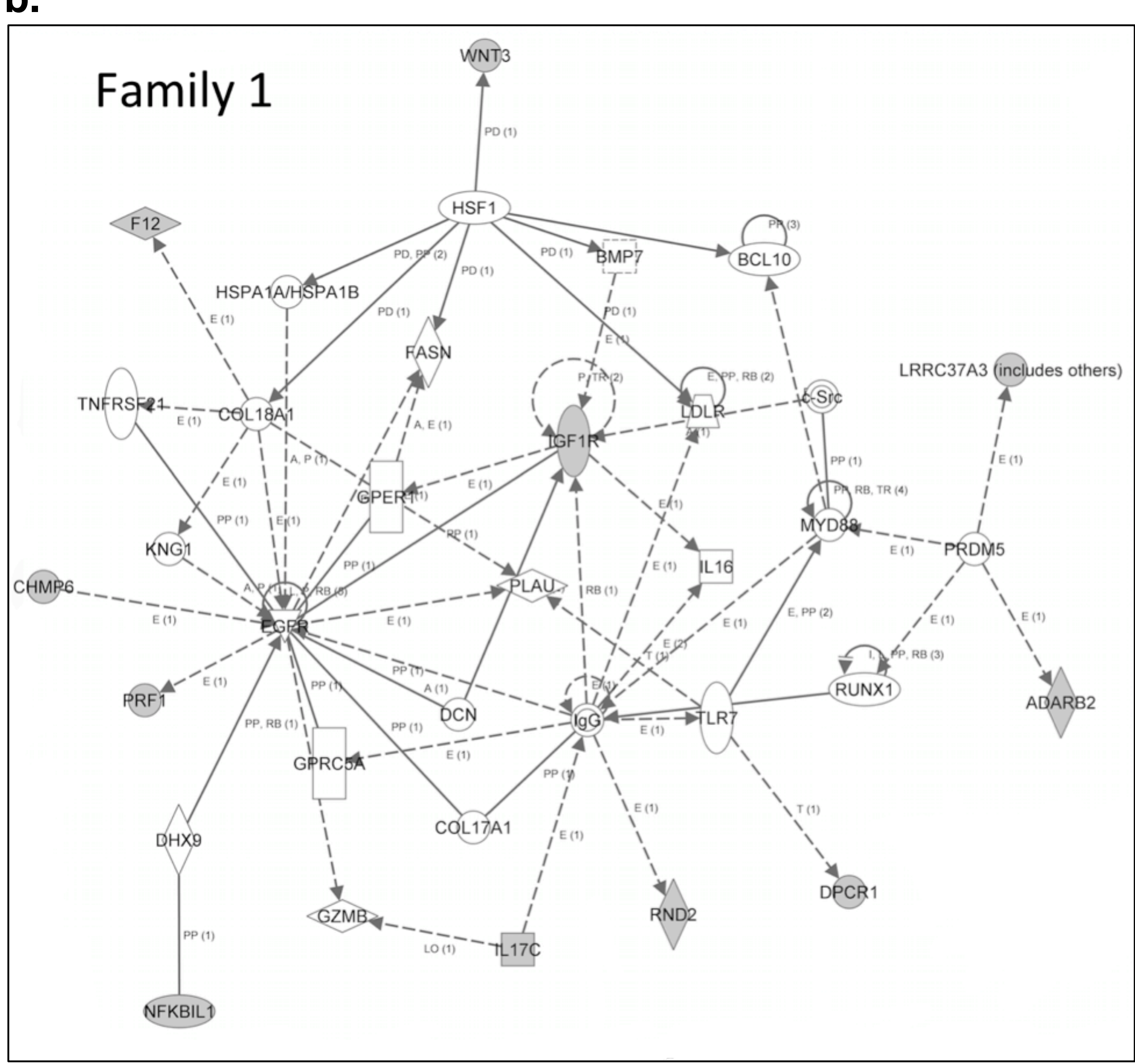


a. continued

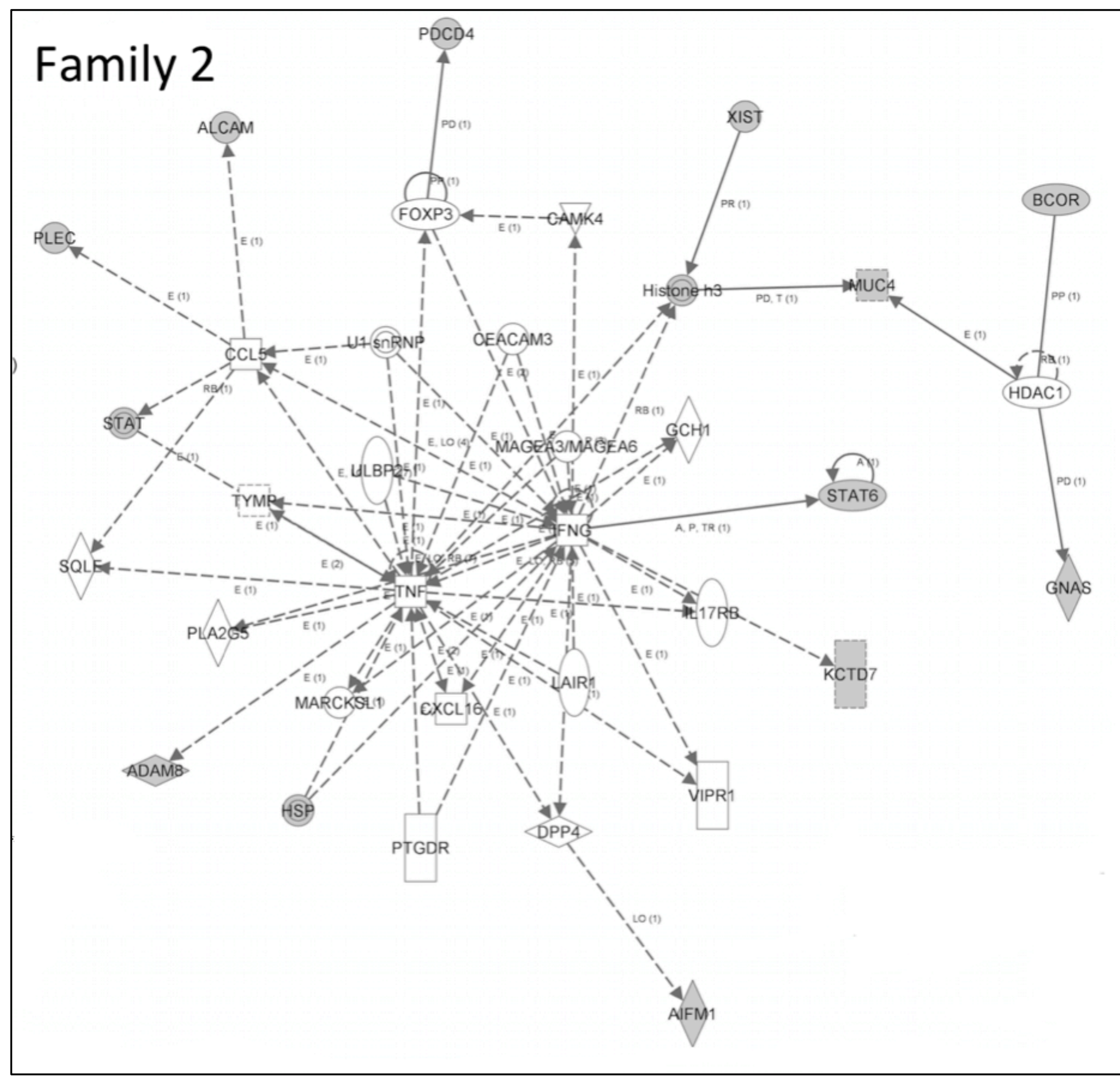


4.6a. Two ‘Cell Death and Survival’ networks independently identified in each affected twin in the study. Ingenuity Pathway Analysis (IPA) identified the networks in both affected twins in the study. Shading represents genes in this study that are differentially methylated.

b.



b. continued



4.6b. Two ‘Cellular Movement and Immune Cell Trafficking’ networks independently identified in each affected twin in the study. Ingenuity Pathway Analysis (IPA) identified the networks in both affected twins in the study. Shading represents genes in this study that are differentially methylated.

4.4 Discussion

The results included in this section show that a) monozygotic twins differ in DNA methylation, b) that this difference is genome-wide, c) that it includes a relatively large number of *provisional de novo* events, and d) with some exceptions, the differences are pair-specific. The results support the argument that *de novo* methylation changes are common during development and aging (Abdolmalcky *et al*., 2004; Wockner *et al*., 2014). Further, the results obtained in this report are specific to the two patients and will not necessarily cover the whole spectrum of the disease. Although the two MZD pairs for schizophrenia are unrelated, they share differences in DNA methylation in 27 genes and genomic locations. Also, the differentially methylated genes in each family affect two shared networks that are compatible with the development of this neurodevelopmental disease. Finally, the genes identified have the potential to explain the discordance of both twin pairs studied. The two patients share identified networks affecting cell death and immune cell trafficking. Of special interest to this discussion are (i) *HIST* genes primarily located on chromosome 1 and (ii) *SNORD115* and *SNORD116* genes located on chromosome 15.

i. HIST Genes

It is noteworthy that my analysis identified 5 *HIST* genes as genes of interest in the discordance for schizophrenia. The histone coding gene family, has already been implicated in the causation of schizophrenia (Dempster *et al*., 2011; Mill *et al*., 2008; Wockner *et al*., 2014). A Histone gene cluster on Chromosome 6p22.1 has also been implicated in a meta-analysis of schizophrenia associated loci in individuals of European ancestry (Shi *et al*., 2009). In addition, altered histone methylation has been found in olfactory cells, implicating oxidative stress in schizophrenia (Kano *et al*., 2013). Lastly, post-mortem brain tissue from schizophrenia patients has been found to have higher levels of histone deacetylase, *HDAC1*, and the level of *HDAC1* has been shown to be inversely correlated with *GAD67* protein expression, which tends to be

decreased in schizophrenia patients (Gavin and Sharma, 2010; Sharma *et al*., 2008). Together, these results argue that histones may play a role in this complex disease. As mentioned, my observation on altered methylation of a HIST cluster follows previous reports and argues that this alteration may also contribute to the discordance of these two twin pairs for schizophrenia.

ii. SNORD Genes

The *SNPRN-UBE3A* locus, which encompasses the *SNORD115* and *SNORD116* gene families, is a complex non-coding RNA region that spans 15q11-q13 (Runte *et al*., 2001). Noncoding RNAs, including miRNAs, are known to fine-tune gene expression through transcriptional and post-transcriptional regulations including RNA stability and protein translation (Gavin and Akbarian, 2012). In addition to serving as an antisense RNA for *UBE3A*, the polycistronic transcript also undergoes extensive processing, including the production of a number of small nucleolar RNA species (snoRNAs). The HBII-52 snoRNAs (also known as the *SNORD115* family) regulate the alternative splicing of the 5HTR2C serotonin receptor and result in an increased serotonin response in neurons (Leung *et al*., 2009). *SNORD115* is further processed into processed snoRNAs (psnoRNAs) that go on to regulate alternative splicing in a number of other transcripts, including epigenomic modifiers (Kishore *et al*., 2010). However, a conflicting report emerged in 2012 that showed evidence against psnoRNAs, indicating that *SNORD115* and *SNORD116* may generate genuine snoRNAs (Bortolin-Cavaille and Cavaille, 2012). A novel ncRNA species, LncRNAs with snoRNA ends, also originates from this loci. They are functionally distinct from snoRNAs and lncRNAs and are associated with the FOX family of splicing regulators that alter the alternative splicing of a number of other genes. In addition to psnoRNAs, snoRNAs, and lnc-snoRNAs, the snoRNAs are even further processed into snoRNA-derived RNAs (sdRNAs). These sdRNAs are proposed to come in two variations: some resembling miRNAs that associate with argonaute proteins to regulate translation and another longer type that form

complexes to influence gene expression (Falaleeva and Stamm, 2013). Further investigation into the locus has shown that it produces even more ncRNA products, with the introns forming the snoRNA derivatives and the exons forming two distinct but overlapping neuronal lncRNA clouds from the *SNORD115* and *SNORD116* regions that are involved in modulating circadian rhythm and energy expenditure (Powell *et al*., 2013a, 2013b). The lncRNAs are functionally distinct from the earlier identified ncRNA species and are also primarily expressed in developing neurons (Powell *et al*., 2013a).

Interestingly, the lncRNA from the *SNPRN-UBE3A* region has been suggested to regulate another imprinted locus, the *DLK1-DIO3* region, which is the only other known imprinted cluster of ncRNA that produces lncRNA, miRNA, and snoRNA. It is also involved in neurodevelopment and suggests that imprinted ncRNAs are capable of 'genomic cross-talk' (Murrell, 2014; Stelzer *et al*., 2014). Interestingly, while imprinting disorders are known to originate from these loci, a highly resolved and restricted deletion in the *SNORD116* region was identified as the minimal mutation to cause Prader-Willi Syndrome (Bieth *et al*., 2014).

Apart from the HIST and SNORD genes identified and discussed above, there are five other gene promoters that were differentially methylated in both families. These were *PTPRN2, TTYH3, ZFP41, GNAS and XIST. PTPRN2* is a phosphatase that is thought to be involved in the regulation of insulin secretion. *TTYH3* is a member of the Tweety family of proteins, which functions as a chloride anion channel. *ZFP41* is a zinc finger protein; this class of proteins has been reported to be associated with psychosis and related disorders (Sun *et al*., 2015). *GNAS*, an imprinted gene, has been previously associated with deficit schizophrenia in an Italian population (Minoretti *et al*., 2006). Lastly, *XIST,* on chromosome X, is essential for the initiation and spread of X-inactivation.

Also, *DRD4*, which was identified as a top upstream regulator of the shared *DMR* gene set, has been previously implicated in schizophrenia, and is thought to be the target of many antipsychotics (Lai *et al*., 2010).

The identified networks across unrelated twins share common functions supporting the hypothesis that a different set of patient-specific gene insults may lead to disease symptoms. There has been a long held linkage between schizophrenia and immune cell function. This theory gained further support as novel functions of immune molecules in the brain and cross-talk between the immune system and the central nervous system (Debnath *et al*., 2013). In addition, a number of studies have shown up-regulation of immuno-inflammatory genes in the CNS (Debnath *et al*., 2013; Saetre *et al*., 2007; Soderlund *et al*., 2009) as well as immune system gene modulation of synaptic function (Schmitt *et al*., 2011). In the 'cellular movement and immune cell trafficking' networks identified in Family 2, two genes (*TNF* and *IFNG)* immerged as hub genes. The tumor necrosis factor (*TNF*) had been associated with schizophrenia and it has also been reported that immune dysregulation could have a genetic component in schizophrenia patients (Boin *et al*., 2001). In addition, a single nucleotide polymorphism in the interferon gamma gene (*IFNG*) had been associated with paranoid schizophrenia in males (Paul-Samojedny *et al*., 2011), however, the role of these gene in the pathophysiology of the disease remains to be elucidated. Similarly, the other hub gene of the 'cellular movement and immune cell trafficking' network, *EGFR,* identified in Family 1 has also been previously associated with schizophrenia (Benzel *et al*., 2007).

The human *p53* tumor suppressor gene (*TP53*), which is identified as a primary hub gene in the 'cell death and survival' network of Family 1 in this study, plays a role in neurodevelopment and was previously associated with schizophrenia (Ni *et al*., 2005). Similarly, the primary hub gene in the 'cell death and survival' network, *IL1B*, which is differentially methylated in Family 2, has an important role in the development of the central nervous system. The *IL1B* gene is also reported to be associated with schizophrenia (Hanninen *et al*., 2008).

Finally, the question of the effect of the observed DNA methylation on disease must be interpreted with caution. I may add that most schizophrenia patients are exposed to antipsychotic drugs in North America. The patients were

under treatment with medications, though not consistently. I note that such drugs and cellular heterogeneity of the studied samples (Liu *et al*., 2013) may also affect DNA methylation (Melka *et al*., 2014) as confounding factors. Further, although practical, I recognize the use of blood to make inferences regarding a brain-based disorder is not without caution and recommend that this discovery study be complemented by population studies of larger sample size on this disease as well as confirmation of the findings using alternative technologies. At this time, it is not possible to eliminate these and other confounding factors that may affect the results. The epigenetic changes I have identified would not necessarily be expected to remain stable over time nor continue to show specificity for pathways relevant to schizophrenia. However, the observations on genes and pathways relevant to the disease, lend support to the complexity of this neurodevelopmental disorder and its etiology.

4.4.1 Conclusions

I report genome-wide methylation differences between monozygotic twins discordant for schizophrenia. A number of genes and networks identified are twin pair-specific, while others are shared between two unrelated patients. Specifically, the results identify three sets of gene clusters, *HIST* (Chromosome 1), *SNORD115* and *SNORD116* (Chromosome 15), which are differentially methylated in the two twins with schizophrenia as compared to their unaffected counterpart. I also report common networks identified independently in the two patients that relate to cell death/survival and immune cell trafficking networks disrupted in schizophrenia. The results on monozygotic discordant twins argue for a network based rather than gene focused approach in the understanding of schizophrenia and related disorders.

4.5 References

Abdolmaleky, H.M., Smith, C.L., Faraone, S. V, Shafa, R., Stone, W., Glatt, S.J., and Tsuang, M.T. (2004). Methylomics in psychiatry: Modulation of gene-environment interactions may be through DNA methylation. Am. J. Med. Genet. B, Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet. *127B*, 51–59.

Bell, J.T., and Spector, T.D. (2011). A twin approach to unraveling epigenetics. Trends Genet. *27*, 116–125.

Benzel, I., Bansal, A., Browning, B.L., Galwey, N.W., Maycox, P.R., McGinnis, R., Smart, D., St Clair, D., Yates, P., and Purvis, I. (2007). Interactions among genes in the ErbB-Neuregulin signalling network are associated with increased susceptibility to schizophrenia. Behav. Brain Funct. *3*, 31.

Bieth, E., Eddiry, S., Gaston, V., Lorenzini, F., Buffet, A., Conte Auriol, F., Molinas, C., Cailley, D., Rooryck, C., Arveiler, B., *et al.* (2014). Highly restricted deletion of the SNORD116 region is implicated in Prader-Willi Syndrome. Eur. J. Hum. Genet. 23, 252-5.

Boin, F., Zanardini, R., Pioli, R., Altamura, C.A., Maes, M., and Gennarelli, M. (2001). Association between -G308A tumor necrosis factor alpha gene polymorphism and schizophrenia. Mol. Psychiatry *6*, 79–82.

Bonsch, D., Wunschel, M., Lenz, B., Janssen, G., Weisbrod, M., and Sauer, H. (2012). Methylation matters? Decreased methylation status of genomic DNA in the blood of schizophrenic twins. Psychiatry Res. *198*, 533–537.

Bortolin-Cavaille, M.L., and Cavaille, J. (2012). The SNORD115 (H/MBII-52) and SNORD116 (H/MBII-85) gene clusters at the imprinted Prader-Willi locus generate canonical box C/D snoRNAs. Nucleic Acids Res. *40*, 6800–6807.

Bruder, C.E.G., Piotrowski, A., Gijsbers, A.A.C.J., Andersson, R., Erickson, S., Diaz de Ståhl, T., Menzel, U., Sandgren, J., von Tell, D., Poplawski, A., *et al.* (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am. J. Hum. Genet. *82*, 763–771.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V, Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Debnath, M., Cannon, D.M., and Venkatasubramanian, G. (2013). Variation in the major histocompatibility complex [MHC] gene family in schizophrenia:

associations and functional implications. Prog. Neuropsychopharmacol. Biol. Psychiatry *42*, 49–62.

Dempster, E.L., Pidsley, R., Schalkwyk, L.C., Owens, S., Georgiades, A., Kane, F., Kalidindi, S., Picchioni, M., Kravariti, E., Toulopoulou, T., *et al*. (2011). Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. Hum. Mol. Genet. *20*, 4786–4796.

Falaleeva, M., and Stamm, S. (2013). Processing of snoRNAs as a new source of regulatory non-coding RNAs: snoRNA fragments form a new class of functional RNAs. Bioessays *35*, 46–54.

First, M.B., Spitzer, R.L., Gibbon, M., and and Williams, J.B.W. (1996). Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV) (Washington, D.C.: American Psychiatric Press, Inc).

First, M.B., Gibbon, M., Spitzer, R.L., Williams, J.B.W., and Benjamin, L.S. (1997). Structured Clinical Interview for DSM-IV Axis II Personality Disorders, (SCID-II) (Washington, D.C.: American Psychiatric Press, Inc.).

Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suñer, D., Cigudosa, J.C., Urioste, M., Benitez, J., *et al*. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. Proc. Natl. Acad. Sci. U. S. A. *102*, 10604–10609.

Gavin, D.P., and Akbarian, S. (2012). Epigenetic and post-transcriptional dysregulation of gene expression in schizophrenia and related disease. Neurobiol. Dis. *46*, 255–262.

Gavin, D.P., and Sharma, R.P. (2010). Histone modifications, DNA methylation, and schizophrenia. Neurosci. Biobehav. Rev. *34*, 882–888.

Hanninen, K., Katila, H., Saarela, M., Rontu, R., Mattila, K.M., Fan, M., Hurme, M., and Lehtimaki, T. (2008). Interleukin-1 beta gene polymorphism and its interactions with neuregulin-1 gene polymorphism are associated with schizophrenia. Eur. Arch. Psychiatry Clin. Neurosci. *258*, 10–15.

Kano, S., Colantuoni, C., Han, F., Zhou, Z., Yuan, Q., Wilson, A., Takayanagi, Y., Lee, Y., Rapoport, J., Eaton, W., *et al*. (2013). Genome-wide profiling of multiple histone methylations in olfactory cells: further implications for cellular susceptibility to oxidative stress in schizophrenia. Mol. Psychiatry *18*, 740–742.

Kinoshita, M., Numata, S., Tajima, A., Shimodera, S., Ono, S., Imamura, A., Iga, J., Watanabe, S., Kikuchi, K., Kubo, H., *et al*. (2013). DNA methylation signatures of peripheral leukocytes in schizophrenia. Neuromolecular Med. *15*, 95–101.

Kishore, S., Khanna, A., Zhang, Z., Hui, J., Balwierz, P.J., Stefan, M., Beach, C., Nicholls, R.D., Zavolan, M., and Stamm, S. (2010). The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. Hum. Mol. Genet. *19*, 1153–1164.

Lai, J.H., Zhu, Y.S., Huo, Z.H., Sun, R.F., Yu, B., Wang, Y.P., Chai, Z.Q., and Li, S.B. (2010). Association study of polymorphisms in the promoter region of DRD4 with schizophrenia, depression, and heroin addiction. Brain Res. *1359*, 227–232.

Leung, K.N., Vallero, R.O., DuBose, A.J., Resnick, J.L., and LaSalle, J.M. (2009). Imprinting regulates mammalian snoRNA-encoding chromatin decondensation and neuronal nucleolar size. Hum. Mol. Genet. *18*, 4227–4238.

Levesque, M.L., Casey, K.F., Szyf, M., Ismaylova, E., Ly, V., Verner, M.P., Suderman, M., Brendgen, M., Vitaro, F., Dionne, G., *et al*. (2014). Genome-wide DNA methylation variability in adolescent monozygotic twins followed since birth. Epigenetics *9*, 1410–1421.

Ling, C., and Groop, L. (2009). Epigenetics: a molecular link between environmental factors and type 2 diabetes. Diabetes *58*, 2718–2725.

Liu, Y., Aryee, M.J., Padyukov, L., Fallin, M.D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M., *et al*. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat. Biotechnol. *31*, 142–147.

Maiti, S., Kumar, K.H.B.G., Castellani, C.A., O'Reilly, R., and Singh, S.M. (2011). Ontogenetic *de novo* copy number variations (CNVs) as a source of genetic individuality: studies on two families with MZD twins for schizophrenia. PLoS One *6*, e17125.

Manikkam, M., Guerrero-Bosagna, C., Tracey, R., Haque, M.M., and Skinner, M.K. (2012). Transgenerational actions of environmental compounds on reproductive disease and identification of epigenetic biomarkers of ancestral exposures. PLoS One *7*, e31901.

Melka, M.G., Laufer, B.I., McDonald, P., Castellani, C.A., Rajakumar, N., O'Reilly, R., and Singh, S.M. (2014). The effects of olanzapine on genome-wide DNA methylation in the hippocampus and cerebellum. Clin. Epigenetics *6*, 1.

Mill, J., Tang, T., Kaminsky, Z., Khare, T., Yazdanpanah, S., Bouchard, L., Jia, P., Assadzadeh, A., Flanagan, J., Schumacher, A., *et al*. (2008). Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. Am. J. Hum. Genet. *82*, 696–711.

Minoretti, P., Politi, P., Coen, E., Di Vito, C., Bertona, M., Bianchi, M., and Emanuele, E. (2006). The T393C polymorphism of the GNAS1 gene is associated with deficit schizophrenia in an Italian population sample. Neurosci. Lett. *397*, 159–163.

Murrell, A. (2014). Cross-talk between imprinted loci in Prader-Willi syndrome. Nat. Genet. *46*, 528–530.

Ni, X., Trakalo, J., Valente, J., Azevedo, M.H., Pato, M.T., Pato, C.N., and Kennedy, J.L. (2005). Human p53 tumor suppressor gene (TP53) and schizophrenia: case-control and family studies. Neurosci. Lett. *388*, 173–178.

Ollikainen, M., Smith, K.R., Joo, E.J.-H., Ng, H.K., Andronikos, R., Novakovic, B., Abdul Aziz, N.K., Carlin, J.B., Morley, R., Saffery, R., *et al.* (2010). DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. Hum. Mol. Genet. *19*, 4176–4188.

Paul-Samojedny, M., Owczarek, A., Suchanek, R., Kowalczyk, M., Fila-Danilow, A., Borkowska, P., Kucia, K., and Kowalski, J. (2011). Association study of interferon gamma (IFN-gamma) +874T/A gene polymorphism in patients with paranoid schizophrenia. J. Mol. Neurosci. *43*, 309–315.

Petronis, A. (2010). Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature *465*, 721–727.

Petronis, A., Gottesman, I.I., Kan, P., Kennedy, J.L., Basile, V.S., Paterson, A.D., and Popendikyte, V. (2003). Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? Schizophr. Bull. *29*, 169–178.

Powell, W.T., Coulson, R.L., Crary, F.K., Wong, S.S., Ach, R.A., Tsang, P., Alice Yamada, N., Yasui, D.H., and Lasalle, J.M. (2013a). A Prader-Willi locus lncRNA cloud modulates diurnal genes and energy expenditure. Hum. Mol. Genet. *22*, 4318–4328.

Powell, W.T., Coulson, R.L., Gonzales, M.L., Crary, F.K., Wong, S.S., Adams, S., Ach, R.A., Tsang, P., Yamada, N.A., Yasui, D.H., *et al.* (2013b). R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. Proc. Natl. Acad. Sci. U. S. A. *110*, 13938–13943.

Runte, M., Huttenhofer, A., Gross, S., Kiefmann, M., Horsthemke, B., and Buiting, K. (2001). The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. Hum. Mol. Genet. *10*, 2687–2700.

Saetre, P., Emilsson, L., Axelsson, E., Kreuger, J., Lindholm, E., and Jazin, E. (2007). Inflammation-related genes up-regulated in schizophrenia brains. BMC Psychiatry *7*, 46.

Schmitt, A., Leonardi-Essmann, F., Durrenberger, P.F., Parlapani, E., Schneider-Axmann, T., Spanagel, R., Arzberger, T., Kretzschmar, H., Herrera-Marschitz, M., Gruber, O., *et al*. (2011). Regulation of immune-modulatory genes in left superior temporal cortex of schizophrenia patients: a genome-wide microarray study. World J. Biol. Psychiatry *12*, 201–215.

Sharma, R.P., Grayson, D.R., and Gavin, D.P. (2008). Histone deactylase 1 expression is increased in the prefrontal cortex of schizophrenia subjects: analysis of the National Brain Databank microarray collection. Schizophr. Res. *98*, 111–117.

Shi, J., Levinson, D.F., Duan, J., Sanders, A.R., Zheng, Y., Pe'er, I., Dudbridge, F., Holmans, P.A., Whittemore, A.S., Mowry, B.J., *et al*. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature. 460, 753-7.

Singh, S.M., Murphy, B., and O'Reilly, R. (2002). Epigenetic contributors to the discordance of monozygotic twins. Clin. Genet. *62*, 97–103.

Soderlund, J., Schroder, J., Nordin, C., Samuelsson, M., Walther-Jallow, L., Karlsson, H., Erhardt, S., and Engberg, G. (2009). Activation of brain interleukin-1beta in schizophrenia. Mol. Psychiatry *14*, 1069–1071.

Stelzer, Y., Sagi, I., Yanuka, O., Eiges, R., and Benvenisty, N. (2014). The noncoding RNA IPW regulates the imprinted DLK1-DIO3 locus in an induced pluripotent stem cell model of Prader-Willi syndrome. Nat. Genet. *46*, 551–557.

Sun, Y., Hu, D., Liang, J., Bao, Y.-P., Meng, S.-Q., Lu, L., and Shi, J. (2015). Association between variants of zinc finger genes and psychiatric disorders: systematic review and meta-analysis. Schizophr. Res. *162*, 124–137.

Wockner, L.F., Noble, E.P., Lawford, B.R., Young, R.M., Morris, C.P., Whitehall, V.L., and Voisey, J. (2014). Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients. Transl. Psychiatry *4*, e339.

Wong, C.C., Caspi, A., Williams, B., Craig, I.W., Houts, R., Ambler, A., Moffitt, T.E., and Mill, J. (2010). A longitudinal study of epigenetic variation in twins. Epigenetics *5*, 516–526.

Xu, J., He, G., Zhu, J., Zhou, X., St Clair, D., Wang, T., Xiang, Y., Zhao, Q., Xing, Q., Liu, Y., *et al*. (2014). Prenatal nutritional deficiency reprogrammed postnatal gene expression in mammal brains: implications for schizophrenia. Int. J. Neuropsychopharmacol. *18*, 10.

A version of this chapter has been submitted for publication in *Schizophrenia Research*

Chapter 5 - Integration of Genome-Wide DNA Sequence and DNA Methylation Changes in Two Unrelated Monozygotic Twin Pairs Discordant for Schizophrenia

5.0 Overview

Most genetic disorders are associated with aberrant functioning of genes and pathways that may be caused by a variety of mechanisms including sequence changes and aberrant expression caused by epigenetic change (which may arise from environmental factors). In this chapter, I have focused on genes that are affected by sequence changes between co-twins together with DNA methylation differences between co-twins (the most well studied epigenetic alteration). This approach focuses on genes that are variable and pathways that are affected by multiple mechanisms in the patient as compared to her unaffected monozygotic twin. The goal is to identify highly variable genes and pathways that may play a significant role in the development of schizophrenia in a specific patient.

The dataset for this analysis uses the gene lists from the Affymetrix Human SNP 6.0 analysis (Chapter 2), the Complete Genomics Sequencing analysis (Chapter 3) and the MeDIP Microarray analysis (Chapter 4). It includes any gene that showed at least one sequence variant (in the case of the Affymetrix and Complete Genomics data) and at least one differentially methylated region (in the case of the MeDIP data) that was not shared with its co-twin. It has yielded patient-specific genes that differ between the affected twin and her unaffected matched monozygotic twin. This list of genes was used to model the development of schizophrenia in individual patients. This approach has become plausible with our ability to assess genetic and epigenetic variations in individuals, both completely and comprehensively.

5.1 Introduction

Schizophrenia is a complex mental disorder with high heritability (80%), extensive genetic heterogeneity, environmental contributions and only 50% concordance in monozygotic (MZD) twins. The disappointing success in search of the causations of schizophrenia may be due to the limited scope of current lines of investigation that evaluate genetic, epigenetic or environmental contributors individually. They are not suited to fully unravel the complex interaction between DNA sequence as well as their epigenetic modifications. Integration across multiple “omic” datasets is now becoming an informative approach for unraveling complex disorders. In particular, evidence is emerging that epigenetic states may serve to directly mediate the relationship between certain genetic polymorphisms and phenotype (McVicker *et al*., 2013). The proportion of inter-individual variation in methylomes that may be driven by genotype as well as environment is currently unknown. What is known is that DNA methylation plays a critical role in the regulation of gene expression (Razin and Kantor, 2005). DNA methylation is one of the most well investigated epigenetic modifications and is involved in a number of critical biological processes including embryonic development, X-inactivation and imprinting.

Individual differences in DNA methylation have been correlated with DNA sequence polymorphism, labeled as Methylation Quantitative Trait Loci (methQTLs) (Teh *et al*., 2014). As might be expected, genetic variants at CpG sites are able to alter methylation status at that site (Bell *et al*., 2011; Gertz *et al*., 2011; Gibbs *et al*., 2010; Hellman and Chess, 2010). In addition, genotype at one SNV may affect methylation status at multiple neighbouring CpGs (Eckhardt *et al*., 2006; Zhi *et al*., 2013). Accordingly, in humans, inter-individual variation in DNA methylation could be a consequence of nucleotide polymorphism influencing methylation of cytosines, both directly and indirectly (Liu *et al*., 2013). In fact, methQTLs in adult adipose tissue have been shown to correlate with 19% of observed variance in methylation levels (Grundberg *et al*., 2013).

DNA variations leading to both methylation and expression variation at the same locus appear relatively rare but are reproducible findings (Wagner *et al.*, 2014). For example, Patel *et al* (2013) show that candidate SNP and CpG loci with marginal associations in GWAS and EWAS (Epigenome-Wide Association Studies) are correlated with type 2 diabetes (Patel *et al.*, 2013). Also, a SNP within the IL4R gene combined with methylation at a CpG site within the same gene is predictive of childhood asthma risk (Soto-Ramírez *et al.*, 2013). Further, widespread relationships exist between DNA methylation and gene expression especially in developmentally significant genes, including HOX clusters (Wagner *et al.*, 2014). Additionally, recent studies describe evidence for gene by environment interactions on DNA methylation (Yousefi *et al.*, 2013). This evidence is consistent with the emerging view that genotype can determine the degree of environmentally induced phenotypic plasticity via allele specific variation. In humans, where manipulative experiments are out of the question, a logical approach is the use of monozygotic twins.

Combined with our ability to identify almost all genetic variation per individual, MZD twins provide an exceptional opportunity to assess patient specific changes that may account for the disease phenotype. A combined analysis of genetic and epigenetic changes on the same twin pairs is expected to provide a more effective approach for two reasons. First, it is now possible to generate relatively reliable complete genome sequences as well as promoter methylation states on an individual level and second, the unaffected twin that originated from the same zygote provides a near perfect genetic match for contrast and comparison. Here, I report on the joint analysis of sequence variations including structural changes (Complete Genomics Sequencing), and genome-wide DNA methylation (NimbleGen Human DNA Methylation Microarray) in two families with monozygotic twins discordant for schizophrenia. Monozygotic twins showed differences in DNA sequence as well as DNA methylation and some of these differences were co-localized. The results allow

consideration of the interaction of DNA methylation and sequence changes in schizophrenia.

Specifically, results on family 1 show that 58 genes differ in DNA sequence as well as promoter methylation in a schizophrenia-affected twin as compared to her healthy co-twin. The corresponding number for family 2 was 13. The two lists are over represented by neuronal genes and include a number of known schizophrenia candidate genes and drug targets. The results argue that changes in multiple genes via co-localized genetic and epigenetic alteration contribute to a liability threshold that is necessary for development of schizophrenia. This novel hypothesis, although logical, remains to be validated.

5.2 Methods

The experiments performed received approval by the University of Western Ontario's Committee on research involving human subjects. The families in the study provided written informed consent for participation. Once again, the capacity for consent was ensured using three measures 1) Schizophrenic patients gave consent only during a "normal" phase (no psychosis present), 2) Both twins of the twin pair were present and gave consent at the same time (the normal twin and their Schizophrenic sibling), and, 3) If Dr. O'Reilly felt that the capacity to consent was compromised, the patients were not included in the study. They were interviewed and clinically assessed by a single senior Psychiatrist (R. O'Reilly) using the SCID-I and SCID-II. A second senior psychiatrist independently reviewed videotapes of the structured interviews of the twins and confirmed the diagnoses.

The two discordant twin pairs in this study were female monozygotic twins. The twins from Family 1 were Afro-American females aged 53 and the twins from Family 2 were Caucasian females aged 43. The affected member of twin pair 1 was diagnosed with schizophrenia at age 22 and the affected member of twin pair 2 was diagnosed with schizoaffective disorder at age 27. The twins and their parents contributed whole blood samples for DNA isolation. DNA was extracted

from whole blood using the PerfectPure DNA Blood Kit (Gaithersburg, MD), following the manufacturer's protocol. Zygosity was confirmed by Affymetrix 6.0 microarray and specifically using the Affymetrix Genotyping Console 4.0 concordance feature.

As fully described in Chapter 2B, whole genome microarray analysis using the Affymetrix Genome-Wide Human SNP Array 6.0 was performed at the London Regional Genomics Centre (LRGC) following the manufacturer's protocol. For downstream analysis of CEL files, Affymetrix Genotyping Console 4.1.1 (A), Partek Genomics Suite (P), and PennCNV (p) were used. Only variants that were greater than 1 kb in size were classified as CNVs for the purposes of this study and only those identified by all three software programs in the same individual were included in subsequent analysis. Calls were merged if they were adjacent (gap ≤ 20% and of the same gain/loss state) and overlapping calls were identified as such both between programs and between subjects using a 50% reciprocal overlap rule.

As fully described in Chapter 3, the genome sequence of the twin pairs was generated at Complete Genomics Inc. (Mountain View, CA). The sequences met the criteria of high accuracy (99.999%) and were considered suitable for identification of rare variants including somatic mutations as described by Drmanac *et al* (Drmanac *et al*., 2010). These variants included single nucleotide variants (SNVs), indels and block substitutions as well as larger variants classified as Copy Number Variants (CNVs) and Structural Variants (SVs) that were called in comparison to reference sequence (NCBI Build 37/hg19). A read depth of 50 and a call quality of 100 (calculated by complete genomics and based on a phred scale) were chosen as parameters for initial variant filtering.

As fully described in Chapter 4, I assessed genome-wide methylation in the two twin pairs using the NimbleGen Methylation Promoter Microarray**.** The NimbleGen Human DNA Methylation 720k CpG Island Plus RefSeq Promoter Microarray is a multiplex slide with 3 identical arrays per slide. Each Roche Nimblegen Inc (Madison, MI) array covers 27,728 annotated CpG islands as well

as 22,532 promoters of the RefSeq genes derived from the UCSC RefFlat files. Briefly, the pair files were analyzed with the tiling workflow in Partek Genomics Suite version 6.6 (St. Louis, Missouri). Genes that overlapped differentially methylated regions were annotated.

5.2.4 Annotation and candidate gene analysis

This analysis is based on within-pair differences between two pairs of monozygotic twins discordant for schizophrenia. It included a DNA microarray, complete genome sequence features as well as a genome-wide DNA promoter methylation microarray. The results identified gene specific sequence differences that are associated with corresponding differences in DNA methylation. I have annotated the genes identified with two features: schizophrenia related candidate genes published in the literature and gene ontology.

First, OMIM (omim.org) and SZGene (szgene.com) databases were chosen as the starting point for this analysis. Second, three primary articles were chosen on the basis of three main criteria: 1. They were published in a reputable journal (impact factor >15), 2. They included large sample sizes (>2,500 cases) and 3. They were published within the last six years. It included three GWAS studies (Ayalew *et al.*, 2012; Ripke *et al.*, 2014; Stefansson *et al.*, 2009). Genes overlapping in at least two of the five literature/database sources were compiled into a list of candidate genes to be evaluated in this analysis (34 genes).

Second, I used Ingenuity Variant Analysis (QIAGEN; Redwood City, CA) to annotate the gene lists with links to schizophrenia or neurological disease and related functions. Genes had to have a direct connection to these phenotypes (no hops allowed). Genes were also annotated with direction of methylation, imprinting database information (http://www.geneimprint.com/ and http://igc.otago.ac.nz/), variation type, position, breakpoints, reference calls, gene region identity of variant, SIFT functional prediction, PolyPhen-2 functional prediction, call quality, read depth, dbSNP entry, 1000 genomes frequency and complete genomics public genomes frequency.

5.3 Results

5.3.1 Twin Pair 1

Table 5.1 presents a list of 58 genes that showed changes in DNA sequence as well as DNA methylation between the members of twin pair 1 (Figure 5.1). The observed sequence differences include insertions, deletions, substitutions, single nucleotide variants (SNVs) and tandem duplications, most of which are associated with intronic or intergenic sequences. Of special interest are sequence differences in the *FAM120B* and *ZNF717* genes that are located in exonic regions. Also, *OR6B3* has a sequence difference in the promoter region and *ZNF717* and *AHRR* show differences in the 3'UTR regions. Additionally, SNV and substitutions were observed in the 5' UTR region of the *HNRNPCL1* gene and six sequence differences were found in the exonic regions of the *ZNF717* gene that were characterized as missense mutations (Appendix H).

As shown in Table 5.1, a number of genes in this list are related to drug response in schizophrenia or have been previously implicated in psychosis. These include *GABRB3, OPRD1, RXRA* and *SCN5A*. Besides sequence changes, these genes also showed decreased methylation in their promoter regions. Specifically, *GABRB3* and *OPRD1* harbor insertions, *RXRA* has a multi base-pair substitution and *SCN5A* has a single nucleotide variant, all of which are intronic. Yet another feature of the 58 genes is that 30 of them have an increase and 28 of them have a decrease in methylation in their promoter region in the affected twin as compared to their unaffected co-twin (Table 5.1). Interestingly, *ANKLE2, SNORD115-29* (an imprinted gene) and *TSNARE1* are differentially methylated in both the mother and father (Appendix I). Table 5.1 also shows that among the 30 genes that had increased methylation in the patient, *CACNA1B, DIP2C* and *RASA3* are also differentially methylated in the father (Appendix I). Next, I used the 58 genes in Ingenuity Pathway analysis (Table 5.2). The results show that among the most significant findings are '*PPAR* signaling' (p=2.52E-03), 'neurological diseases' (p=2.04E-05 - 4.71E-02) and 'psychological disorders' (p=2.77E-05 - 4.71E-02). Interestingly, they affected a

number of molecular and cellular pathways including ‘cell-to-cell signaling and interaction’ (*p*=2.55E-03 - 3.26E-02), ‘molecular transport’ (*p*=2.55E-03 - 4.13E-02) and ‘cell cycle’ (*p*=3.01E-03 - 6.01E-03).

Among the genes affected by sequence variations in Family 1, eight of them have been implicated in schizophrenia (Figure 5.1). The specific features of the mutations associated with the eight genes are detailed in Table 5.3. It shows that four *de novo* events affect *DISC1*, all confined to introns, of which three (dbSNP: rs821635, rs116832039 and rs200784958) represent polymorphisms and one represents a rare event (Chr 1: 231868731). Six of the other schizophrenia gene variations also differ in common polymorphisms and may or may not be relevant to disease (*DRD3, GRIA1, GRIN2B, DTNBP1, DGKI,* and *NLGN4X*). The remaining difference is a deletion in the intronic region of *FXR1* (Chr 3:180657418) that is a rare event and may contribute to disease discordance (Table 5.3).

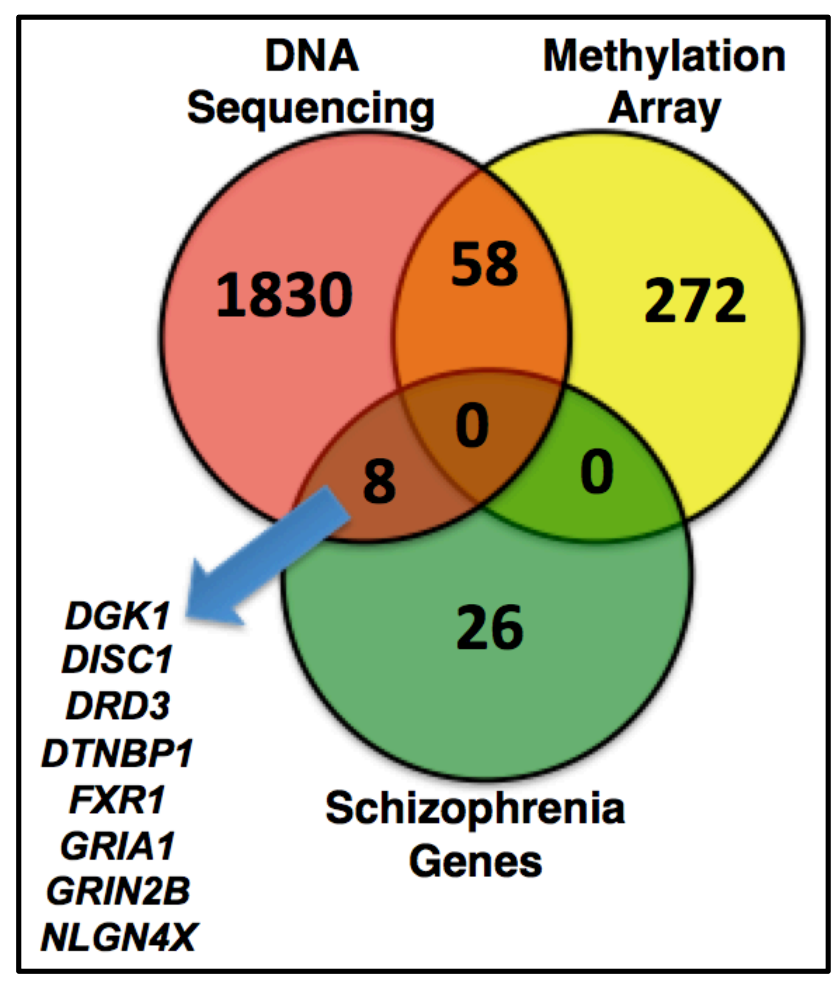


Figure 5.1. Differences between the MZD pair in Family 1 in sequence changes (1896) and changes in DNA methylation (330).

Table 5.1. Genes (58) affected by *de novo* methylation and *de novo* sequence changes in Family 1.

Gene	Neurological Function (from IVA)	Methylation Result	NGS Variation Type	NGS Gene Region Affected
ABCG8	No known neurological function	Decrease	SNV	Intronic
ACSM5	Mutant human ACSM5 gene is associated with CTCAE grade 4 astrocytoma in human brain.	Increase	Insertion	Intronic
ADARB2	Mutant human ADARB2 gene (rs7073579) is observed with sleepiness in human (unknown geographic location). Mutant human ADAR3 [ADARB2] protein is observed with early-onset obsessive-compulsive disorder in humans.	Increase	SNV	Intronic
AHRR	Mutant human AHRR gene (germline c.505C>G) is observed with autism spectrum disorder in humans.	Increase	Insertion	Intronic
ANKLE2	No known neurological function	Increase	Substitution	Intronic
ATP6V1E1	Downregulation of human ATP6V1E1 mRNA in substantia nigra from human is associated with Parkinson's disease in human.	Decrease	Substitution	Intronic
BANP	Upregulation of human BANP mRNA in vastus lateralis muscle is associated with Huntington's disease in human.	Increase	Deletion	Intronic
C7orf50	No known neurological function	Increase	SNV	Intronic
CACNA1B	Mutant human CACNA1B gene (germline c.5199G>A) is observed with schizophrenia in human.	Increase	SNV	Intronic
CAMTA1	Downregulation of human CAMTA1 mRNA in peritumoral white matter is associated with glioblastoma in human. Mutant human CAMTA1 gene is associated with nonprogressive cerebellar ataxia with mental retardation.	Decrease	SNV	Intronic
CBFA2T3	No known neurological function	Decrease	SNV	Intronic
CD2AP	Mutant human CD2AP gene (SNP substitution mutation (rs9349407)) is associated with Alzheimer's disease in human.	Decrease	Deletion	Intronic
CLN6	Change of function germline mutant human CLN6 gene is observed with childhood-onset recessive variant late-infantile neuronal ceroid-lipofuscinosis in human.	Decrease	Insertion	Intronic
CNPY1	No known neurological function	Increase	Substitution	Intronic
COL20A1	No known neurological function	Decrease	Deletion	Intronic

CSMD2	Mutant human CSMD2 gene (germline c.6329A>G) is observed with schizophrenia in humans.	Decrease	Deletion	Intronic
DFFB	No known neurological function	Decrease	Tandem Duplication	N/A
DIP2C	Mutant human DIP2C gene (germline c.1969dupG) is observed with autism in human. Upregulation of human KIAA0934 [DIP2C] mRNA in Brodmann's area 10 from prefrontal cortex is associated with major depression in adult human. Mutant human DIP2C protein (rs10904051) is observed with early-onset obsessive-compulsive disorder in human.	Increase	Substitution	Intronic
DVL1	No known neurological function	Decrease	Tandem Duplication	N/A
EXOC2	No known neurological function	Increase	Deletion	Intronic
FAM120B	Mutant human FAM120B gene (germline c.1178C>G (rs200873057)) is observed with autism spectrum disorder in human.	Increase	SNV	Exonic; Intronic
GABRB3	Olanzapine, an antagonist of human GABRB3 protein -I for the treatment of schizophrenia in human. Muscimol, an agonist of human GABRB3 protein, has been approved for the treatment of schizophrenia.	Decrease	Insertion	Intronic
GRID2IP	Downregulation of mouse Delphilin [Grid2ip] mRNA in cerebellum is associated with purkinje cell degeneration in mouse.	Decrease	SNV	Intronic
HMHA1	Mutant human HMHA1 gene (germline c.1577T>C) is observed with autism in human.	Decrease	Deletion	Intronic
HNF1A	Loss-of-function change of function heterozygous germline mutant human HNF1A protein (p.M626K) is observed with mild mental retardation in humans. Germline mutant human HNF1A protein (p.G574S) is observed with diabetic neuropathy in humans.	Decrease	Substitution	Intronic
HNRNPCP5	Mutant human HNRNPCL1 gene (SNP substitution mutation, allelic variations: A/G (rs1856638)) is associated with bipolar affective disorder in human.	Increase	Substitution; SNV	Intronic; 5'UTR
IL1RAPL2	In mouse, homozygous mutant mouse Il1r gene(s) (knockout) decreases experimental autoimmune encephalomyelitis in mouse.	Increase	SNV	Intronic

KATNAL2	Mutant human KATNAL2 gene (germline c.510+1G>A) is observed with autism in human.	Increase	SNV	Intronic
KAZN	Downregulation of mouse W84167 [Kazn] mRNA in striatum from brain is associated with Huntington's disease in 6 week-old r6/2 mouse.	Increase	SNV	Intronic
KIAA1751	No known neurological function	Increase	Tandem Duplication	N/A
MAD1L1	No known neurological function	Increase	SNV	Intronic
MYT1L	Mutant human MYT1L gene (germline c.2636+1G>A) is observed with intellectual disability in human. Downregulation of human MYT1L mRNA in caudate is associated with Huntington's disease in human.	Increase	Substitution; Deletion	Intronic; N/A
NFATC1	No known neurological function	Increase	Substitution	Intronic
NGFR	Mutant human NGFR gene (germline c.364T>C) is observed with intellectual disability in human. Mutant human NGFR gene (common germline c.614C>T (rs2072446)) is associated with familial Alzheimer's disease in human.	Decrease	Substitution	Intronic
OBSCN	Mutant human OBSCN gene (germline c.6839G>A (rs562143677)) is observed with schizophrenia in human.	Decrease	Insertion	Intronic
OCA2	No known neurological function	Decrease	Deletion	Intronic
OPRD1	Naltrexone, an antagonist of human OPRD1 protein, has been approved for the treatment of schizophrenia. Mutant human OPRD1 gene (allelic variations: C/T (rs678849)) is associated with amyotrophic lateral sclerosis in humans.	Decrease	Insertion	Intronic
OR6B3	No known neurological function	Increase	SNV	Promoter
PPP1R13L	No known neurological function	Decrease	SNV	Intronic
PRKAR1B	Blockade of active mouse Pka complex(es) prevents loss of cholinergic fibers in lesioned ipsilateral region cerebral cortex from mouse missing a ovary that is decreased by injection of E2 [beta-estradiol].	Decrease	Insertion; Tandem Duplication	Intronic; N/A
PTPRN2	Downregulation of human PTPRN2 mRNA in clonal cultured T lymphocytes is associated with Rett syndrome in humans.	Increase	SNV	Intronic
RAB40C	No known neurological function	Increase	SNV	Intronic
RASA3	No known neurological function	Increase	SNV	Intronic
RPH3AL	No known neurological function	Increase	SNV	Intronic

RPTOR	In mouse oligodendrocytes, homozygous mutant mouse Raptor [Rptor] gene increases dysmyelination in mouse spinal cord dorsal column. In neuronal progenitor cells from 17.5-19.5 day-old embryonic mouse of CNS, homozygous mutant mouse Rptor gene (knockout) increases microencephaly in 17.5-19.5 day-old embryonic mouse.	Decrease	SNV	Intronic
RXRA	Bexarotene, an activator of human RXRA protein, is in Phase III clinical trial for the treatment of schizophrenia in human (Bexarotene). Upregulation of human Rxr protein(s) in brain from human is associated with Alzheimer's disease in human.	Decrease	Substitution	Intronic
SCN5A	Gain-of-function germline mutant human SCN5A protein (p.R800L, alternately c.2399G>T) is observed with schizophrenia in human (Caucasian). Oxcarbazepine, an inhibitor of human SCN5A protein, is in Phase 2/Phase 3 clinical trial for the treatment of schizophrenia in human.	Decrease	SNV	Intronic
SHOX	No known neurological function	Decrease	SNV	Intronic
SLC4A3	In 129S1/Sv * 129X1/SvJ * C57BL/6 mouse, homozygous mutant mouse Slc4a3 gene (allele Slc4a3tm1Cahb/Slc4a3tm1Cahb) (knockout) increases seizures in mice.	Decrease	Deletion	Intronic
SLC6A3	Mutant human DAT [SLC6A3] gene is associated with schizophrenia in human. Modafinil, an inhibitor of human SLC6A3 protein, is in Phase IV clinical trial for the treatment of schizophrenia in human. Sibutramine, an inhibitor of human SLC6A3 protein, is in Phase IV clinical trial for the treatment of schizophrenia in human.	Increase	Insertion	Intronic
SNORD115-29	No known neurological function	Increase	Tandem Duplication	N/A
SNX29	No known neurological function	Increase	Insertion	Intronic
STXBP2	No known neurological function	Decrease	SNV	Intronic
TMEM255B	No known neurological function	Decrease	SNV	Intronic

TRAPPC9	Mutant human TRAPPC9 gene (germline c.3203G>A) is observed with schizophrenia in human. Mutant human TRAPPC9 gene (germline c.1708C>T (rs267607137)) increases the risk of autosomal recessive nonsyndromic mental retardation in human.	Decrease	Substitution	Intronic
TSNARE1	No known neurological function	Increase	Insertion	Intronic
TSSC1	No known neurological function	Increase	SNV	Intronic
ZNF717	No known neurological function	Increase	SNV; Substitution; Deletion; Insertion	Exonic; Intronic

Table 5.2. Ingenuity Pathway Analysis results showing the most significant pathways and networks for co-localized sequence variations and promoter methylation differences in Family 1.

Diseases and disorders		**# of molecules**
Neurological Disease	2.04E-05 - 4.71E-02	18
Skeletal and Muscular Disorders	2.76E-05 - 3.91E-02	15
Psychological Disorders	2.77E-05 - 4.71E-02	12
Molecular and Cellular Functions		**# of molecules**
Cell-To-Cell Signaling and Interaction	2.55E-03 - 3.26E-02	5
Molecular Transport	2.55E-03 - 4.13E-02	10
Small Molecule Biochemistry	2.55E-03 - 4.42E-02	8
Cell Cycle	3.01E-03 - 6.01E-03	2
Cell Morphology	3.01E-03 - 1.50E-02	2
Top networks	**Score**	
Tissue morphology, cell death and survival	7	
Neurological diseases	2	
Canonical pathways	***p*-value**	**Ratio**
Hepatic Cholestasis	7.76E-06	6/158 (0.038)
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis	4.54E-04	5/214 (0.023)
LXR/RXR Activation	4.81E-04	4/121 (0.033)
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	1.69E-03	5/287 (0.017)
PPAR Signaling	2.52E-03	3/90 (0.033)

5.3.2 Twin Pair 2

Figure 5.2 includes a number of genes that differ for sequence changes and methylation differences between the members of twin pair 2. The observed sequence differences included insertions, deletions, substitutions and SNVs. Once again, most of these differences are confined to introns or intergenic regions. In total, 13 genes are affected by both mechanisms; sequence change and DNA methylation. The specific features of the 13 genes are detailed in Table 5.4. It shows that only the *MUC4* and *UBTFL1* genes contained sequence differences in exonic regions and that *AK2* has a sequence difference in the 3'UTR region (Appendix J).

Two of the genes identified in Family 2 have been previously related to schizophrenia. These include *SORBS2* and *RBFOX1* that showed increased methylation in their promoter regions and corresponding *de novo* sequence variations in the affected twin (Table 5.4). Specifically, *SORBS2* harbors a 13 bp insertion and a single base pair substitution; similarly *RBFOX1* harbors 8 sequence variations in intronic regions throughout the gene comprising small sequence deletions, substitutions and single base pair variants (Appendix J).

Yet another feature of the 13 genes is that 11 of them have an increase and 2 of them (*BCOR* and *PLEC*) have a decrease in methylation in their promoter region in the affected twin as compared to their unaffected co-twin (Table 5.4). Interestingly, *BCOR* and *PLEC* were also differentially methylated in both Mom and Dad of twin pair 2 (Appendix K). In addition, one of the 13 genes, *PRDM16*, is an imprinted gene (http://www.geneimprint.com). Ingenuity Pathway Analysis using the 13 genes did not yield any statistically significant pathways or networks of interest, likely due to the small number of genes inputted.

Table 5.3. *De novo* sequence variations in the affected twin in Family 1 which have been previously associated with schizophrenia. (NCBI Build 37/hg19)

Chr	Start	Gene	Ref	Affected Twin	Type	Gene Region	dbSNP ID	1000 Genomes Frequency	CG Genomes Frequency
1	231868731	*DISC1*		TAGGAGTGGTG GTGGTGATGA	Insertion	Intronic			
1	231869328	*DISC1*	G	A	SNV	Intronic	116832039		
1	232006400	*DISC1*		C	Insertion	Intronic	200784958	1.38	0.92
1	232149541	*DISC1*	C	T	SNV	Intronic	821635	38.56	33.33
3	113852521	*DRD3*	T	A	SNV	Intronic	12491384		42.59
3	180657418	*FXR1*	AG		Deletion	Intronic			
5	152929315	*GRIA1*	AGATC	TT	Substitution	Intronic	386693748		23.14
6	15636417	*DTNBP1*		CTCT	Insertion	Intronic	58627855		
7	137140536	*DGKI*	G	A	SNV	Intronic	2351386	62.4	63.88
12	13779752	*GRIN2B*		T	Insertion	Intronic	151115453	86.38	76.85
X	5952756	*NLGN4X*	T	G	SNV	Intronic	10284218		

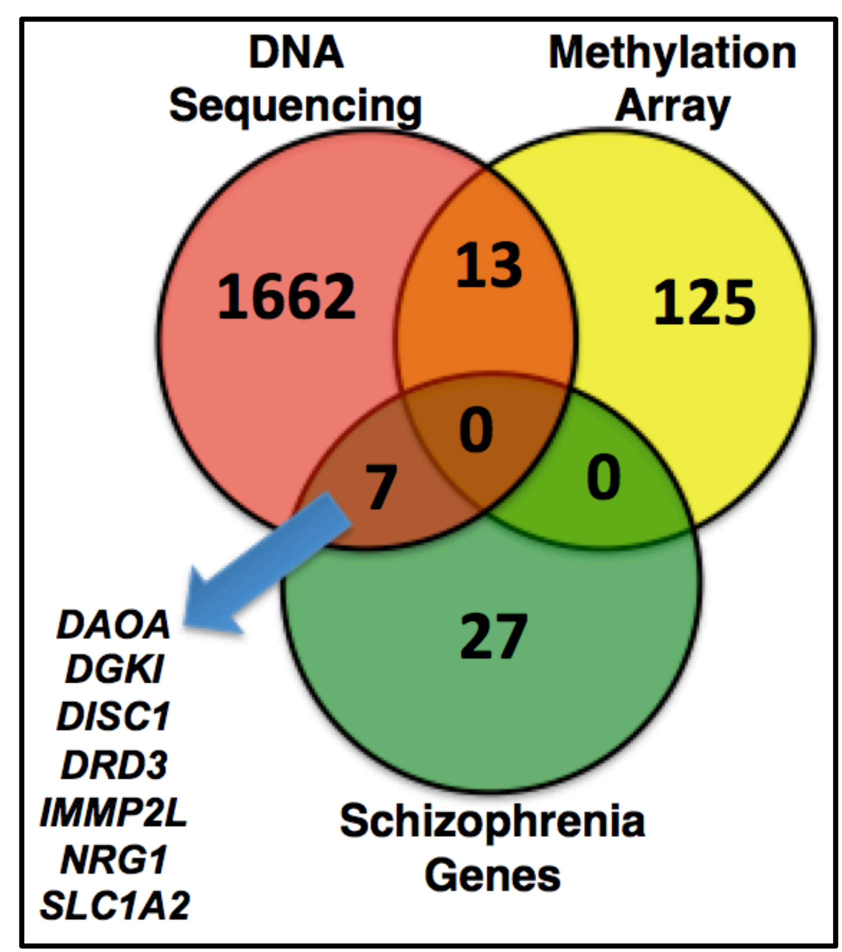


Figure 5.2. Differences between the MZD pair in Family 2 in sequence changes (1682) and changes in DNA methylation (138).

Among the genes affected by sequence variations only in Family 2, seven of them have been previously implicated in schizophrenia (Figure 5.2). The specific features of the mutations associated with the seven genes are listed in Table 5.5. It shows that three *de novo* events affect *SLC1A2*, all confined to introns, of which all have been found in dbSNP and represent polymorphisms (dbSNP: rs4755399, rs4755400, rs386752435). Four of the other variations in schizophrenia genes between affected and unaffected twins of this family also differ in well-known polymorphisms and may or may not be relevant to disease (*DRD3, IMMP2L, NRG1, DISC1*). I note that the remaining differences include two rare changes affecting *DGK1* (Chr 7: 137483119) with a three base-pair intronic substitution, and *DAOA* (Chr 13: 106118100), which has a unique deletion in its promoter region. These rare variations may partially explain disease discordance in this family (Table 5.5).

5.3.3 Twin Pair 1 and Twin Pair 2

The results have also allowed me to assess sharing of changes seen in the two patients from two different MZD twin pairs. This analysis has identified one gene, *PTPRN2*. Here, the two patients harbor unique sequence variants in their intronic sequences as well as increased methylation of the promoter region of this gene.

Table 5.4. Genes (13) affected by *de novo* methylation and *de novo* sequence changes in Family 2.

Gene	Neurological Function (from IVA)	Methylation Result	NGS Variation Type	NGS Gene Region Affected
AK2	Downregulation of human AK2 mRNA in prefrontal cortex is associated with bipolar disorder in human.	Increase	SNV	ncRNA; 3'UTR
BCOR	Somatic mutant human BCOR gene is associated with medulloblastoma in human.	Decrease	Insertion	Intronic
KCNT1	Heterozygous *de novo* mutant human KCNT1 protein (p.R428Q, alternately c.1283G>A) is observed with malignant migrating partial seizures of infancy syndrome in human (France). Germline mutant human KCNT1 gene (c.2800G>A (rs397515403)) increases early infantile epileptic encephalopathy type 14 in human.	Increase	Deletion	Intronic
MUC4	Mutant human MUC4 gene (silent somatic c.6783C>T translating to p.D2261D) is associated with neuroblastoma in human autonomic ganglion (observed in 2 of 238 samples).	Increase	SNV	Intronic; Exonic
PLEC	Germline mutant human PLEC1 [PLEC] protein (frameshift p.E4015Gfs*69, alternately c.12043dupG) is observed with childhood-onset myasthenic syndrome in human (African American).	Decrease	Insertion	Intronic
PRDM16	Mutant human PRDM16 gene (rs2651899) is observed with migraine with aura in human	Increase	Deletion	Intronic
PTPRN2	Downregulation of human PTPRN2 mRNA in clonal cultured T lymphocytes is associated with Rett syndrome in human.	Increase	Substitution; Deletion	Intronic; N/A
RBFOX1	Mutant human A2BP1 [RBFOX1] gene is associated with sporadic schizophrenia in male human. Downregulation of human A2BP1 [RBFOX1] mRNA in caudate nucleus from brain is associated with Huntington's disease in human. Mutant human A2BP1 [RBFOX1] gene is observed with childhood-onset autism in human.	Increase	Deletion	Intronic
RSU1	Downregulation of human RSP-1 [RSU1] mRNA in Brodmann's area 10 from prefrontal cortex is associated with major depression in adult human.	Increase	SNV	Intronic

SETD5	Mutant human SETD5 gene increases the risk of intellectual disability in human. Mutant human SETD5 gene (germline c.1405G>A) is observed with autism in human.	Increase	Deletion	Intronic
SORBS2	Mutant human SORBS2 gene (germline c.2451G>T) is observed with schizophrenia in human.	Increase	Insertion	Intronic
TENM2	Mutant human TENM2 gene is associated with glioma in human brain (observed in 5 of 268 samples). Mutant human TENM2 gene is associated with CTCAE grade 4 astrocytoma in human brain (observed in 6 of 298 samples).	Increase	Insertion	Intronic
UBTFL1	No known neurological function	Increase	SNV	Exonic

Table 5.5. *De novo* sequence variations in the affected twin in Family 2 which have been previously associated with schizophrenia. (NCBI Build 37/hg19)

Chr	Start	Gene	Ref	Affected Twin	Type	Gene Region	dbSNP ID	1000 Genomes Project	CG Genomes Frequency
1	232167072	*DISC1*		AAAGGA	Insertion	Intronic	147764269		7.4
3	113883058	*DRD3*	C	A	SNV	Intronic	111248254	32.13	35.18
7	110651008	*IMMP2L*	G		Deletion	Intronic	59572830	31.67	33.33
7	137483119	*DGK1*	CC	ACA	Substitution	Intronic			
8	32170561	*NRG1*	C	T	SNV	Intronic	4336564	27.06	26.85
11	35425416	*SLC1A2*	T	A	SNV	Intronic	4755399		30.55
11	35425423	*SLC1A2*	T	C	SNV	Intronic	4755400		31.48
11	35425428	*SLC1A2*	TAA	GGC	Substitution	Intronic	386752435		37.97
13	106118100	*DAOA*	GTATCCA		Deletion	Promoter			

5.4 Discussion

The results included in this chapter identify almost all genes that differ between the affected and unaffected members of two monozygotic twin pairs discordant for schizophrenia. These genes harbor sequence differences as well as differences in DNA promoter methylation between MZD twins. The results highlight four insights. First, MZ twins differ in rare DNA sequences as well as DNA methylation caused by *de novo* events. Second, some genes that are affected by differential promoter methylation between MZD twins also harbor various types of sequence variations between the same twins. These genes may be viewed as highly variable in structure and expression, where some of the variation may represent *de novo* events. Third, a number of the changes identified are in known candidate genes for schizophrenia. Therefore, the genes affected in this study may contribute to neuropsychiatric disorders, particularly schizophrenia. Additionally, when the DNA sequence differences are analyzed independently, additional previously identified schizophrenia candidate genes arise. The latter finding may be due to the fact that most candidate genes reported to date have been identified by their differences in DNA sequences. Finally, the comprehensive data on all possible sequence and methylation differences between an affected and unaffected monozygotic twin pair is compatible with a threshold model for this complex disease (Hennah *et al.*, 2009; McGue *et al.*, 1983). Specifically, it may explain the etiology of schizophrenia in a given patient. In the threshold model, any sequence variation or promoter methylation not shared between the patient and her healthy co-twin is labeled as a potential predisposing factor for the disease. An accumulation of predisposing factors and additional factors would lead to the critical threshold of variation necessary for disease manifestation. The results on the two patients in this report are encouraging and support this assertion.

In Family 1, 32 of the 58 genes with co-localized sequence and methylation differences are implicated in neurologically related functions, including schizophrenia. This overrepresentation of neurological genes in this list

is highly suggestive. Specifically, I identified patient-specific pathways, such as '*PPAR* signaling pathway' (*p*=2.52E-03). In addition, I found that 'Neurological Diseases' (*p*=2.04E-05 - 4.71E-02) and 'Psychological Disorders' (*p*=2.77E-05 - 4.71E-02) were two of the top three diseases and disorders identified. Further to the pathways identified, it also includes a number of genes that have been the focus of attention in studies on neurodevelopmental disorders including schizophrenia. In fact, a number of them have been specifically implicated in the features that separate the affected and healthy members of this twin pair. For example, *SNORD115* is differently methylated in this twin pair and also has an observed *de novo* CNV difference that covers this gene. Functionally, *SNORD115* changes the alternative splicing of the serotonin receptor 2C pre-mRNA, a receptor that is known to be involved in mental disorders including autism and schizophrenia (Feinberg *et al*., 2015; Kishore *et al*., 2010). Yet another example is the sequence variations and altered promoter methylations observed in *GABRB3* in this study. This gene is known to be involved in a number of mental disorders including schizophrenia. *De novo* mutations in *GABRB3* have been reported in panic disorders (Hodges *et al*., 2014) and intellectual disability (Hamdan *et al*., 2014). Also, olanzapine, an antagonist of human *GABRB3* protein and pregnenolone, an inhibitor of human *GABRB3,* have been used in the treatment of schizophrenia and bipolar/major depressive disorder, respectively (Mosby, 2003). Additionally, the genomic region that harbors *SNORD115* and *GABRB3* (15q11) has been shown to have strong genetic and epigenetic underpinning for simultaneous presentation of Prader-Willi syndrome (PWS) and psychosis (Krefft *et al*., 2014).

Similarly in Family 1, a *de novo* sequence variation (insertion) was observed in the opioid receptor gene, *OPRD1,* in this study. Naltrexone, an antagonist of human *OPRD1* protein, is a schizophrenia drug (Mosby, 2003) and mutant human *OPRD1* gene was found to be associated with amyotrophic lateral sclerosis in human, which is a fatal neurological disorder that attacks neurons (Lin *et al*., 1995).

Support from the literature is also provided for a number of other systems affected in twin pair 1 including GABAergic signaling molecules (Fujimori and Yoneda, 2004) and oxcarbazepine, an inhibitor of human SCN5A protein that is in clinical trials to treat schizophrenia (http://clinicaltrials.gov/). Moreover, loss-of-function germline mutant human *SCN5A* protein has been implicated in schizophrenia in a Caucasian population (Hu *et al*., 2013). The *RXRA* gene is also a target of Bexarotene, a schizophrenia drug (Xiong *et al*., 2008). Therefore, the disease state of the patient in this study could be hypothesized to reach the disease threshold by some or all of the changes in sequence variations and DNA methylations observed in the patient as compared to her unaffected MZ twin sister.

The conclusions are also compatible with observations on Family 2 that has yielded 13 genes that carry differences in DNA sequences and promoter methylations in the affected as compared to unaffected MZD twin discordant for schizophrenia. Twelve of the 13 genes identified are involved in neurological functions and abnormalities including schizophrenia. Although the number of genes affected is not sufficient to assess them in pathway analysis, gene specific functions further support their potential involvement in schizophrenia. For example, *BCOR,* that carries an insertion and hypermethylation of its promoter in the patient has been shown to form a complex (*BCL6*/*BCOR*/*SIRT1*) which serves as a potent repressor of the SHH pathway in neural development (Tiberi *et al*., 2014). Further, *BCOR* specific class I and II histone deacetylases (HDACs) have been shown to interact with this protein, which suggests a possible link between the two classes of HDACs.

Of particular relevance is Disrupted-in-Schizophrenia-1 (DISC1) that harbors an insertion in the patient of Family 2 as well as an insertion and SNV in the patient of Family 1. *DISC1* is known to be involved in cell proliferation, differentiation, and migration as well as neuronal axon and dendrite outgrowth (Ozeki *et al*., 2003). This gene is also involved in neural development and has an established role in the etiology of schizophrenia (Tohyama *et al*., 2015).

Furthermore, other genes implicated in schizophrenia include *DRD3* (Dai *et al*., 2014) and DAOA (Bass *et al*., 2009; Muller *et al*., 2011; Prata *et al*., 2008; Shi *et al*., 2009).

Interestingly, I found both sequence and methylation changes in the *PTPRN2* gene in both patients. This gene is thought to be a protein tyrosine phosphatase and has been shown to be associated with Rett syndrome, yet another neurodevelopmental disorder (Delgado *et al*., 2006). In addition, rare CNVs related to *ADHD* have been identified to overlap the *PTPRN2* gene (Lionel *et al*., 2011), making this gene a potential contributor to the disease liability for schizophrenia, specifically in these patients.

The identification of CNVs from Affymetrix Human SNP 6.0 arrays did not yield any shared gene regions with the methylation dataset. This is likely due to the reduced genomic coverage of the microarray as compared to DNA sequencing.

5.5 Conclusion

Most research in search of genes causing schizophrenia has used limited DNA sequence differences in a large number of patients matched with unaffected controls from a variety of collections and populations. The results argue for extensive heterogeneity of causation even within a single family making every patient different and any definitive conclusions about this disorder problematic. This research has identified all/almost all changes in the patient in contrast to matched unaffected MZ co-twins. The results have identified a number of genes that differ in DNA sequence as well as promoter methylations. As expected, the list of genes differs between two MZD pairs. They predominantly represent genes involved in neurodevelopment and function. More importantly, a number of genes affected by DNA sequence changes have been implicated in schizophrenia and related disorders. The results confirm uniqueness of individual patients whose etiology could be explained by a threshold model involving multiple genes in this complex disorder. Also, it supports the suggestion that a variety of mechanisms

including genetic, epigenetic and environmental factors could contribute to this threshold and may be co-localized. Although logical, this model remains to be experimentally authenticated.

5.6 References

Ayalew, M., Le-Niculescu, H., Levey, D.F., Jain, N., Changala, B., Patel, S.D., Winiger, E., Breier, A., Shekhar, A., Amdur, R., *et al*. (2012). Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. Mol. Psychiatry *17*, 887–905.

Bass, N.J., Datta, S.R., McQuillin, A., Puri, V., Choudhury, K., Thirumalai, S., Lawrence, J., Quested, D., Pimm, J., Curtis, D., *et al*. (2009). Evidence for the association of the DAOA (G72) gene with schizophrenia and bipolar disorder but not for the association of the DAO gene with schizophrenia. Behav. Brain Funct. *5*, 28.

Bell, J.T., Pai, A.A., Pickrell, J.K., Gaffney, D.J., Pique-Regi, R., Degner, J.F., Gilad, Y., and Pritchard, J.K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol. *12*, R10.

Dai, D., Cheng, J., Zhou, K., Lv, Y., Zhuang, Q., Zheng, R., Zhang, K., Jiang, D., Gao, S., and Duan, S. (2014). Significant association between DRD3 gene body methylation and schizophrenia. Psychiatry Res. *220*, 772–777.

Delgado, I.J., Kim, D.S., Thatcher, K.N., LaSalle, J.M., and Van den Veyver, I.B. (2006). Expression profiling of clonal lymphocyte cell cultures from Rett syndrome patients. BMC Med. Genet. *7*, 61.

Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., *et al*. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science *327*, 78–81.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T. V, Davies, R., Down, T.A., *et al*. (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. Nat. Genet. *38*, 1378–1385.

Feinberg, J.I., Bakulski, K.M., Jaffe, A.E., Tryggvadottir, R., Brown, S.C., Goldman, L.R., Croen, L.A., Hertz-Picciotto, I., Newschaffer, C.J., Daniele Fallin,

M., *et al*. (2015). Paternal sperm DNA methylation associated with early signs of autism risk in an autism-enriched cohort. Int. J. Epidemiol.

Fujimori, S., and Yoneda, Y. (2004). Neuropsychiatric disorders and GABA. Nihon Shinkei Seishin Yakurigaku Zasshi *24*, 265–271.

Gertz, J., Varley, K.E., Reddy, T.E., Bowling, K.M., Pauli, F., Parker, S.L., Kucera, K.S., Willard, H.F., and Myers, R.M. (2011). Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. PLoS Genet. *7*, e1002228.

Gibbs, J.R., van der Brug, M.P., Hernandez, D.G., Traynor, B.J., Nalls, M.A., Lai, S.-L., Arepalli, S., Dillman, A., Rafferty, I.P., Troncoso, J., *et al*. (2010). Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet. *6*, e1000952.

Grundberg, E., Meduri, E., Sandling, J.K., Hedman, A.K., Keildson, S., Buil, A., Busche, S., Yuan, W., Nisbet, J., Sekowska, M., *et al*. (2013). Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. Am. J. Hum. Genet. *93*, 876–890.

Hamdan, F.F., Srour, M., Capo-Chichi, J.M., Daoud, H., Nassif, C., Patry, L., Massicotte, C., Ambalavanan, A., Spiegelman, D., Diallo, O., *et al*. (2014). *De novo* mutations in moderate or severe intellectual disability. PLoS Genet. *10*, e1004772.

Hellman, A., and Chess, A. (2010). Extensive sequence-influenced DNA methylation polymorphism in the human genome. Epigenetics Chromatin *3*, 11.

Hennah, W., Thomson, P., McQuillin, A., Bass, N., Loukola, A., Anjorin, A., Blackwood, D., Curtis, D., Deary, I.J., Harris, S.E., *et al*. (2009). DISC1 association, heterogeneity and interplay in schizophrenia and bipolar disorder. Mol. Psychiatry *14*, 865–873.

Hodges, L.M., Fyer, A.J., Weissman, M.M., Logue, M.W., Haghighi, F., Evgrafov, O., Rotondo, A., Knowles, J.A., and Hamilton, S.P. (2014). Evidence for linkage and association of GABRB3 and GABRA5 to panic disorder. Neuropsychopharmacology *39*, 2423–2431.

Hu, R.M., Tan, B.H., Orland, K.M., Valdivia, C.R., Peterson, A., Pu, J., and Makielski, J.C. (2013). Digenic inheritance novel mutations in SCN5a and SNTA1 increase late I(Na) contributing to LQT syndrome. Am. J. Physiol. Circ. Physiol. *304*, H994–H1001.

Kishore, S., Khanna, A., Zhang, Z., Hui, J., Balwierz, P.J., Stefan, M., Beach, C., Nicholls, R.D., Zavolan, M., and Stamm, S. (2010). The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. Hum. Mol. Genet. *19*, 1153–1164.

Krefft, M., Frydecka, D., Adamowski, T., and Misiak, B. (2014). From Prader-Willi syndrome to psychosis: translating parent-of-origin effects into schizophrenia research. Epigenomics *6*, 677–688.

Lin, Y., Lin, X., Hong, L., Foundling, S., Heinrikson, R.L., Thaisrivongs, S., Leelamanit, W., Raterman, D., Shah, M., and Dunn, B.M. (1995). Effect of point mutations on the kinetics and the inhibition of human immunodeficiency virus type 1 protease: relationship to drug resistance. Biochemistry *34*, 1143–1152.

Lionel, A.C., Crosbie, J., Barbosa, N., Goodale, T., Thiruvahindrapuram, B., Rickaby, J., Gazzellone, M., Carson, A.R., Howe, J.L., Wang, Z., *et al*. (2011). Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. Sci. Transl. Med. *3*, 95ra75.

Liu, Y., Aryee, M.J., Padyukov, L., Fallin, M.D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M., *et al*. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat. Biotechnol. *31*, 142–147.

McGue, M., Gottesman, I.I., and Rao, D.C. (1983). The transmission of schizophrenia under a multifactorial threshold model. Am. J. Hum. Genet. *35*, 1161–1178.

McVicker, G., van de Geijn, B., Degner, J.F., Cain, C.E., Banovich, N.E., Raj, A., Lewellen, N., Myrthil, M., Gilad, Y., and Pritchard, J.K. (2013). Identification of genetic variants that affect histone modifications in human cells. Science *342*, 747–749.

Mosby (2003). Mosby’s Drug Consult 2003: The Comprehensive Reference for Generic and Brand Name Drugs with CDRO.

Muller, D.J., Zai, C.C., Shinkai, T., Strauss, J., and Kennedy, J.L. (2011). Association between the DAOA/G72 gene and bipolar disorder and meta-analyses in bipolar disorder and schizophrenia. Bipolar Disord. *13*, 198–207.

Ozeki, Y., Tomoda, T., Kleiderlein, J., Kamiya, A., Bord, L., Fujii, K., Okawa, M., Yamada, N., Hatten, M.E., Snyder, S.H., *et al*. (2003). Disrupted-in-Schizophrenia-1 (DISC-1): mutant truncation prevents binding to NudE-like (NUDEL) and inhibits neurite outgrowth. Proc. Natl. Acad. Sci. U. S. A. *100*, 289–294.

Patel, C.J., Chen, R., Kodama, K., Ioannidis, J.P.A., and Butte, A.J. (2013). Systematic identification of interaction effects between genome- and environment-wide associations in type 2 diabetes mellitus. Hum. Genet. *132*, 495–508.

Prata, D., Breen, G., Osborne, S., Munro, J., St Clair, D., and Collier, D. (2008). Association of DAO and G72(DAOA)/G30 genes with bipolar affective disorder. Am. J. Med. Genet. B, Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet. *147B*, 914–917.

Razin, A., and Kantor, B. (2005). DNA methylation in epigenetic control of gene expression. Prog. Mol. Subcell. Biol. *38*, 151–167.

Ripke, S., Neale, B.M., Corvin, A., Walters, J.T.R., Farh, K.-H., Holmans, P.A., Lee, P., Bulik-Sullivan, B., Collier, D.A., Huang, H., *et al*. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

Shi, J., Badner, J.A., Gershon, E.S., Chunyu, L., Willour, V.L., and Potash, J.B. (2009). Further evidence for an association of G72/G30 with schizophrenia in Chinese. Schizophr. Res. *107*, 324–326.

Soto-Ramírez, N., Arshad, S.H., Holloway, J.W., Zhang, H., Schauberger, E., Ewart, S., Patil, V., and Karmaus, W. (2013). The interaction of genetic variants and DNA methylation of the interleukin-4 receptor gene increase the risk of asthma at age 18 years. Clin. Epigenetics *5*, 1.

Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietilainen, O.P., Mors, O., Mortensen, P.B., *et al*. (2009). Common variants conferring risk of schizophrenia. Nature.

Teh, A.L., Pan, H., Chen, L., Ong, M.L., Dogra, S., Wong, J., MacIsaac, J.L., Mah, S.M., McEwen, L.M., Saw, S.M., *et al*. (2014). The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. Genome Res. *24*, 1064–1074.

Tiberi, L., Bonnefont, J., van den Ameele, J., Le Bon, S.D., Herpoel, A., Bilheu, A., Baron, B.W., and Vanderhaeghen, P. (2014). A BCL6/BCOR/SIRT1 complex triggers neurogenesis and suppresses medulloblastoma by repressing Sonic Hedgehog signaling. Cancer Cell *26*, 797–812.

Tohyama, M., Miyata, S., Hattori, T., Shimizu, S., and Matsuzaki, S. (2015). Molecular basis of major psychiatric diseases such as schizophrenia and depression. Anat. Sci. Int.

Wagner, J.R., Busche, S., Ge, B., Kwan, T., Pastinen, T., and Blanchette, M. (2014). The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol. *15*, R37.

Xiong, H., Callaghan, D., Jones, A., Walker, D.G., Lue, L.F., Beach, T.G., Sue, L.I., Woulfe, J., Xu, H., Stanimirovic, D.B., *et al.* (2008). Cholesterol retention in Alzheimer's brain is responsible for high beta- and gamma-secretase activities and Abeta production. Neurobiol. Dis. *29*, 422–437.

Yousefi, M., Karmaus, W., Zhang, H., Ewart, S., Arshad, H., and Holloway, J.W. (2013). The methylation of the LEPR/LEPROT genotype at the promoter and body regions influence concentrations of leptin in girls and BMI at age 18 years if their mother smoked during pregnancy. Int. J. Mol. Epidemiol. Genet. *4*, 86–100.

Zhi, D., Aslibekyan, S., Irvin, M.R., Claas, S.A., Borecki, I.B., Ordovas, J.M., Absher, D.M., and Arnett, D.K. (2013). SNPs located at CpG sites modulate genome-epigenome interaction. Epigenetics *8*, 802–806.

Chapter 6 - Discussion

6.0 Overview

The studies presented in this PhD Thesis have identified CNV (Chapter 2), genome sequence (Chapter 3) and methylation (Chapter 4) alterations between sets of monozygotic twins that are discordant for schizophrenia. We have followed the families for over five years since sample collection and established that the twins have remained discordant. The analysis used three technologies: a CNV array, whole genome sequencing and a DNA methylation array. The results have identified genomic variations between twin pairs. Where possible, based on availability of parental data, these variations were established to be *de novo*. A number of exonic regions were affected as well as intergenic and intronic regions. Some of the affected genes and regions have been previously reported to be involved in schizophrenia and others are presented here for the first time. Together, these genes and regions may serve as candidate genes for future studies. Further, the presence of multiple variations with possible links to schizophrenia pathology supports the contention that they may in fact be involved in the development of schizophrenia in these patients.

Further, I have modeled the results in a threshold model for schizophrenia where an accumulation of variation increases the liability towards disease (McGue *et al.*, 1983). I postulate that genetic or epigenetic changes can act as the source of variation that increases the liability towards disease. Also, in some cases the mutations may be co-localized and affect the same genes. The model assumes that the inherited predisposition towards disease may lie in both monozygotic twins. However, the threshold is reached in the affected twin only by virtue of additional changes (Figure 6.1). The results also offer a number of insights into twin genomics as discussed below.

A number of major conclusions can be drawn from the research presented in this thesis:

First, the results presented here confirm that rare *de novo* events are relatively common. They take place during growth, development and aging making even MZ twins genetically discordant. If a mutation occurs in the early stages of development it may result in distinct differences between monozygotic twins while a later occurrence may generate twins with mosaicism.

Second, the study design has enabled the identification of a wide variety of genome differences. These include CNVs, genome sequence variation and DNA methylation changes.

Third, monozygotic twin pairs discordant for schizophrenia show *de novo* differences at the genetic and epigenetic levels. Many of these differences are rare and patient-specific. Further, some of these differences are shared between unrelated affected twins. This comprehensive and complete assessment of individual genomes is unique in the literature: for the first time it has offered a near complete genetic/epigenetic profile of changes between monozygotic twins. Collectively, the results argue that genetic models and interpretations used over the last 100 years, that are based on 100% genetic concordance of monozygotic twins, deserve reevaluation.

Fourth, many of the *de novo* genes and pathways unique to the affected twin(s) are involved in neurodevelopmental networks/pathways relevant to disease manifestation or have previously been implicated in schizophrenia. These include 'cell death and survival', 'cellular movement and immune cell trafficking', 'glutamate receptor signaling' and 'dopamine feedback in cAMP signaling', among others.

Finally, the genomic results on families with monozygotic twins discordant for schizophrenia are compatible with a threshold model where the accumulation of variants increases the liability towards schizophrenia.

6.1 Threshold Model

The model proposed (Figure 6.1) states that inherited mutations (Girard *et al*., 2011; Gulsuner *et al*., 2013; Xu *et al*., 2011) place an individual on a liability scale for the development of schizophrenia. Additional contributors to this scale may include *de novo* mutations, environmental factors and/or epigenetic (McCarthy *et al*., 2014) events, most occurring during ontogenic development (Singh and O'Reilly, 2009). The model predicts that the liability threshold may be met by inherited factors alone or require additional random genetic and epigenetic events, including environmental events. I have tested this model using monozygotic twins discordant for schizophrenia with the assumption that any inherited components will be shared and that non-inherited components (*de novo* mutations, epigenetic and environmental) will be acquired during ontogenetic development (Figure 6.1). The model also argues that the determinants of the disease may be individual-specific and differ across patients. The results included in this thesis identify a number of events that may have played a role in the development of schizophrenia in the affected member of the two twin pairs studied. A large majority of the genes/regions identified are involved in post-synaptic complexes (Kirov *et al*., 2012) and synaptic strength (Szatkiewicz *et al*., 2014); processes that are known to be defective in patients with schizophrenia.

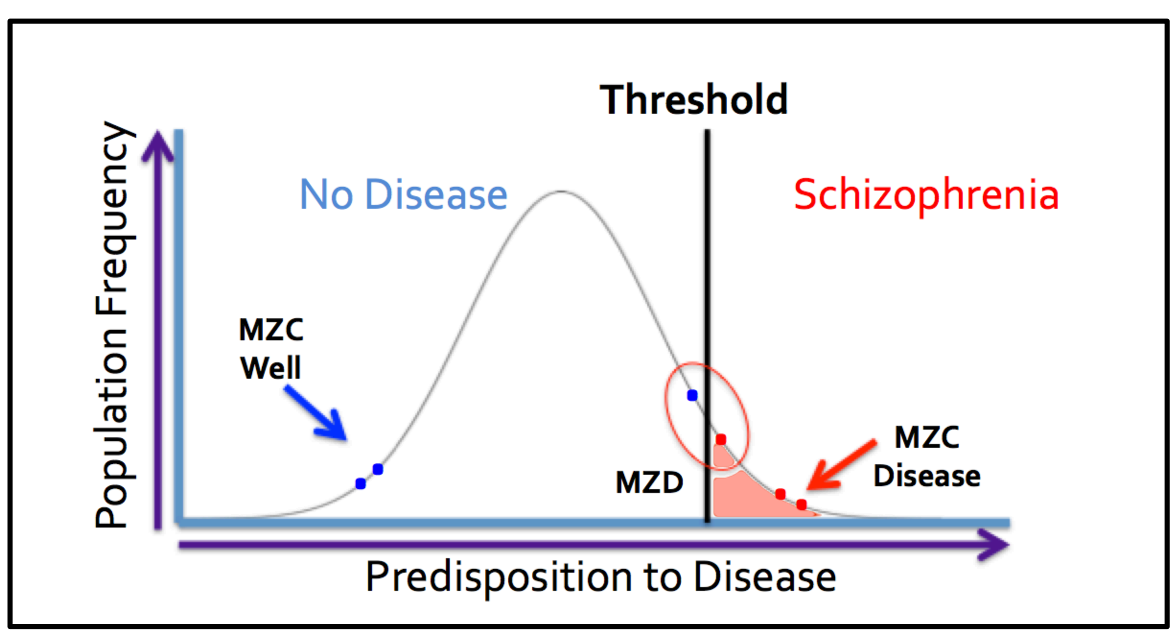


Figure 6.1. Visual representation of the proposed threshold model.

This model was first suggested by McGue *et al,* (McGue *et al*., 1983). Here, I apply its principles to monozyotic twin concordance/discordance for schizophrenia. A blue dot indicates a well twin and a red dot indicates an affected twin. MZC = Monozygotic Concordant, MZD = Monozygotic Discordant.

Thus, the development of schizophrenia is compatible with a threshold model. Previous population and family studies have implicated a number of mutational mechanisms, as well as many genes of small effect, in the causation of this disease (Kendler, 2015; Picchioni and Murray, 2007; Ripke *et al.*, 2013). Conceptually, an accumulation of mutations affecting a number of genes and/or regions affecting disease-specific pathways may establish a genetic liability gradient for schizophrenia where a critical threshold is required for the manifestation of the disease state (Bouchard and McGue, 1981; McGue *et al.*, 1983). Further, the underlying genetic liability is expected to be higher in families with schizophrenia patients. At the individual level, this liability will be much higher if a family history is observed. In the case of MZ twins where one is affected with the disease, the affected member is expected to have acquired additions to the liability scale via genomic mutations and/or environmental factors that contribute to threshold crossing and thus to the development of the disease. It remains an attractive hypothesis but has remained untested, primarily because of the lack of appropriate data. It should also be noted that threshold models are often assessed at the level of the population. This study is unique in that it applies this model to cases representing individual patients representing MZD twins.

The model argues that a gene or region involved in a critical pathway may be affected by mutations involving different sets of variants in different individuals. More importantly, the two members of a MZD twin pair would share a common genetic liability for the disease with the affected twin acquiring additional defects causing him/her to cross the disease threshold on the liability continuum and develop schizophrenia.

6.1.1 Role of *de novo* Genetic and Epigenetic Mutations in the Threshold Model of Schizophrenia

The results included in this thesis, therefore, show that *de novo* genetic and epigenetic mutations may play a significant role in the development of

schizophrenia. These observations are backed by extensive literature that argue for the existence of somatic mutation in our genomes (Breckpot *et al*., 2012; Li *et al*., 2014; O'Huallachain *et al*., 2012; Weber-Lehmann *et al*., 2014). In this model, MZD twins with the disease are argued to have acquired additional *de novo* errors onto an already existing background of genetic and epigenetic predisposition and thus contributing to disease liability, as discussed above. These *de novo* errors must involve somatic changes during ontogeny (Maiti *et al*., 2011; Singh *et al*., 2009). Such rare post-zygotic variants have been ignored in the past. Mutations that occur later in development are likely to be very low in frequency and may even resemble technological error (Insel, 2014). I propose that *de novo* genetic and epigenetic mutations in the affected twin are responsible for the generation of threshold reaching insults in monozygotic twins discordant for schizophrenia and thus contribute to disease manifestation and ultimately, disease discordance.

6.1.2 Further Support for The Threshold Model of Schizophrenia

This model will account for a variety of results and observations in the literature. For example, *de novo* mutations have been found to be more likely to occur in schizophrenia patients as compared to unaffected siblings (Gulsuner and McClellan, 2014). In addition, damaging *de novo* mutations have been found in postmortem brains of schizophrenia patients that disrupted genes regulating neurogenesis in the prefrontal cortex (Gulsuner *et al*., 2013) and an increased number of *de novo* CNVs have been implicated in schizophrenia and bipolar disorder patients (Georgieva *et al*., 2014). Further, schizophrenia has not been the only neurodevelopmental disorder to implicate *de novo* mutation, it has also been shown that patients with autism or intellectual disability have more *de novo* mutations than matched controls (Insel, 2014).

Also, the involvement of a relatively large number of gene mutations reported in comprehensive publications have been difficult to replicate (Farrell *et al*., 2015) and argue for far-reaching heterogeneity and patient-centered

approaches towards the understanding of this neurodevelopmental disease. In addition, various phenotypic variables have been investigated through genome-wide association studies and results reveal that the patients who carried *de novo* CNVs and SNVs displayed the most severe phenotypes (for example, learning disabilities leading to reduced work skills) (Malherbe *et al.*, 2015). Interestingly, the genes harboring *de novo* mutations in schizophrenia studies, including those involving twins, have been found to be enriched in networks and pathways relevant to psychiatric disorders (Bloom *et al.*, 2013; Castellani *et al.*, 2014, 2015; Kirov, 2010; Maiti *et al.*, 2011).

Of critical importance is the ability of this model to explain the persistance of schizophrenia in the population despite its reduced fecundity. In fact, the presence of somatic mutation has been suggested to explain the limited heritability of many cancers (Insel, 2014) and somatic mutations (particularly after inheritance of a predisposition) are believed to be the driving force behind the manifestation of a number of cancers (Lengauer *et al.*, 1998). Further, mosaicism is a well-known occurrence that has been associated with at least 30 Mendelian diseases (O'Huallachain *et al.*, 2012) which supports the primary mechanism implicated in the threshold model of schizophrenia: *de novo* genetic and epigenetic change.

6.2 Context of Study Findings

Evidence from a variety of sources suggests that schizophrenia is a highly heterogeneous neurodevelopmental disorder, caused by many different genes with small effects (Craddock *et al.*, 2006; Kendler, 2015; Schwab and Wildenauer, 2013). The genes and regions identified in this study to be affected by *de novo* genetic and epigenetic mutations in the affected co-twins are the most likely candidates to contribute to observed discordance for schizophrenia through the accumulation of small effects contributing to a threshold. Rare, low frequency variations are enriched for potentially functional mutations because they are under weak purifying selection and have increased levels of population

variation due to the fact that they are recent in origin (Mathieson and McVean, 2012; Nelson *et al*., 2012). Moreover, the function of the genes and pathways/networks identified from *de novo* mutation in the affected co-twins were more likely than genes from the unaffected co-twins to be associated with schizophrenia and more likely to have been previously identified as schizophrenia candidates in the literature, lending support for their potential contribution to schizophrenia manifestation in these discordant pairs.

Differential *de novo* mutations offer an alternative explanation for the discordance of MZ twins. They would add new somatic genetic and epigenetic mutations to individual twins with identical predisposing genotypes. Depending on the developmental time of mutation (very early causing all cells to be affected versus later causing mosaicism) and tissue/organ involved, these *de novo* mutations will accumulate and help reach or cross the threshold producing the disease. The novel results could also explain comparable risk of the disease in the offspring of affected as well as unaffected twins. The findings presented in this thesis suggest that every individual will have a unique combination of inherited and acquired gene mutations, with the potential to make each individual, even from a single family, genetically unique. This notion is supported by recent evidence from the 1000 Genomes Pilot Project which has found considerable variation in mutation rates within and between families (Conrad *et al*., 2011).

Our group has previously suggested that the presence of *de novo* CNV showers (Singh *et al*., 2009) might help explain both the discordance of MZ twins and the persistent high prevalence of schizophrenia despite its selective disadvantage. In this study, an increased burden of *de novo* mutation was not identified between affected twins and their unaffected co-twin. However, other groups have reported a significant difference in the level of *de novo* mutation in cases versus controls. For example, increased *de novo* mutations in exome sequences in schizophrenia (Girard *et al*., 2011; Xu *et al*., 2011).

Many genes implicated in schizophrenia have also been implicated in autism, intellectual disability, bipolar disorder and major depressive disorder (Kavanagh *et al*., 2014), suggesting that similar *de novo* insults may be relevant to multiple complex disorders, including schizophrenia. Further, despite clear genetic contributions to the causation of schizophrenia, fecundity in schizophrenia is significantly reduced (Nimgaonkar, 1998), yet the lifetime morbidity risk remains high at approximately 0.75% (Saha *et al*., 2005). As previously mentioned, one possible explanation for this combination of selective disadvantage and high prevalence is *de novo* mutation which would act against selection pressure (Doi *et al*., 2009).

6.3 Methodological Considerations

Any study of this magnitude is not without its strengths and limitations. Here, I discuss a few major strengths of the experiments presented in this thesis as well as some caveats.

6.3.1 Study Strengths

First, the use of monozygotic twins and parental samples (where available) allows for the assessment of a patient with schizophrenia as compared to nature's best possible matched control – their identical twin. Although heterogeneity in environmental and random effects cannot be eliminated, a significant reduction in heterogeneity is achieved given that twins start life from the same egg and sperm cell and have shared in-utero environments, maternal nutrition, gestation lengths, and early life experiences, among other similarities.

Second, the technologies used to generate the datasets in this thesis comprise some of the most comprehensive and highest resolution methods currently available in this field, allowing for the assessment of the genome at the single base pair level.

Finally, the results allow for the interpretation and analysis of genomic change at the individual/patient-specific level, which is important when assessing

causation to a disease that presents with a wide spectrum of symptoms and manifestations.

6.3.2 Study Caveats

There are a number of added considerations that must be taken into account in the interpretation of the results generated during this Ph.D. research. Firstly, most schizophrenia patients, including those analyzed in this thesis, are treated with antipsychotic medications prior to the time of sample collection for genomic analysis. It is understood that drugs used to treat psychiatric disorders may cause epigenetic changes in the genome (Dong *et al*., 2009) and thus it will be expected that some of the epigenetic changes presented in this report may have arisen as a result of the effect of medication (Melka *et al*., 2013, 2014a, 2014b). The questions remain: are these changes the causes or the effects of the disease process and what is the role, if any, of drugs used by patients? However, it should be noted that the possibility that all of the observed differences are an effect of medication is not compatible with several facts: firstly, the novel differences that have been detected in patients affect genes and/or pathways that are implicated (and affected) in schizophrenia and secondly, studies on medication-free patients have shown aberrant genomic profiles between schizophrenic patients and healthy controls (Bönsch *et al*., 2012; Kinoshita *et al*., 2013).

Second, this study primarily used blood samples for assessment of genetic and epigenetic profiles of participants. It is important to note that it is likely that some *de novo* mutation would exist in all/most tissues of the same individual and that other *de novo* mutation would exist only in a subset of tissue, or just one tissue depending on the timing of the mutation. However, with current methods and limited tissue types, I cannot precisely ascertain the timing of *de novo* change. As mentioned above, the additional use of buccal cells for confirmatory experiments did allow the estimation of timing for some variants. I note that some of the variants identified in this study occurred before the

differentiation of these two tissues and others occurred after. Thus, I would expect some variants to be brain-specific (which I have not been able to measure in this study) (Poduri *et al.*, 2013). Of note however, is the increased potential for *de novo* mutation to occur during development due to the high rates of replication observed at this critical time. Genome-wide studies have suggested that the methylation status of blood reflects that in the brain, which supports the use of blood as a surrogate tissue (Aberg *et al.*, 2013). However, until the field's understanding of somatic change is more thoroughly defined, the precise validity of surrogate tissues in studies of disease etiology remains unclear. In fact, evidence is beginning to emerge for the potential that even at the neuronal level, one neuron may be distinctly different in DNA sequence from another neuron directly beside it (Insel, 2014).

Third, the use of twins addresses the issue of heterogeneity in schizophrenia patients, allowing for a significant reduction in inter-individual variability. However, despite this unique match between a schizophrenic twin and their healthy co-twin, there is no precedent with regards to sample size of twins needed to detect potential disease contributing variations. Rather than assessing a fraction of variations in an increasing number of samples to increase effectiveness, my focus is to assess all variations in individual patients and genetically matched controls. In theory, a single pair may be enough to identify a causal element(s). I have demonstrated the effectiveness of this approach by reporting on Cadherin as a candidate gene for schizophrenia (Singh *et al.*, 2010). For this reason, I have chosen to primarily report patient and family-specific profiles in each family. In cases where I have identified shared variants between unrelated patients, I have proceeded with caution and identified these variants only as potential candidates for the disease that require further assessment in population level GWAS analysis.

Finally, the current resolution available in the area of genetics and genomics is limiting. For example, somatic variants could be present in just a few reads in a tissue sample but still be important for disease causation, yet current

technologies will likely identify mosacism as nothing more than sequencing or hybridization error. Thus, as technology progresses so too will the reliability with which we can ascertain precise information about the genome.

Whole genome sequencing is estimated to have an error rate ranging from 0.1-0.6% (Wall *et al*., 2014). The technology used for sequencing in this study from Complete Genomics has an estimated error rate of 0.001% (Drmanac *et al*., 2010). When the size of the genome is taken into account, this means that there are up to 3 million potential errors in each genome sequence analyzed. This reminds us to proceed with caution when interpreting large-scale databases and reinforces the importance of downstream confirmatory methods.

In fact, most groups that are working with large-scale datasets report concerning results with regards to validation rates in population studies as well as twin studies. For example, a genome-wide CNV study using Affymetrix 6.0 array in 1,097 monozygotic twins chose 20 CNV candidates for confirmation (of 153 putative *de novo* CNVs) and were able to validate 2 of the 20 candidates (Abdellaoui *et al*., 2015). In addition, a comprehensive report by Dal *et al* uncovered *de novo* SNVs through whole genome deep sequencing in two twin pairs and were able to validate 8 of 19 variants in one twin pair and 1 of 15 variants in another twin pair (Dal *et al*., 2014). In addition, the latter report describes the presence of mosaicism at twin-specific *de novo* SNVs (Dal *et al*., 2014). These results in the literature are very similar to the validation rates that I have identified in my Affymetrix 6.0 and whole genome sequencing datasets in this thesis.

The sequencing dataset poses a particular challenge as regions of low sequence complexity, satellite regions and large repeats, continue to present major difficulty for short read sequencing technologies such as the Complete Genomics method. Despite an average fold coverage of over 47x in this study, this average is not equally distributed along each and every chromosome; some repeat regions have significantly less while other euchromatic regions have significantly more fold coverage.

6.4 Implications

The results have provided a reminder that clinical research should no longer be about nature versus nurture, but instead about the complex interplay of nature and nurture. The findings have a number of potential implications. First, these conclusions may help develop earlier and/or better diagnosis and have the potential to identify diagnostic sub-groups through the grouping of patients with similar genetic and epigenetic profiles. Earlier diagnosis alone could assist greatly in disease management and early treatment strategies. Second, the findings may help to identify better therapeutic strategies through the identification of drug targets as well as candidate regions for targeted methylation and/or genetic therapy strategies. Finally, the findings contribute to the understanding of schizophrenia causes. Knowledge of the causes of schizophrenia will assist patients in uncovering the factors behind why they are experiencing symptoms of this life-altering mental health disorder.

The identification of genetic and epigenetic changes involved in even a subset of schizophrenia patients will be a major breakthrough. Identification of genomic changes that are contributing to this disease also have the potential to identify affected pathways and etiological mechanisms that will allow characterization of cytoarchitecture with diagnostic and therapeutic implications.

6.5 Future Directions

The field of psychiatric genetics is vast and incredibly complex. Although genetic technologies and advances are moving forward at unprecedented rates, the challenges that lie ahead are many. Expanding on the research presented here is paramount to understanding the genetic variation between and within individuals, as well as the critical drivers of disease manifestation. Future studies should focus on the following objectives:

1. Gene expression and other functional studies on the candidate genes and regions presented in this study to uncover which of the identified changes may be causal to the disease.
2. Further characterization of schizophrenia patients into more homogenous clinical subgroups would allow for genetic studies to be performed that focus on one subtype of schizophrenia at a time thus significantly reducing clinical heterogeneity. This aim will also lead to better diagnosis of patients.
3. Technological advances that will allow for better identification of the timing of *de novo* genetic/epigenetic mutations and a significant reduction in the error rates of sequencing and array-based technologies.
4. Investigation of somatic variation in the brains of monozygotic twins, particularly using single cell technologies that are now emerging.
5. Assessment of other types of epigenetic modifications including histone modification, DNA hydroxymethylation, and MicroRNA profiles between identical twins discordant for schizophrenia.
6. As our understanding of non-coding regions of the genome becomes clearer, the impact of non-coding variants on disease will be revealed. Studies which focus on the potential for non-coding variants to have a regulatory effect on important genes may uncover novel genomic candidates for disease presentation.
7. Expanded studies on medication-free patients will allow for better assessment of epigenetic responses to treatment. A database of epigenetic profiles in healthy controls, patients treated with antipsychotics and patients without medication may assist in this aim.
8. Combination of multiple data types towards an integrative genomics approach that includes better understanding of the functional biological pathways using functional studies, gene expression, DNA sequence and epigenetic data.

9. Experiments that work towards better understanding of the interactions between environmental factors and genomic insults will provide the background for better understanding of how environmental factors mediate the risk of schizophrenia.

6.6 Final Conclusions

In conclusion, these collective experiments using a unique twin model and experimental design, alongside comprehensive genomic techniques, identified differences between monozygotic twins at both the genetic and genomic level. These findings highlight the variability that exists in our genomes and emphasizes that no two individuals are identical; not even twins who begin life as a single zygote. In addition, the results presented in this thesis identify genetic candidates for the development of schizophrenia in monozygotic twins discordant for schizophrenia with some of these candidates having been previously identified and others being novel. These novel insults may act on an already existing background of predisposition via a threshold model (McGue *et al*., 1983) (Castellani *et al*., 2015, *under review*).

The inherent dynamic and responsive nature of the landscape of the human genome makes it a candidate for the explanation of disease discordance in monozygotic twins for a number of phenotypes, yet also presents an experimental challenge given its complexity and the vast number of factors that are known to contribute to genomic change. Discordant twins and their parents provide exceptionally matched controls for assessment of genomic contributors to disease. Accordingly, the approach presented here may be applied to other complex diseases that show twin discordance.

The human genome is not static. We do not simply acquire parental genomes and then pass them on to the next generation; rather, the genome undergoes a variety of alterations at the genetic and epigenetic level. These changes may represent a way for our genomes to remain both fluid and malleable as we encounter a variety of environmental impacts, however, these

changes have the potential to be detrimental and may be leading factors towards the formation of complex diseases.

Furthering our knowledge of genomic architecture and change, as well as the timing and source of genomic change, will help to lead to a better understanding of the dynamic nature of the human genome. Ultimately, determining the contributing factors to complex diseases, like schizophrenia, will aid in the diagnosis, treatment and understanding of this and other life-altering disorders.

6.7 References

Abdellaoui, A., Ehli, E.A., Hottenga, J.-J., Weber, Z., Mbarek, H., Willemsen, G., van Beijsterveldt, T., Brooks, A., Hudziak, J.J., Sullivan, P.F., *et al*. (2015). CNV Concordance in 1,097 MZ Twin Pairs. Twin Res. Hum. Genet. *18*, 1–12.

Aberg, K.A., Xie, L.Y., McClay, J.L., Nerella, S., Vunck, S., Snider, S., Beardsley, P.M., and van den Oord, E.J.C.G. (2013). Testing two models describing how methylome-wide studies in blood are informative for psychiatric conditions. Epigenomics *5*, 367–377.

Bloom, R.J., Kahler, A.K., Collins, A.L., Chen, G., Cannon, T.D., Hultman, C., and Sullivan, P.F. (2013). Comprehensive analysis of copy number variation in monozygotic twins discordant for bipolar disorder or schizophrenia. Schizophr. Res. *146*, 289–290.

Bönsch, D., Wunschel, M., Lenz, B., Janssen, G., Weisbrod, M., and Sauer, H. (2012). Methylation matters? Decreased methylation status of genomic DNA in the blood of schizophrenic twins. Psychiatry Res. *198*, 533–537.

Bouchard, T.J., and McGue, M. (1981). Familial studies of intelligence: a review. Science *212*, 1055–1059.

Breckpot, J., Thienpont, B., Gewillig, M., Allegaert, K., Vermeesch, J.R., and Devriendt, K. (2012). Differences in Copy Number Variation between Discordant Monozygotic Twins as a Model for Exploring Chromosomal Mosaicism in Congenital Heart Defects. Mol. Syndromol. *2*, 81–87.

Castellani, C.A., Awamleh, Z., Melka, M.G., O'Reilly, R.L., and Singh, S.M. (2014). Copy number variation distribution in six monozygotic twin pairs discordant for schizophrenia. Twin Res. Hum. Genet. *17*, 108–120.

Castellani, C.A., Laufer, B.I., Melka, M.G., Diehl, E.J., O'Reilly, R.L., and Singh, S.M. (2015). DNA methylation differences in monozygotic twin pairs discordant for schizophrenia identifies psychosis related genes and networks. BMC Med. Genomics *8*, 17.

Conrad, D.F., Keebler, J.E.M., DePristo, M.A., Lindsay, S.J., Zhang, Y., Casals, F., Idaghdour, Y., Hartl, C.L., Torroja, C., Garimella, K. V, *et al*. (2011). Variation in genome-wide mutation rates within and between human families. Nat. Genet. *43*, 712–714.

Craddock, N., O'Donovan, M.C., and Owen, M.J. (2006). Genes for schizophrenia and bipolar disorder? Implications for psychiatric nosology. Schizophr. Bull. *32*, 9–16.

Dal, G.M., Ergüner, B., Sağıroğlu, M.S., Yüksel, B., Onat, O.E., Alkan, C., and Özçelik, T. (2014). Early postzygotic mutations contribute to *de novo* variation in a healthy monozygotic twin pair. J. Med. Genet. *51*, 455–459.

Doi, N., Hoshi, Y., Itokawa, M., Usui, C., Yoshikawa, T., and Tachikawa, H. (2009). Persistence criteria for susceptibility genes for schizophrenia: a discussion from an evolutionary viewpoint. PLoS One *4*, e7799.

Dong, E., Grayson, D.R., Guidotti, A., and Costa, E. (2009). Antipsychotic subtypes can be characterized by differences in their ability to modify GABAergic promoter methylation. Epigenomics *1*, 201–211.

Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., *et al*. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science *327*, 78–81.

Farrell, M.S., Werge, T., Sklar, P., Owen, M.J., Ophoff, R.A., O'Donovan, M.C., Corvin, A., Cichon, S., and Sullivan, P.F. (2015). Evaluating historical candidate genes for schizophrenia. Mol. Psychiatry.

Georgieva, L., Rees, E., Moran, J.L., Chambert, K.D., Milanova, V., Craddock, N., Purcell, S., Sklar, P., McCarroll, S., Holmans, P., *et al*. (2014). *De novo* CNVs in bipolar affective disorder and schizophrenia. Hum. Mol. Genet. *23*, 6677–6683.

Girard, S.L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., Dionne-Laporte, A., Spiegelman, D., Henrion, E., Diallo, O., *et al*. (2011). Increased exonic *de novo* mutation rate in individuals with schizophrenia. Nat. Genet. *43*, 860–863.

Gulsuner, S., and McClellan, J.M. (2014). *De novo* mutations in schizophrenia disrupt genes co-expressed in fetal prefrontal cortex. Neuropsychopharmacology *39*, 238–239.

Gulsuner, S., Walsh, T., Watts, A.C., Lee, M.K., Thornton, A.M., Casadei, S., Rippey, C., Shahin, H., Nimgaonkar, V.L., Go, R.C.P., *et al.* (2013). Spatial and temporal mapping of *de novo* mutations in schizophrenia to a fetal prefrontal cortical network. Cell *154*, 518–529.

Insel, T.R. (2014). Brain somatic mutations: the dark matter of psychiatric genetics? Mol. Psychiatry *19*, 156–158.

Kavanagh, D.H., Tansey, K.E., O'Donovan, M.C., and Owen, M.J. (2014). Schizophrenia genetics: emerging themes for a complex disorder. Mol. Psychiatry *20*, 72–76.

Kendler, K.S. (2015). A joint history of the nature of genetic variation and the nature of schizophrenia. Mol. Psychiatry *20*, 77–83.

Kinoshita, M., Numata, S., Tajima, A., Shimodera, S., Ono, S., Imamura, A., Iga, J., Watanabe, S., Kikuchi, K., Kubo, H., *et al.* (2013). DNA methylation signatures of peripheral leukocytes in schizophrenia. Neuromolecular Med. *15*, 95–101.

Kirov, G. (2010). The role of copy number variation in schizophrenia. Expert Rev. Neurother. *10*, 25–32.

Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., *et al.* (2012). *De novo* CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol. Psychiatry *17*, 142–153.

Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. Nature *396*, 643–649.

Li, R., Montpetit, A., Rousseau, M., Wu, S.Y.M., Greenwood, C.M.T., Spector, T.D., Pollak, M., Polychronakos, C., and Richards, J.B. (2014). Somatic point mutations occurring early in development: a monozygotic twin study. J. Med. Genet. *51*, 28–34.

Maiti, S., Kumar, K.H.B.G., Castellani, C.A., O'Reilly, R., and Singh, S.M. (2011). Ontogenetic *de novo* copy number variations (CNVs) as a source of genetic individuality: studies on two families with MZD twins for schizophrenia. PLoS One *6*, e17125.

Malherbe, P.J., Roos, J.L., Ehlers, R., and Karayiorgou, M. (2015). Phenotypic features of patients with schizophrenia carrying *de novo* gene mutations: a pilot study. Psychiatry Res. *225*, 108–114.

Mathieson, I., and McVean, G. (2012). Differential confounding of rare and common variants in spatially structured populations. Nat. Genet. *44*, 243–246.

McCarthy, S.E., Gillis, J., Kramer, M., Lihm, J., Yoon, S., Berstein, Y., Mistry, M., Pavlidis, P., Solomon, R., Ghiban, E., *et al*. (2014). *De novo* mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. Mol. Psychiatry *19*, 652–658.

McGue, M., Gottesman, I.I., and Rao, D.C. (1983). The transmission of schizophrenia under a multifactorial threshold model. Am. J. Hum. Genet. *35*, 1161–1178.

Melka, M.G., Castellani, C.A., Laufer, B.I., Rajakumar, R.N., O'Reilly, R., and Singh, S.M. (2013). Olanzapine induced DNA methylation changes support the dopamine hypothesis of psychosis. J. Mol. Psychiatry *1*, 19.

Melka, M.G., Laufer, B.I., McDonald, P., Castellani, C.A., Rajakumar, N., O'Reilly, R., and Singh, S.M. (2014a). The effects of olanzapine on genome-wide DNA methylation in the hippocampus and cerebellum. Clin. Epigenetics *6*, 1.

Melka, M.G., Castellani, C.A., Rajakumar, N., O'Reilly, R., and Singh, S.M. (2014b). Olanzapine-induced methylation alters cadherin gene families and associated pathways implicated in psychosis. BMC Neurosci. *15*, 112.

Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., St Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S.-A., Fraser, D., *et al.* (2012). An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science *337*, 100–104.

Nimgaonkar, V.L. (1998). Reduced fertility in schizophrenia: here to stay? Acta Psychiatr. Scand. *98*, 348–353.

O'Huallachain, M., Karczewski, K.J., Weissman, S.M., Urban, A.E., and Snyder, M.P. (2012). Extensive genetic variation in somatic human tissues. Proc. Natl. Acad. Sci. U. S. A. *109*, 18018–18023.

Picchioni, M.M., and Murray, R.M. (2007). Schizophrenia. BMJ *335*, 91–95.

Poduri, A., Evrony, G.D., Cai, X., and Walsh, C.A. (2013). Somatic mutation, genomic variation, and neurological disease. Science *341*, 1237758.

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kahler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., *et al*. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat. Genet. *45*, 1150–1159.

Saha, S., Chant, D., Welham, J., and McGrath, J. (2005). A systematic review of the prevalence of schizophrenia. PLoS Med. *2*, e141.

Schwab, S.G., and Wildenauer, D.B. (2013). Genetics of psychiatric disorders in the GWAS era: an update on schizophrenia. Eur. Arch. Psychiatry Clin. Neurosci. *263 Suppl* , S147–S154.

Singh, S.M., and O'Reilly, R. (2009). (Epi)genomics and neurodevelopment in schizophrenia: monozygotic twins discordant for schizophrenia augment the search for disease-related (epi)genomic alterations. Genome *52*, 8–19.

Singh, S.M., Castellani, C.A., and O'Reilly, R.L. (2009). Copy number variation showers in schizophrenia: an emerging hypothesis. Mol. Psychiatry *14*, 356–358.

Singh, S.M., Castellani, C., and O'Reilly, R. (2010). Autism meets schizophrenia via cadherin pathway. Schizophr. Res. *116*, 293–294.

Szatkiewicz, J.P., O'Dushlaine, C., Chen, G., Chambert, K., Moran, J.L., Neale, B.M., Fromer, M., Ruderfer, D., Akterin, S., Bergen, S.E., *et al*. (2014). Copy number variation in schizophrenia in Sweden. Mol. Psychiatry *19*, 762–773.

Wall, J.D., Tang, L.F., Zerbe, B., Kvale, M.N., Kwok, P.-Y., Schaefer, C., and Risch, N. (2014). Estimating genotype error rates from high-coverage next-generation sequence data. Genome Res. *24*, 1734–1739.

Weber-Lehmann, J., Schilling, E., Gradl, G., Richter, D.C., Wiehler, J., and Rolf, B. (2014). Finding the needle in the haystack: differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. Forensic Sci. Int. Genet. *9*, 42–46.

Xu, B., Roos, J.L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., Gogos, J.A., and Karayiorgou, M. (2011). Exome sequencing supports a *de novo* mutational paradigm for schizophrenia. Nat. Genet. *43*, 864–868.

Appendices

Appendix A - Availability of Supporting Data

The datasets supporting the results of this thesis are available in the Gene Expression Omnibus (GEO) repository, [GSE61862, http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE61862] and [GSE33598, http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33598].

Appendix B - Ethics Approval

Research Western

Use of Human Participants - Ethics Approval Notice

Principal Investigator: Dr. Richard O'Reilly
Review Number: 18017E
Review Level: Delegated
Approved Local Participants: 0
Protocol Title: A Twin Study to Locate Genetic Differences that Cause Schizophrenia
Department & Institution: Psychiatry, Regional Mental Health Care St. Thomas
Sponsor: Canadian Institutes of Health Research

Ethics Approval Date: May 10, 2011 **Expiry Date:** December 31, 2016
Documents Reviewed & Approved & Documents Received for Information:

Document Name	Comments	Version Date
UWO Protocol		
Letter of Information & Consent	Discordant	
Letter of Information & Consent	Discordant - re-consent	
Letter of Information & Consent	Concordant	
Letter of Information & Consent	Concordant - re-consent	
Letter of Information & Consent	Non-affected	
Letter of Information & Consent	Non-affected - re-consent	
Letter of Information & Consent	Parent	
Letter of Information & Consent	Parent - re-consent	
Advertisement		
Advertisement		

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the UWO Updated Approval Request Form.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The UWO HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Ethics Officer to Contact for Further Information

Janice Sutherland (jsutherl@uwo.ca)	Grace Kelly (grace.kelly@uwo.ca)

This is an official document. Please retain the original in your files.

The University *of* Western Ontario
Office of Research Ethics
Support Services Building Room 5150 • London, Ontario • CANADA - N6A 3K7
PH: 519-661-3036 • F: 519-850-2466 • ethics@uwo.ca • www.uwo.ca/research/ethics

Appendix C - Structural Variants unique to affected member of Family 1. (NCBI Build 37/hg19)

Chr	Start	End	Size (bp)	SV Type	Gene(s)	Father	Mother	Inheritance
1	5717300	5725895	8595	Deletion	0	Present	Absent	Inherited
1	5727877	5728315	438	Deletion	0	Present	Present	Inherited
1	26460142	26464816	4674	Deletion	0	Present	Absent	Inherited
1	57242625	57242964	339	Deletion	C1orf168	Present	Present	Inherited
1	62654769	62657187	2418	Deletion	0	Present	Absent	Inherited
1	71237363	71240327	2964	Deletion	0	Present	Absent	Inherited
1	112691792	112704705	12913	Deletion	0	Present	Present	Inherited
1	144894289	144895285	996	Deletion	PDE4DIP	Present	Present	Inherited
1	145026746	145027069	323	Deletion	PDE4DIP	Present	Present	Inherited
1	158867539	158869984	2445	Deletion	0	Present	Present	Inherited
1	238852322	238853265	943	Deletion	0	Present	Absent	Inherited
1	248051492	248057652	6160	Deletion	0	Present	Present	Inherited
2	1801743	1802067	324	Deletion	MYT1L	Present	Absent	Inherited
2	3822800	3823598	798	Deletion	0	Present	Present	Inherited
2	33091738	33095634	3896	Deletion	LOC285045	Present	Present	Inherited
2	182565264	182565592	328	Deletion	0	Present	Absent	Inherited
2	239626179	239627520	1341	Deletion	0	Present	Present	Inherited
3	228847	234481	5634	Deletion	0	Present	Present	Inherited
3	8718847	8722982	4135	Deletion	0	Present	Absent	Inherited
3	16239409	16241210	1801	Deletion	GALNTL2	Present	Absent	Inherited
3	49494287	49494616	329	Deletion	0	Present	Present	Inherited
3	73939035	73939381	346	Deletion	0	Present	Absent	Inherited
4	4568030	4568628	598	Deletion	0	Present	Absent	Inherited
4	28511981	28513202	1221	Deletion	0	Present	Absent	Inherited
4	59428999	59435542	6543	Deletion	0	Present	Absent	Inherited
4	73831249	73831821	572	Deletion	0	Present	Absent	Inherited

4	76743108	76744645	1537	Deletion	0	Present	Present	Inherited
4	80888062	80894107	6045	Deletion	ANTXR2	Present	Present	Inherited
4	90427886	90428672	786	Deletion	0	Present	Absent	Inherited
4	190535505	190538363	2858	Deletion	0	Present	Absent	Inherited
5	668063	668593	530	Deletion	LOC100132605;TPPP	Present	Present	Inherited
5	25959078	25959871	793	Deletion	0	Present	Absent	Inherited
5	42628539	42631218	2679	Deletion	GHR	Present	Absent	Inherited
5	95343326	95343680	354	Deletion	0	Present	Present	Inherited
5	112224253	112226906	2653	Deletion	REEP5	Present	Absent	Inherited
5	126126039	126127716	1677	Deletion	LMNB1	Present	Absent	Inherited
5	127172332	127173282	950	Deletion	LOC728586	Present	Present	Inherited
5	176387587	176390188	2601	Deletion	UIMC1	Present	Absent	Inherited
6	14745281	14745697	416	Deletion	0	Present	Present	Inherited
6	31193512	31194660	1148	Deletion	0	Present	Absent	Inherited
6	51739577	51745615	6038	Deletion	PKHD1	Present	Absent	Inherited
6	95193324	95194338	1014	Deletion	0	Present	Present	Inherited
6	133371501	133371845	344	Deletion	0	Present	Absent	Inherited
6	149409430	149410432	1002	Deletion	0	Present	Absent	Inherited
6	168059226	168059550	324	Deletion	0	Present	Absent	Inherited
7	33898856	33899187	331	Deletion	0	Present	Present	Inherited
7	36441222	36441996	774	Deletion	ANLN	Present	Absent	Inherited
7	66193913	73382120	7188207	Deletion	ABHD11;AUTS2;BAZ1B;BCL7B;C7orf42;CALN1;CLDN3;CLDN4;DNAJC30;FKBP6;FZD9;GTF2IRD2P;LOC100093631;LOC100128031;LOC100131972;LOC100288540;LOC100289307;LOC100289339;LOC442572;LOC644794;LOC729156;MLXIPL;NCF1B;NSUN5;NSUN5C;PMS2L4;POM121;RABGEF1;SBDS;SBDSP;SPDYE7P;SPDYE8P;STAG3L3;STAG3L4;STX1A;TBL2;TRIM50;TRIM74;TYW1;TYW1B;VP	Present	Present	Inherited

					S37D;WBSCR17;WBSCR22;WBSCR26;WBSCR27;WBSCR28			
7	81441673	81442580	907	Deletion	0	Present	Present	Inherited
7	100613739	100614479	740	Deletion	MUC12	Present	Absent	Inherited
7	102800742	102801402	660	Deletion	0	Present	Absent	Inherited
7	155199777	155201525	1748	Deletion	0	Present	Absent	Inherited
8	40774677	40779833	5156	Deletion	0	Present	Absent	Inherited
8	59078464	59078800	336	Deletion	0	Present	Absent	Inherited
8	113085436	113085766	330	Deletion	0	Present	Present	Inherited
8	144056582	144057161	579	Deletion	0	Present	Absent	Inherited
9	6700565	6710685	10120	Deletion	0	Present	Present	Inherited
9	24502071	24519103	17032	Deletion	0	Present	Present	Inherited
9	110018286	110020857	2571	Deletion	0	Present	Absent	Inherited
9	124411502	124411804	302	Deletion	DAB2IP	Present	Absent	Inherited
9	129492472	129492919	447	Deletion	0	Present	Absent	Inherited
9	135622738	135627436	4698	Deletion	C9orf98	Present	Absent	Inherited
10	22081665	22082212	547	Deletion	DNAJC1	Present	Absent	Inherited
10	78346590	78351578	4988	Deletion	0	Present	Absent	Inherited
10	78800911	78801230	319	Deletion	KCNMA1	Present	Absent	Inherited
10	119634283	119634614	331	Deletion	0	Present	Absent	Inherited
11	5784571	5809284	24713	Deletion	OR52N1;OR52N5	Present	Absent	Inherited
11	12152558	12152882	324	Deletion	MICAL2	Present	Present	Inherited
11	63698909	63701654	2745	Deletion	0	Present	Present	Inherited
11	69400843	69401405	562	Deletion	0	Present	Present	Inherited
11	102751130	102752828	1698	Deletion	LOC100288111	Present	Absent	Inherited
12	9017095	9018122	1027	Deletion	A2ML1	Present	Present	Inherited
12	16420123	16421283	1160	Deletion	0	Present	Absent	Inherited
12	99793948	99802771	8823	Deletion	ANKS1B	Absent	Absent	*De novo*
12	112296544	112297279	735	Deletion	MAPKAPK5	Present	Absent	Inherited
12	122634715	122650584	15869	Deletion	0	Present	Absent	Inherited

12	127199963	127200302	339	Deletion	0	Absent	Present	Inherited
13	21951418	21951834	416	Deletion	ZDHHC20	Present	Absent	Inherited
13	81747255	81747583	328	Deletion	0	Absent	Absent	*De novo*
13	108560717	108561673	956	Deletion	0	Present	Absent	Inherited
14	66157661	66158245	584	Deletion	FUT8	Present	Absent	Inherited
14	73389502	73390207	705	Deletion	0	Absent	Absent	*De novo*
14	85139146	85139626	480	Deletion	0	Present	Absent	Inherited
14	106932640	107174930	242290	Deletion	0	Absent	Present	Inherited
15	39267016	39267370	354	Deletion	0	Present	Present	Inherited
15	93897281	93897642	361	Deletion	0	Absent	Absent	*De novo*
15	101051762	101052327	565	Deletion	LASS3	Present	Present	Inherited
16	11683765	11685325	1560	Deletion	0	Present	Absent	Inherited
16	25340111	25343129	3018	Deletion	0	Present	Present	Inherited
17	801626	802228	602	Deletion	NXN	Absent	Present	Inherited
17	5587702	5588656	954	Deletion	0	Absent	Absent	*De novo*
17	48427538	48427922	384	Deletion	LOC100288444;XYLT2	Present	Present	Inherited
18	108872	109083	211	Deletion	0	Absent	Absent	*De novo*
18	52447413	52447773	360	Deletion	0	Present	Absent	Inherited
18	63766868	63769201	2333	Deletion	0	Present	Present	Inherited
19	1182661	1183179	518	Deletion	0	Absent	Absent	*De novo*
19	12753340	12754497	1157	Deletion	0	Absent	Absent	*De novo*
19	14732344	14734142	1798	Deletion	EMR3	Present	Present	Inherited
20	23671899	23674222	2323	Deletion	0	Present	Absent	Inherited
20	26214307	26220376	6069	Deletion	0	Absent	Absent	*De novo*
20	44535505	44535941	436	Deletion	PLTP	Absent	Present	Inherited
22	35645166	35646141	975	Deletion	0	Absent	Present	Inherited
22	43101594	43102348	754	Deletion	A4GALT	Present	Absent	Inherited
22	49063815	49064594	779	Deletion	FAM19A5	Present	Present	Inherited
22	49764594	49769018	4424	Deletion	0	Absent	Absent	*De novo*
X	899553	899984	431	Deletion	0	Present	Absent	Inherited

X	16427688	16428223	535	Deletion	0	Absent	Present	Inherited
X	49731220	49733608	2388	Deletion	CLCN5	Absent	Absent	*De novo*
1	825765	5726936	4901171	Tandem Dup	ACAP3;ACTRT2;AGRN;AJAP1;ARHGEF16;ATAD3A;ATAD3B;ATAD3C;AURKAIP1;B3GALT6;C1orf159;C1orf170;C1orf174;C1orf222;C1orf70;C1orf86;C1orf93;CALML6;CCDC27;CCNL2;CDC2L1;CDC2L2;CPSF3L;DFFB;DVL1;FAM132A;FLJ14100;FLJ39609;FLJ42875;GABRD;GLTPD1;GNB1;HES4;HES5;ISG15;KIAA0495;KIAA0562;KIAA1751;KLHL17;LOC100128003;LOC100128838;LOC100129381;LOC100129534;LOC100131742;LOC100132814;LOC100133612;LOC100287685;LOC100287750;LOC100287848;LOC100287898;LOC100288202;LOC100288271;LOC100288313;LOC100288379;LOC100288479;LOC115110;LOC148413;LOC284661;LOC388588;LOC401934;LOC441869;LOC643988;LOC728661;LOC728716;LRRC47;MEGF6;MIB2;MMEL1;MMP23A;MMP23B;MORN1;MRPL20;MXRA8;NADK;NOC2L;PANK4;PEX10;PLCH2;PLEKHN1;PRDM16;PRKCZ;PUSL1;RER1;SAMD11;SCNN1D;SDF4;SKI;SLC35E2;SSU72;TAS1R3;TMEM52;TMEM88B;TNFRSF14;TNFRSF18;TNFRSF4;TP73;TPRG1L;TTLL10;UBE2J2;VWA1;WDR8	Absent	Absent	*De novo*
1	230107606	230107783	177	Tandem Dup	0	Absent	Absent	*De novo*
2	1426068	1522926	96858	Tandem Dup	TPO	Absent	Present	Inherited
2	2649502	2649583	81	Tandem Dup	0	Present	Absent	Inherited
3	75699636	75699709	73	Tandem Dup	0	Present	Absent	Inherited
4	80273622	80274729	1107	Tandem Dup	0	Absent	Absent	*De novo*

5	178012521	178012708	187	Tandem Dup	COL23A1	Absent	Present	Inherited
6	1053869	1054275	406	Tandem Dup	LOC285768	Absent	Absent	*De novo*
7	605977	606269	292	Tandem Dup	PRKAR1B	Present	Present	Inherited
7	140181586	140189220	7634	Tandem Dup	0	Present	Present	Inherited
10	125919520	125919748	228	Tandem Dup	0	Absent	Absent	*De novo*
10	132294621	132294747	126	Tandem Dup	0	Absent	Absent	*De novo*
11	131550263	131550706	443	Tandem Dup	NTM	Absent	Absent	*De novo*
14	24369410	24369615	205	Tandem Dup	0	Absent	Absent	*De novo*
14	66258294	66258410	116	Tandem Dup	0	Present	Present	Inherited
15	25468396	25470000	1604	Tandem Dup	SNORD115-29	Absent	Absent	*De novo*
15	63185677	63185745	68	Tandem Dup	0	Absent	Present	Inherited
19	544936	545072	136	Tandem Dup	GZMM	Absent	Absent	*De novo*
19	23262787	23265543	2756	Tandem Dup	0	Present	Absent	Inherited
20	24510425	24510694	269	Tandem Dup	C20orf39	Present	Absent	Inherited
2	61702932/ 61700753	61703455/ 61700765	523	Distal Dup	0	Absent	Present	Inherited
5	150319009/ 150319494	150319103/ 150319759	94	Distal Dup	LOC134466	Present	Absent	Inherited
12	12544853/ 12546495	12544955/ 12546622	102	Distal Dup	LOH12CR1	Present	Present	Inherited
12	87205701/ 87204798	87205819/ 87204806	118	Distal Dup	MGAT4C	Present	Absent	Inherited
13	21538303/ 21545834	21541177/ 21545834	2874	Distal Dup	0	Present	Absent	Inherited
16	69761819/ 69762887	69762131/ 69763048	312	Distal Dup	0	Absent	Present	Inherited
18	69711799/ 69712885	69712015/ 69712888	216	Distal Dup	0	Absent	Present	Inherited
23	2141240/ 2143040	2141701/ 2143092	461	Distal Dup	DHRSX;LOC100288477; LOC100288983	Present	Absent	Inherited
23	52886722/ 55678952	52892121/ 679111	5399	Distal Dup	XAGE-4; XAGE3	Absent	Present	Inherited
2	133022029	17	45213172	Interchrom	CDC27	Absent	Absent	*De novo*

Appendix D - Structural Variants unique to affected member of Family 2. (NCBI Build 37/hg19)

Chr	Start	End	Size (bp)	SV Type	Gene(s)
1	1142719	1143140	421	Deletion	0
1	56401293	56401614	321	Deletion	0
1	62390602	62390951	349	Deletion	INADL
1	85748634	85748964	330	Deletion	0
1	91914111	91914566	455	Deletion	0
1	207292362	207293195	833	Deletion	C4BPA
1	222034158	222034490	332	Deletion	0
1	228572175	228572504	329	Deletion	0
2	8616813	8617120	307	Deletion	0
2	16271827	16274059	2232	Deletion	0
2	39071490	39071832	342	Deletion	DHX57
2	53625772	53628009	2237	Deletion	0
2	88821957	88822489	532	Deletion	0
2	194689502	194698766	9264	Deletion	0
2	206731199	206731522	323	Deletion	0
2	226039583	226040060	477	Deletion	0
3	10397475	10397834	359	Deletion	ATP2B2
3	23764073	23764403	330	Deletion	0
3	144888449	144888910	461	Deletion	0
3	185017344	185017669	325	Deletion	0
4	3612027	3612659	632	Deletion	0
4	29661122	29661450	328	Deletion	0
4	43931496	43931837	341	Deletion	0
4	44509394	44509712	318	Deletion	0
4	58087343	58087741	398	Deletion	0
4	92280245	92285214	4969	Deletion	KIAA1680

4	100023050	100023385	335	Deletion	0
4	149990582	149992581	1999	Deletion	0
4	181549440	181549754	314	Deletion	0
4	189969695	189970049	354	Deletion	0
4	190624640	190624987	347	Deletion	0
5	42165941	42167782	1841	Deletion	0
5	45004820	45014528	9708	Deletion	0
5	68724047	68724498	451	Deletion	MARVELD2
5	93916299	93916636	337	Deletion	C5orf36
5	127172332	127173282	950	Deletion	LOC728586
5	177821921	177823857	1936	Deletion	COL23A1
5	178106931	178111943	5012	Deletion	0
6	857494	857907	413	Deletion	0
6	11556106	11556426	320	Deletion	TMEM170B
6	32624895	32625860	965	Deletion	0
6	33026728	33028610	1882	Deletion	0
6	33125920	33126843	923	Deletion	0
6	50765971	50766282	311	Deletion	0
6	57297190	57301246	4056	Deletion	PRIM2
6	106983636	106983955	319	Deletion	AIM1
6	119417519	119417847	328	Deletion	FAM184A
6	133371501	133371845	344	Deletion	0
6	134062939	134063433	494	Deletion	0
6	154446930	154447253	323	Deletion	OPRM1
7	9634650	9635993	1343	Deletion	0
7	42549714	42550036	322	Deletion	0
7	158029478	158030452	974	Deletion	PTPRN2
8	98595409	98595748	339	Deletion	0
9	71895487	71896647	1160	Deletion	0
9	87775285	87776256	971	Deletion	0

9	129520461	129523267	2806	Deletion	0
9	136379222	136379564	342	Deletion	0
10	35593222	35593552	330	Deletion	CCNY
10	57702043	57702383	340	Deletion	0
10	107253002	107253355	353	Deletion	0
10	114157801	114158139	338	Deletion	ACSL5
11	7716915	7717231	316	Deletion	OVCH2
11	31916727	31917056	329	Deletion	0
11	32286653	32286986	333	Deletion	0
11	33537687	33538010	323	Deletion	0
11	65642111	65643526	1415	Deletion	0
11	104755159	104762566	7407	Deletion	CASP12
12	124458567	124496090	37523	Deletion	ZNF664
12	130060303	130060634	331	Deletion	TMEM132D
13	23055219	23055526	307	Deletion	0
13	81747255	81747583	328	Deletion	0
13	93056847	93057422	575	Deletion	GPC5
14	35428814	35432457	3643	Deletion	0
14	38643511	38643828	317	Deletion	0
14	42486335	42486650	315	Deletion	0
14	76813774	76814209	435	Deletion	0
14	78828245	78828596	351	Deletion	0
14	100845014	100845347	333	Deletion	WDR25
15	22336853	22344093	7240	Deletion	LOC727924
15	39372537	39373645	1108	Deletion	0
15	41864762	41865200	438	Deletion	TYRO3
15	42017748	42018080	332	Deletion	MGA
15	50946438	50947428	990	Deletion	TRPM7
15	67083771	67084069	298	Deletion	0
15	83773883	83774212	329	Deletion	0

15	91173990	91174326	336	Deletion	CRTC3
15	91981580	91989268	7688	Deletion	0
15	93515798	93516111	313	Deletion	CHD2
16	12002336	12002661	325	Deletion	GSPT1
16	23910887	23911193	306	Deletion	PRKCB
16	86284321	86284629	308	Deletion	0
17	136772	138726	1954	Deletion	RPH3AL
17	193726	197057	3331	Deletion	RPH3AL
17	41382745	41399871	17126	Deletion	0
17	41383489	41466010	82521	Deletion	LOC100130581
17	46904824	46907714	2890	Deletion	0
17	47228439	47228769	330	Deletion	B4GALNT2
17	57363991	57365775	1784	Deletion	0
17	69852154	69852481	327	Deletion	0
18	1199564	1200310	746	Deletion	TFG/GPR128
18	3997022	3997479	457	Deletion	0
18	22892494	22892818	324	Deletion	ZNF521
18	27629690	27630016	326	Deletion	0
18	37061646	37061962	316	Deletion	LOC647946
18	43261014	43261339	325	Deletion	SLC14A2
18	44338433	44338770	337	Deletion	0
18	50462567	50463032	465	Deletion	DCC
19	1182661	1183179	518	Deletion	0
19	2128535	2129051	516	Deletion	AP3D1
19	45473177	45474322	1145	Deletion	CLPTM1
20	25583242	25583678	436	Deletion	0
20	33241939	33244364	2425	Deletion	PIGU
20	53292665	53292978	313	Deletion	0
21	16588380	16591454	3074	Deletion	0
21	36072707	36073562	855	Deletion	CLIC6

21	39590542	39590888	346	Deletion	0
22	33295772	33296090	318	Deletion	SYN3
22	49622515	49624301	1786	Deletion	0
X	31577478	31577794	316	Deletion	DMD
X	85605177	85609869	4692	Deletion	DACH2
X	138384911	138385237	326	Deletion	0
1	30569469	30965397	395928	Tandem Dup	0
2	242929421	242929637	216	Tandem Dup	0
3	139670597	139674454	3857	Tandem Dup	CLSTN2
4	190478831	190479023	192	Tandem Dup	0
6	160877802	160956684	78882	Tandem Dup	LOC100289195;LPA;LPAL2
7	606117	606324	207	Tandem Dup	PRKAR1B
7	156124573	156124795	222	Tandem Dup	0
9	44070664	44070787	123	Tandem Dup	LOC100289454
9	140563766	140564031	265	Tandem Dup	0
10	27224456	27229304	4848	Tandem Dup	C10orf51
13	34811550	34815591	4041	Tandem Dup	0
14	105943990	105944114	124	Tandem Dup	CRIP2
19	14706528	14706835	307	Tandem Dup	CLEC17A
X	52886720	55678950	2792230	Tandem Dup	XAGE4;ALAS2;APEX2;FAM104B;FAM120C;FAM156A;FAM156B;FGD1;FOXR2;GNL3L;GPR173;HSD17B10;HUWE1;IQSEC2;ITIH5L;KDM5C;LOC100132984;LOC100288024;LOC100288052;LOC100288498;LOC100288560;LOC644893;MAGED2;MAGEH1;PAGE2;PAGE2B;PAGE3;PAGE5;PFKFB1;PHF8;RIBC1;SMC1A;SNORA11;TRO;TSPYL2;TSR2;USP51;WNK3;XAGE3
2	16407764/ 16406392	16407988/ 16406410	224	Distal Dup	0
3	194546260/	194546429/	169	Distal Dup	0

	194543294	194543309			
7	100550585/ 100551244	100550785/ 100551247	200	Distal Dup	MUC3A
9	107816635/ 107817341	107816979/ 107817348	344	Distal Dup	0
23	17060848/ 17063276	17061023/ 17063295	175	Distal Dup	REPS2

Appendix E - Selected Regions for Sanger Sequencing. (NCBI Build 37/hg19)

Family	Chr	Position of Targeted Variant	Gene Symbol	Gene Region	Forward Primer	Reverse Primer	Size (bp)	Result
1	3	75787298	LOC100287163	Exonic	5'TTTCCACATTCG TTACATGCGTAT3'	5'GTGGAAAAACTTTT CACTGTAAGTCAC3'	511	Identical between co-twins
1	3	75787304	LOC100287163	Exonic				Identical between co-twins
1	4	85996	ZNF595	Intronic; Exonic	5'TCGTTTTACATG TCACACCTAACT3'	5'TTATGTTCGTT CAGAACTGTGG3'	199	Identical between co-twins
1	4	86004	ZNF595	Intronic; Exonic				Mosaicism?
1	4	86022	ZNF595	Intronic; Exonic				Identical between co-twins
1	4	86043	ZNF595	Intronic; Exonic				Identical between co-twins
1	11	48346662	OR4C3	Exonic	5'CTCTACATGGTT TTCTCAAGTGTG3'	5'ACAATGAGTCAG CAATGAGTTTAG3'	634	Identical between co-twins
1	12	40880542	MUC19	Exonic	5'TAGAACATCGG TTGAAGAATCA3'	5'AGCTGATGG CCGAATTGT3'	523	Mosaicism?
1	12	40880545	MUC19	Exonic				Identical between co-twins
1	13	25671607	PABPC3	Exonic	5'CCAGGCTTAC CTCACTAACG3'	5'TCCCGCAGCA TATTTATACC3'	412	Identical between co-twins
1	17	74288421	QRICH2	Exonic	5'TACCTTGCTGA CCTATTCCAG3'	5'AGATTGATGT GGTGCAACCT3'	672	Mosaicism?
1	17	74288565	QRICH2	Exonic	5'TACCTTGCTGA CCTATTCCAG3'	5'AGATTGATGT GGTGCAACCT3'	672	Identical between co-twins

<table>
<tr><td>1</td><td>22</td><td>29885594</td><td>NEFH</td><td>Exonic</td><td rowspan="2">5'GGAAAAGGCCA AGTCTCCAACG3'</td><td rowspan="2">5'TTCCTTTTTTGG GATCTCCTTCTCA3'</td><td rowspan="2">550</td><td>Identical between co-twins</td></tr>
<tr><td>1</td><td>22</td><td>29885599</td><td>NEFH</td><td>Exonic</td><td>Identical between co-twins</td></tr>
<tr><td>2</td><td>1</td><td>1650787</td><td>CDK11A/CDK11B</td><td>Exonic; 5'UTR</td><td rowspan="3">5'CTTTCAGC TAGTTTGCTC TCTCTGGTT3'</td><td rowspan="3">5'TCGTTCCTATCTG AATCATGCATTTT3'</td><td rowspan="3">516</td><td>Identical between co-twins</td></tr>
<tr><td>2</td><td>1</td><td>1650797</td><td>CDK11A/CDK11B</td><td>Splice Site; Exonic; 5'UTR</td><td>Identical between co-twins</td></tr>
<tr><td>2</td><td>1</td><td>1650801</td><td>CDK11A/CDK11B</td><td>Exonic; 5'UTR</td><td>Identical between co-twins</td></tr>
<tr><td>2</td><td>1</td><td>111957583</td><td>OVGP1</td><td>Exonic</td><td rowspan="2">5'TATGTTTCTG GAGGGGACAG3'</td><td rowspan="2">5'TGTGACCACT GGACAGAAGA3'</td><td rowspan="2">241</td><td>Identical between co-twins</td></tr>
<tr><td>2</td><td>1</td><td>111957592</td><td>OVGP1</td><td>Exonic</td><td>Identical between co-twins</td></tr>
<tr><td>2</td><td>1</td><td>144916673</td><td>PDE4DIP</td><td>Exonic</td><td>5'CCTGAGGAGTA TGGGGTAATCA3'</td><td>5'TGCCTCCAC TTCTTTGTTCC3'</td><td>239</td><td>Mosaicism?</td></tr>
<tr><td>2</td><td>8</td><td>103573011</td><td>ODF1</td><td>Exonic</td><td>5'TCAAGATCGA GTCTCCTTGC3'</td><td>5'CTTTCACACA ACACCAGCAG3'</td><td>317</td><td>Identical between co-twins</td></tr>
<tr><td>2</td><td>11</td><td>56143963</td><td>OR8U1</td><td>Exonic</td><td>5'ATGCTGGCAG TCACCATATT3'</td><td>5'CAGAGCTTCT TTCACCTCCTT3'</td><td>171</td><td>Identical between co-twins</td></tr>
<tr><td>2</td><td>19</td><td>4511197</td><td>PLIN4</td><td>Exonic</td><td>5'ATTTACGGCAC CAGTGACTC3'</td><td>5'GACCCAAAAT ATCGCAACAG3'</td><td>262</td><td>Mosaicism?</td></tr>
<tr><td>2</td><td>22</td><td>38120180</td><td>TRIOBP</td><td>Exonic</td><td>5'CTCTTCTACC CAGCAGGACA3'</td><td>5'TGTCTCGCT GGATGGTTC3'</td><td>1160</td><td>Identical between co-twins</td></tr>
</table>

Appendix F - Promoters differentially methylated between co-twins in Family 1. (NCBI Build 36/hg18)

The regions represent 330 unique genes.

Transcript	Chromosome	Region Start	Region End	*p*-value (region)	Methylation Status
DVL1	1	1275260	1275280	0.000909697	Decreased
KIAA1751	1	1901884	1901904	0.000849051	Decreased
CEP104	1	3763670	3763690	0.000606465	Decreased
DFFB	1	3763670	3763690	0.000606465	Decreased
CAMTA1	1	7687801	7687821	0.000363879	Decreased
CLSTN1	1	9807577	9807597	0.000363879	Decreased
PTCHD2	1	11461175	11461195	0.000606465	Decreased
HNRNPCP5	1	13107459	13107479	0.000849051	Decreased
KAZN	1	15122441	15122461	0.000121293	Decreased
KAZN	1	15122441	15122461	0.000121293	Decreased
RUNX3	1	25107194	25107214	0.000545818	Decreased
RUNX3	1	25107194	25107214	0.000545818	Decreased
HMGN2	1	26670728	26670748	0.000121293	Decreased
WDTC1	1	27432713	27432733	0.000909697	Decreased
RAB42	1	28790511	28790531	0.000909697	Decreased
RAB42	1	28790511	28790531	0.000909697	Decreased
OPRD1	1	29062206	29062226	0.000667111	Decreased
CSMD2	1	34403910	34403930	0.000788404	Decreased
CSMD2	1	34403910	34403930	0.000788404	Decreased
C1orf94	1	34403910	34403930	0.000788404	Decreased
TCTEX1D4	1	45046099	45046119	0.000970344	Decreased
BTBD19	1	45046099	45046119	0.000970344	Decreased
TMEM61	1	55218127	55218147	0.000788404	Decreased
MAN1A2	1	117710682	117710702	0.000909697	Increased
NBPF25P	1	147169424	147169444	0.000970344	Increased
DRD5P2	1	147169424	147169444	0.000970344	Increased

LOC101060524	1	147169424	147169444	0.000970344	Increased
HIST2H2AA3	1	148085850	148085870	0.000121293	Increased
HIST2H2AA4	1	148085850	148085870	0.000121293	Increased
HIST2H2AA3	1	148085850	148085870	0.000121293	Increased
HIST2H2AA4	1	148085850	148085870	0.000121293	Increased
HIST2H3A	1	148085850	148085870	0.000121293	Increased
HIST2H3C	1	148085850	148085870	0.000121293	Increased
HIST2H2BE	1	148125357	148125377	0.000242586	Increased
HIST2H2BE	1	148123735	148123755	0.000363879	Increased
HIST2H2AC	1	148125357	148125377	0.000242586	Increased
HIST2H2AC	1	148123735	148123755	0.000363879	Increased
HIST2H2AB	1	148125357	148125377	0.000242586	Increased
HIST2H2AB	1	148123735	148123755	0.000363879	Increased
LOC100505666	1	153301911	153301931	0.000909697	Increased
EFNA4	1	153301911	153301931	0.000909697	Increased
CD48	1	158931804	158931824	0.000545818	Increased
FCRLB	1	159963081	159963101	0.000788404	Increased
FCRLB	1	159963081	159963101	0.000788404	Increased
C1orf145	1	226467367	226467387	0.000788404	Increased
C1orf145	1	226467367	226467387	0.000788404	Increased
OBSCN	1	226570772	226570792	0.000909697	Increased
OBSCN	1	226467367	226467387	0.000788404	Increased
MYT1L	2	1920311	1920331	0.000242586	Decreased
TSSC1	2	3270302	3270322	0.000121293	Decreased
CCDC121	2	27705190	27705210	0.000849051	Decreased
CCDC121	2	27705190	27705210	0.000849051	Decreased
GPN1	2	27705190	27705210	0.000849051	Decreased
GPN1	2	27705190	27705210	0.000849051	Decreased
ZFP36L2	2	43307309	43307329	0.000727758	Decreased
LINC01126	2	43307309	43307329	0.000727758	Decreased

ABCG5	2	43919524	43919544	0.000606465	Decreased
ABCG8	2	43919524	43919544	0.000606465	Decreased
SIX2	2	45094469	45094489	0.000606465	Decreased
PSME4	2	54052386	54052406	0.000909697	Decreased
LIMS1	2	108517237	108517257	0.000667111	Decreased
ACVR2A	2	148318114	148318134	0.000909697	Increased
ACVR2A	2	148318114	148318134	0.000909697	Increased
OBSL1	2	220136462	220136482	0.000606465	Increased
SLC4A3	2	220199610	220199630	0.000363879	Increased
PRR21	2	240631724	240631744	0.000424525	Increased
OR6B3	2	240631724	240631744	0.000424525	Increased
ANO7	2	241798453	241798473	0.000727758	Increased
CAND2	3	12826659	12826679	0.000849051	Decreased
SCN5A	3	38667000	38667020	0.000849051	Decreased
SCN5A	3	38667000	38667020	0.000849051	Decreased
RPL29	3	52005903	52005923	0.000121293	Decreased
ARHGEF3	3	57068583	57068603	0.000909697	Decreased
SPATA12	3	57068583	57068603	0.000909697	Decreased
ATXN7	3	63873213	63873233	0.000909697	Decreased
ATXN7	3	63873213	63873233	0.000909697	Increased
ZNF717	3	75917850	75917870	0.000363879	Increased
ZNF717	3	75917850	75917870	0.000363879	Increased
ARGFX	3	122768542	122768562	0.000060647	Increased
ZBTB38	3	142604007	142604027	0.000303232	Increased
PCGF3	4	706619	706639	0.000242586	Decreased
UVSSA	4	1356565	1356585	0.000849051	Decreased
FGFR3	4	1774162	1774182	0.000788404	Decreased
C4orf48	4	2012702	2012722	0.000363879	Decreased
C4orf48	4	2012702	2012722	0.000363879	Decreased
MSANTD1	4	3219674	3219694	0.000667111	Decreased

C4orf17	4	100650297	100650317	0.000545818	Decreased
FRG1	4	191098042	191098062	0.000849051	Decreased
LOC283788	4	191098042	191098062	0.000849051	Increased
PDCD6	5	356365	356385	0.000667111	Decreased
AHRR	5	356365	356385	0.000667111	Decreased
SLC6A18	5	1296782	1296802	0.000909697	Decreased
TERT	5	1346428	1346448	0.000242586	Decreased
SLC6A3	5	1469888	1469908	0.000849051	Decreased
LOC100133050	5	99752762	99752782	0.000242586	Decreased
SOWAHA	5	132176644	132176664	0.000606465	Decreased
TIFAB	5	134816893	134816913	0.000242586	Decreased
DRD1	5	174804674	174804694	0.000242586	Decreased
LOC102577424	5	176103116	176103136	0.000606465	Decreased
PFN3	5	176763553	176763573	0.000303232	Increased
F12	5	176763553	176763573	0.000303232	Increased
EXOC2	6	465198	465218	0.000849051	Decreased
TUBB2B	6	3173705	3173725	0.000909697	Decreased
HIST1H2AA	6	25834901	25834952	0.000667111	Increased
HIST1H2BA	6	25834901	25834952	0.000667111	Increased
HIST1H1C	6	26164302	26164322	0.000667111	Increased
HIST1H1E	6	26264920	26264940	0.000788404	Increased
HIST1H2BD	6	26264920	26264940	0.000788404	Increased
DPCR1	6	31026612	31026632	0.000545818	Decreased
ATP6V1G2-DDX39B	6	31620675	31620695	0.000909697	Decreased
DDX39B	6	31620675	31620695	0.000909697	Decreased
SNORD84	6	31620675	31620695	0.000909697	Decreased
ATP6V1G2	6	31620675	31620695	0.000909697	Increased
NFKBIL1	6	31620675	31620695	0.000909697	Increased
NFKBIL1	6	31620675	31620695	0.000909697	Increased
VEGFA	6	43845650	43845670	0.000727758	Increased

VEGFA	6	43845650	43845670	0.000727758	Increased
CD2AP	6	47552558	47552578	0.000667111	Increased
DDO	6	110828027	110828047	0.000363879	Increased
ECHDC1	6	127706751	127706771	0.000242586	Increased
ECHDC1	6	127706751	127706771	0.000242586	Increased
WTAP	6	160067450	160067470	0.000788404	Increased
WTAP	6	160067450	160067470	0.000788404	Increased
RPS6KA2-AS1	6	167234121	167234141	0.000545818	Increased
TCP10	6	167718893	167718913	0.000667111	Increased
FAM120B	6	170528419	170528439	0.000121293	Increased
FAM120B	6	170528419	170528439	0.000121293	Increased
FAM120B	6	170528419	170528439	0.000121293	Increased
PRKAR1B	7	717528	717548	0.000667111	Decreased
PRKAR1B	7	717528	717548	0.000667111	Decreased
PRKAR1B	7	717528	717548	0.000667111	Decreased
PRKAR1B	7	717528	717548	0.000667111	Decreased
PRKAR1B	7	717528	717548	0.000667111	Decreased
C7orf50	7	1122890	1122910	0.000121293	Decreased
TMEM184A	7	1557219	1557239	0.000424525	Increased
MAD1L1	7	1877342	1877362	0.000545818	Increased
MAD1L1	7	2056253	2056273	0.000121293	Increased
TTYH3	7	2664585	2664605	0.000606465	Decreased
GRID2IP	7	6510144	6510164	0.000970344	Increased
SCIN	7	12594747	12594767	0.000363879	Increased
SCIN	7	12594747	12594767	0.000363879	Increased
MEOX2	7	15693738	15693758	0.000849051	Increased
LOC101927524	7	15693738	15693758	0.000849051	Increased
HIBADH	7	27670032	27670052	0.000363879	Increased
MYO1G	7	44986098	44986118	0.000606465	Increased
HIP1	7	75106363	75106383	0.000667111	Increased

SND1	7	127423872	127423892	0.000849051	Increased
SND1-IT1	7	127423872	127423892	0.000849051	Increased
CNPY1	7	154995456	154995476	0.000242586	Increased
PTPRN2	7	157352628	157352648	0.000909697	Increased
PTPRN2	7	157941351	157941371	0.000181939	Increased
ERICH1-AS1	8	972376	972396	0.000121293	Decreased
ERICH1-AS1	8	1040976	1040996	0.000242586	Decreased
ERICH1-AS1	8	977324	977344	0.000242586	Decreased
ERICH1-AS1	8	780350	780370	0.000363879	Decreased
ERICH1-AS1	8	905345	905365	0.000242586	Decreased
MCPH1	8	6411949	6411969	0.000667111	Decreased
ANGPT2	8	6411949	6411969	0.000667111	Decreased
FAM66D	8	12025118	12025138	0.000545818	Decreased
FAM66D	8	12034583	12034603	0.000849051	Decreased
LOC392196	8	12025118	12025138	0.000545818	Decreased
USP17L2	8	12034583	12034603	0.000849051	Decreased
TACC1	8	38733972	38733992	0.000606465	Decreased
ADAM5	8	39290413	39290433	0.000970344	Decreased
UTP23	8	117846997	117847017	0.000727758	Decreased
TRAPPC9	8	141428560	141428580	0.000909697	Decreased
TRAPPC9	8	141428560	141428580	0.000909697	Decreased
SLC45A4	8	142332019	142332039	0.000545818	Decreased
PTP4A3	8	142509287	142509307	0.000909697	Decreased
TSNARE1	8	143405544	143405564	0.000485172	Decreased
PSCA	8	143757951	143757971	0.000788404	Increased
PSCA	8	143757951	143757971	0.000788404	Increased
SLURP1	8	143821736	143821756	0.000121293	Increased
LOC100133669	8	144171219	144171239	0.000909697	Increased
LY6E	8	144171219	144171239	0.000909697	Increased
ZFP41	8	144409425	144409445	0.000121293	Increased

FAM83H	8	144881577	144881597	0.000727758	Increased
SCRIB	8	144965172	144965192	0.000606465	Increased
SCRIB	8	144959364	144959384	0.000909697	Increased
MIR937	8	144965172	144965192	0.000606465	Increased
PGM5-AS1	9	70162095	70162115	0.000363879	Decreased
PGM5	9	70162095	70162115	0.000363879	Decreased
C9orf3	9	96807113	96807133	0.000727758	Decreased
C9orf3	9	96807113	96807133	0.000727758	Decreased
FSD1L	9	107249215	107249235	0.000909697	Decreased
ZBTB43	9	128635281	128635301	0.000060647	Decreased
GPR107	9	131854890	131854910	0.000909697	Decreased
RXRA	9	136440826	136440846	0.000667111	Decreased
GPSM1	9	138356958	138356978	0.000909697	Decreased
GPSM1	9	138351213	138351233	0.000909697	Increased
INPP5E	9	138448648	138448668	0.000485172	Increased
MAN1B1	9	139117758	139117778	0.000363879	Increased
FAM166A	9	139263786	139263806	0.000727758	Increased
C9orf173	9	139263786	139263806	0.000727758	Increased
CACNA1B	9	139936085	139936105	0.000849051	Increased
DIP2C	10	620619	620639	0.000545818	Decreased
DIP2C	10	684962	684982	0.000849051	Decreased
DIP2C	10	382685	382705	0.000363879	Decreased
PRR26	10	684962	684982	0.000849051	Decreased
PRR26	10	684962	684982	0.000849051	Decreased
ADARB2	10	1732764	1732784	0.000242586	Increased
PRF1	10	72030426	72030446	0.000909697	Increased
GPR123	10	134776326	134776346	0.000242586	Increased
GPR123	10	134776326	134776346	0.000242586	Increased
PKP3	11	383291	383311	0.000727758	Decreased
C11orf35	11	545626	545646	0.000909697	Decreased

H19	11	1974422	1974442	0.000667111	Decreased
MIR675	11	1974422	1974442	0.000667111	Decreased
INS-IGF2	11	2139920	2139940	0.000606465	Decreased
INS	11	2139920	2139940	0.000606465	Decreased
HIPK3	11	33235017	33235037	0.000909697	Decreased
HIPK3	11	33235017	33235037	0.000909697	Decreased
HIPK3	11	33235017	33235037	0.000909697	Increased
TTC17	11	43336141	43336161	0.000909697	Increased
MDK	11	46358724	46358744	0.000909697	Increased
MDK	11	46358724	46358744	0.000909697	Increased
MDK	11	46358724	46358744	0.000909697	Increased
MDK	11	46358724	46358744	0.000909697	Increased
OR5I1	11	55461357	55461377	0.000788404	Increased
NRXN2	11	64168118	64168138	0.000363879	Increased
NRXN2	11	64168118	64168138	0.000363879	Increased
KCNK7	11	65123625	65123645	0.000727758	Increased
MAP3K11	11	65123625	65123645	0.000727758	Increased
CARNS1	11	66945049	66945069	0.000181939	Increased
FAM86C2P	11	67330288	67330308	0.000727758	Increased
MIR3649	12	1642474	1642494	0.000788404	Decreased
SPX	12	21571952	21571972	0.000727758	Increased
ASCL4	12	106693398	106693418	0.000667111	Increased
ACADS	12	119661601	119661621	0.000727758	Increased
HNF1A	12	119900006	119900026	0.000303232	Increased
EP400	12	131086579	131086599	0.000667111	Increased
ANKLE2	12	131819474	131819494	0.000242586	Increased
DGKH	13	41512904	41512924	0.000909697	Increased
TFDP1	13	113308369	113308389	0.000849051	Increased
TFDP1	13	113308369	113308389	0.000849051	Increased
TMEM255B	13	113606174	113606194	0.000606465	Increased

RASA3	13	113824201	113824221	0.000485172	Increased
RASA3	13	113856967	113856987	0.000242586	Increased
HNRNPC	14	20808383	20808403	0.000788404	Increased
ZBTB25	14	64040665	64040685	0.000727758	Increased
ZBTB1	14	64040665	64040685	0.000727758	Increased
SERPINA12	14	94054839	94054859	0.000363879	Increased
C14orf180	14	104116175	104116195	0.000667111	Increased
C14orf180	14	104116175	104116195	0.000667111	Increased
C14orf80	14	105036181	105036201	0.000363879	Increased
C14orf80	14	105036181	105036201	0.000363879	Increased
C14orf80	14	105036181	105036201	0.000363879	Increased
SNORD116-8	15	22868070	22868090	0.000121293	Increased
SNORD116-3	15	22868070	22868090	0.000121293	Increased
SNORD116-9	15	22868070	22868090	0.000121293	Increased
SNORD116-10	15	22868070	22868090	0.000121293	Increased
SNORD116-11	15	22868070	22868090	0.000121293	Increased
SNORD115-10	15	22983806	22983826	0.000788404	Increased
SNORD115-12	15	22983806	22983826	0.000788404	Increased
SNORD115-9	15	22983806	22983826	0.000788404	Increased
SNORD115-5	15	22983806	22983826	0.000788404	Increased
SNORD115-10	15	22983806	22983826	0.000788404	Increased
SNORD115-11	15	22983806	22983826	0.000788404	Increased
SNORD115-29	15	22983806	22983826	0.000788404	Increased
SNORD115-36	15	22983806	22983826	0.000788404	Increased
SNORD115-43	15	22983806	22983826	0.000788404	Increased
SNORD115-10	15	22983806	22983826	0.000788404	Increased
SNORD115-12	15	22983806	22983826	0.000788404	Increased
SNORD115-9	15	22983806	22983826	0.000788404	Increased
SNORD115-5	15	22983806	22983826	0.000788404	Increased
SNORD115-33	15	23030052	23030072	0.000545818	Increased

SNORD115-34	15	23030052	23030072	0.000545818	Increased
SNORD115-35	15	23030052	23030072	0.000545818	Increased
SNORD115-11	15	23030052	23030072	0.000545818	Increased
SNORD115-29	15	23030052	23030072	0.000545818	Increased
SNORD115-36	15	23030052	23030072	0.000545818	Increased
SNORD115-43	15	23030052	23030072	0.000545818	Increased
SNORD115-37	15	23030052	23030072	0.000545818	Increased
GABRB3	15	24425395	24425415	0.000788404	Decreased
GABRB3	15	24425395	24425415	0.000788404	Decreased
GABRB3	15	24425395	24425415	0.000788404	Decreased
GABRB3	15	24425395	24425415	0.000788404	Decreased
OCA2	15	25689155	25689175	0.000363879	Decreased
BAHD1	15	38520115	38520135	0.000363879	Decreased
RHOV	15	38954636	38954656	0.000727758	Decreased
CALML4	15	66286407	66286427	0.000667111	Decreased
CLN6	15	66286407	66286427	0.000667111	Decreased
CELF6	15	70369057	70369077	0.000909697	Decreased
CELF6	15	70369057	70369077	0.000909697	Increased
CELF6	15	70369057	70369077	0.000909697	Increased
CD276	15	71762795	71762815	0.000909697	Increased
MTHFS	15	77977330	77977350	0.000909697	Increased
MTHFS	15	77977330	77977350	0.000909697	Increased
MTHFS	15	77977330	77977350	0.000909697	Increased
ST20-MTHFS	15	77977330	77977350	0.000909697	Increased
ZNF710	15	88345120	88345140	0.000606465	Increased
IGF1R	15	97009952	97009972	0.000909697	Increased
PRKXP1	15	98914519	98914539	0.000242586	Increased
RAB40C	16	604302	604322	0.000424525	Decreased
RAB40C	16	604302	604322	0.000424525	Decreased
RAB40C	16	604302	604322	0.000424525	Decreased

RAB40C	16	604302	604322	0.000424525	Decreased
WDR90	16	657369	657389	0.000909697	Decreased
RHOT2	16	657369	657389	0.000909697	Decreased
CHTF18	16	773832	773852	0.000667111	Decreased
TRAF7	16	2155184	2155204	0.000121293	Decreased
ZSCAN10	16	3079775	3079795	0.000909697	Decreased
ZSCAN10	16	3079775	3079795	0.000909697	Decreased
SNX29	16	12118466	12118486	0.000363879	Decreased
PDILT	16	20325935	20325955	0.000667111	Decreased
ACSM5	16	20325935	20325955	0.000667111	Decreased
TP53TG3	16	32596474	32596494	0.000667111	Decreased
TP53TG3B	16	32596474	32596494	0.000667111	Decreased
TP53TG3C	16	32596474	32596494	0.000667111	Decreased
TP53TG3	16	33111540	33111560	0.000909697	Decreased
TP53TG3B	16	33111540	33111560	0.000909697	Decreased
TP53TG3C	16	33111540	33111560	0.000909697	Decreased
TP53TG3	16	33168075	33168095	0.000424525	Decreased
TP53TG3B	16	33168075	33168095	0.000424525	Decreased
TP53TG3C	16	33168075	33168095	0.000424525	Decreased
CTCF	16	66153279	66153299	0.000667111	Increased
BANP	16	86591053	86591073	0.000121293	Increased
BANP	16	86591053	86591073	0.000121293	Increased
BANP	16	86591053	86591073	0.000121293	Increased
IL17C	16	87228505	87228525	0.000667111	Increased
CBFA2T3	16	87489347	87489367	0.000667111	Increased
CBFA2T3	16	87489347	87489367	0.000667111	Increased
SLC22A31	16	87795446	87795466	0.000909697	Increased
CPNE7	16	88174559	88174579	0.000424525	Increased
SPIRE2	16	88428574	88428594	0.000242586	Increased
TCF25	16	88488219	88488239	0.000606465	Increased

RPH3AL	17	206061	206081	0.000545818	Decreased
FLJ36000	17	21827263	21827283	0.000485172	Decreased
SRCIN1	17	33968132	33968152	0.000606465	Decreased
CNTNAP1	17	38092471	38092491	0.000788404	Decreased
VAT1	17	38429338	38429358	0.000909697	Decreased
RND2	17	38429338	38429358	0.000909697	Decreased
ATXN7L3	17	39632047	39632067	0.000909697	Decreased
ARL17A	17	41944466	41944486	0.000545818	Decreased
LRRC37A2	17	41944466	41944486	0.000545818	Decreased
WNT3	17	42251986	42252006	0.000667111	Decreased
SP6	17	43280191	43280211	0.000363879	Decreased
SP6	17	43280191	43280211	0.000363879	Decreased
NGFR	17	44927754	44927774	0.000667111	Decreased
TEX2	17	59695170	59695190	0.000667111	Decreased
RPTOR	17	76132310	76132330	0.000909697	Decreased
CHMP6	17	76579427	76579447	0.000727758	Increased
LDLRAD4	18	13489230	13489250	0.000363879	Decreased
LDLRAD4	18	13252790	13252810	0.000363879	Decreased
LDLRAD4	18	13489230	13489250	0.000363879	Decreased
KATNAL2	18	42804957	42804977	0.000242586	Decreased
KATNAL2	18	42810872	42810892	0.000545818	Decreased
KATNAL2	18	42798912	42798932	0.000303232	Decreased
TCEB3CL	18	42798912	42798932	0.000303232	Decreased
TCEB3CL2	18	42798912	42798932	0.000303232	Decreased
TCEB3C	18	42804957	42804977	0.000242586	Decreased
TCEB3CL	18	42804957	42804977	0.000242586	Decreased
TCEB3CL2	18	42804957	42804977	0.000242586	Decreased
TCEB3C	18	42810872	42810892	0.000545818	Increased
TCEB3CL	18	42810872	42810892	0.000545818	Increased
RAB27B	18	50645912	50645932	0.000545818	Increased

BCL2	18	59136592	59136612	0.000909697	Increased
SALL3	18	74840102	74840122	0.000667111	Increased
NFATC1	18	75281489	75281509	0.000667111	Increased
NFATC1	18	75281489	75281509	0.000667111	Increased
CTDP1	18	75555426	75555446	0.000788404	Increased
CTDP1	18	75555426	75555446	0.000788404	Increased
PQLC1	18	75777134	75777154	0.000303232	Increased
FSTL3	19	631247	631267	0.000667111	Decreased
HMHA1	19	1017228	1017248	0.000909697	Decreased
HMHA1	19	1017228	1017248	0.000909697	Decreased
HMHA1	19	1017228	1017248	0.000909697	Decreased
DAZAP1	19	1371068	1371088	0.000303232	Decreased
MKNK2	19	1989481	1989501	0.000121293	Decreased
AP3D1	19	2066512	2066532	0.000667111	Decreased
KDM4B	19	4995039	4995059	0.000363879	Decreased
PTPRS	19	5244600	5244620	0.000606465	Decreased
STXBP2	19	7618186	7618206	0.000121293	Increased
ZNF564	19	12527452	12527472	0.000363879	Increased
FCHO1	19	17722410	17722430	0.000727758	Increased
FCHO1	19	17722410	17722430	0.000727758	Increased
MAST3	19	18095801	18095821	0.000909697	Increased
POP4	19	34793172	34793192	0.000909697	Increased
PLEKHF1	19	34847241	34847261	0.000242586	Increased
ZNF829	19	42099041	42099061	0.000788404	Increased
ZNF829	19	42099041	42099061	0.000788404	Increased
ZNF568	19	42099041	42099061	0.000788404	Increased
PPP1R13L	19	50591427	50591447	0.000909697	Increased
PPP1R13L	19	50591427	50591447	0.000909697	Increased
RPL13A	19	54683534	54683554	0.000849051	Increased
RPL13AP5	19	54683534	54683554	0.000849051	Increased

SNORD32A	19	54683534	54683554	0.000849051	Increased
SNORD33	19	54683534	54683554	0.000849051	Increased
SNORD34	19	54683534	54683554	0.000849051	Increased
SNORD35A	19	54683534	54683554	0.000849051	Increased
PRR12	19	54789900	54789920	0.000909697	Increased
SHANK1	19	55881768	55881788	0.000606465	Increased
FIZ1	19	60805831	60805851	0.000667111	Increased
ZNF524	19	60805831	60805851	0.000667111	Increased
NLRP4	19	61061344	61061364	0.000121293	Increased
ZNF552	19	63018998	63019018	0.000788404	Increased
MIR4754	19	63589522	63589542	0.000909697	Increased
RPS5	19	63589522	63589542	0.000909697	Increased
BMP2	20	6695915	6695935	0.000363879	Decreased
RRBP1	20	17611434	17611454	0.000727758	Decreased
FAM182B	20	25725783	25725803	0.000485172	Decreased
FAM182B	20	25725783	25725803	0.000485172	Decreased
FRG1B	20	28224614	28224634	0.000545818	Decreased
EPB41L1	20	34143906	34143926	0.000606465	Decreased
EPB41L1	20	34143906	34143926	0.000606465	Decreased
GNAS	20	56879799	56879819	0.000121293	Increased
GNAS	20	56879799	56879819	0.000121293	Increased
CABLES2	20	60405040	60405060	0.000121293	Decreased
LOC63930	20	61145881	61145901	0.000242586	Decreased
LINC01056	20	61145881	61145901	0.000242586	Decreased
YTHDF1	20	61318845	61318865	0.000667111	Decreased
ARFGAP1	20	61389848	61389868	0.000606465	Decreased
ARFGAP1	20	61389848	61389868	0.000606465	Decreased
MIR4326	20	61389848	61389868	0.000606465	Decreased
COL20A1	20	61394057	61394077	0.000667111	Decreased
ZBTB46	20	61848829	61848849	0.000909697	Decreased

LINC00310	21	34473922	34473942	0.000303232	Decreased
MORC3	21	36613431	36613451	0.000667111	Decreased
LINC00111	21	41971605	41971625	0.000606465	Decreased
DNMT3L	21	44507432	44507452	0.000788404	Decreased
PCNT	21	46628030	46628050	0.000121293	Decreased
ATP6V1E1	22	16492493	16492513	0.000909697	Decreased
C22orf34	22	48411744	48411764	0.000667111	Decreased
C22orf34	22	48395329	48395349	0.000242586	Decreased
TRABD	22	48965689	48965709	0.000667111	Increased
SHOX	X	523098	523118	0.000909697	Decreased
XAGE1A	X	52277169	52277189	0.000849051	Decreased
XAGE1A	X	52277169	52277189	0.000849051	Decreased
XAGE1A	X	52277169	52277189	0.000849051	Decreased
XAGE1B	X	52277169	52277189	0.000849051	Decreased
XAGE1B	X	52277169	52277189	0.000849051	Decreased
XAGE1B	X	52277169	52277189	0.000849051	Decreased
XAGE1C	X	52277169	52277189	0.000849051	Decreased
XAGE1C	X	52277169	52277189	0.000849051	Decreased
XAGE1C	X	52277169	52277189	0.000849051	Decreased
XAGE1D	X	52277169	52277189	0.000849051	Decreased
XAGE1D	X	52277169	52277189	0.000849051	Decreased
XAGE1D	X	52277169	52277189	0.000849051	Increased
XAGE1E	X	52277169	52277189	0.000849051	Increased
XAGE1E	X	52277169	52277189	0.000849051	Increased
XAGE1E	X	52277169	52277189	0.000849051	Increased
XAGE1A	X	52550109	52550129	0.000970344	Increased
XAGE1A	X	52550109	52550129	0.000970344	Increased
XAGE1A	X	52550109	52550129	0.000970344	Increased
XAGE1B	X	52550109	52550129	0.000970344	Increased
XAGE1B	X	52550109	52550129	0.000970344	Increased

XAGE1B	X	52550109	52550129	0.000970344	Increased
XAGE1C	X	52550109	52550129	0.000970344	Increased
XAGE1C	X	52550109	52550129	0.000970344	Increased
XAGE1C	X	52550109	52550129	0.000970344	Increased
XAGE1D	X	52550109	52550129	0.000970344	Increased
XAGE1D	X	52550109	52550129	0.000970344	Increased
XAGE1D	X	52550109	52550129	0.000970344	Increased
XAGE1E	X	52550109	52550129	0.000970344	Increased
XAGE1E	X	52550109	52550129	0.000970344	Increased
XAGE1E	X	52550109	52550129	0.000970344	Increased
XAGE1A	X	52563003	52563023	0.000545818	Increased
XAGE1A	X	52563003	52563023	0.000545818	Increased
XAGE1A	X	52563003	52563023	0.000545818	Increased
XAGE1B	X	52563003	52563023	0.000545818	Increased
XAGE1B	X	52563003	52563023	0.000545818	Increased
XAGE1B	X	52563003	52563023	0.000545818	Increased
XAGE1C	X	52563003	52563023	0.000545818	Increased
XAGE1C	X	52563003	52563023	0.000545818	Increased
XAGE1C	X	52563003	52563023	0.000545818	Increased
XAGE1D	X	52563003	52563023	0.000545818	Increased
XAGE1D	X	52563003	52563023	0.000545818	Increased
XAGE1D	X	52563003	52563023	0.000545818	Increased
XAGE1E	X	52563003	52563023	0.000545818	Increased
XAGE1E	X	52563003	52563023	0.000545818	Increased
XAGE1E	X	52563003	52563023	0.000545818	Increased
XIST	X	72974756	72974776	0.000121293	Increased
H2BFWT	X	103153633	103153653	0.000242586	Increased
IL1RAPL2	X	104352919	104352939	0.000849051	Increased
TEX13A	X	104352919	104352939	0.000849051	Increased
SPANXN1	X	144135873	144135893	0.000909697	Increased

HCFC1	X	152873319	152873339	0.000606465	Increased

Appendix G - Promoters differentially methylated between co-twins in Family 2. (NCBI Build 36/hg18)

The regions represent 138 unique genes.

Transcript	Chromosome	Region Start	Region End	*p*-value (region)	Methylation Status
HIST2H2AA3	1	148085850	148085870	0.000303232	Increased
HIST2H2AA4	1	148085850	148085870	0.000303232	Increased
HIST2H2AA3	1	148085850	148085870	0.000303232	Increased
HIST2H2AA4	1	148085850	148085870	0.000303232	Increased
HIST2H3A	1	148085850	148085870	0.000303232	Increased
HIST2H3C	1	148085850	148085870	0.000303232	Increased
AK2	1	33275973	33275993	0.000485172	Decreased
CNTN2	1	203305581	203305601	0.000606465	Decreased
GCSAML	1	245778148	245778168	0.000667111	Increased
GCSAML	1	245778148	245778168	0.000667111	Increased
GCSAML	1	245778148	245778168	0.000667111	Increased
LINC00466	1	63559010	63559030	0.000727758	Increased
FOXD3	1	63559010	63559030	0.000727758	Increased
PRDM16	1	3061919	3061939	0.000788404	Increased
TBX19	1	166515976	166515996	0.000849051	Increased
CMPK1	1	47571304	47571324	0.000909697	Increased
KCNA3	1	111019682	111019702	0.000909697	Increased
KCNA3	1	111019682	111019702	0.000909697	Increased
KCNA3	1	111019682	111019702	0.000909697	Increased
SLC41A1	1	204049689	204049709	0.000909697	Increased
DIEXF	1	208067030	208067050	0.000909697	Increased
PIK3CD	1	9638136	9638156	0.000970344	Increased
C1orf200	1	9638136	9638156	0.000970344	Increased
KDM4A-AS1	1	43944879	43944899	0.000970344	Increased
ST3GAL3	1	43944879	43944899	0.000970344	Increased
FAM98A	2	33678771	33678791	0.000727758	Decreased

WASH2P	2	114057003	114057023	0.000849051	Increased
LYG1	2	99284976	99284996	0.000909697	Increased
PTPN18	2	130846574	130846594	0.000909697	Increased
ATP6V1B1	2	71015580	71015600	0.000970344	Increased
SLC25A36	3	142142590	142142610	0.000485172	Decreased
THUMPD3-AS1	3	9414278	9414298	0.000788404	Increased
SETD5	3	9414278	9414298	0.000788404	Increased
NKIRAS1	3	23933581	23933601	0.000909697	Increased
RPL15	3	23933581	23933601	0.000909697	Increased
RPL15	3	23933581	23933601	0.000909697	Increased
RPL15	3	23933581	23933601	0.000909697	Increased
RPL15	3	23933581	23933601	0.000909697	Increased
MUC4	3	196975464	196975484	0.000909697	Increased
LHFPL4	3	9569188	9569208	0.000970344	Increased
C3orf18	3	50581039	50581059	0.000970344	Increased
C3orf18	3	50581039	50581059	0.000970344	Increased
HEMK1	3	50581039	50581059	0.000970344	Increased
ALCAM	3	106567543	106567563	0.000970344	Increased
SORBS2	4	187115769	187115789	0.000667111	Decreased
TMEM154	4	153821546	153821566	0.000849051	Increased
SNX25	4	186367353	186367373	0.000970344	Increased
SMAD5-AS1	5	135499383	135499403	0.000303232	Decreased
SMAD5	5	135499383	135499403	0.000303232	Decreased
COMMD10	5	115447700	115447720	0.000545818	Increased
TENM2	5	166643495	166643515	0.000909697	Increased
TRIM27	6	29000548	29000568	0.000667111	Decreased
HIST1H1C	6	26164302	26164322	0.000909697	Increased
PACSIN1	6	34607414	34607434	0.000970344	Increased
PACSIN1	6	34607414	34607434	0.000970344	Increased
ENPP5	6	46247581	46247601	0.000970344	Increased

PPIL4	6	149909769	149909789	0.000970344	Increased
TTYH3	7	2653547	2653567	0.000667111	Increased
PTPRN2	7	157141154	157141174	0.000667111	Increased
KCTD7	7	65730399	65730419	0.000788404	Increased
TBRG4	7	45112319	45112339	0.000909697	Increased
SNORA5A	7	45112319	45112339	0.000909697	Increased
SNORA5C	7	45112319	45112339	0.000909697	Increased
SNORA5B	7	45112319	45112339	0.000909697	Increased
DNAJB6	7	156895498	156895518	0.000909697	Increased
UNCX	7	1242530	1242550	0.000970344	Increased
GTF2I	7	73709238	73709258	0.000970344	Increased
PLEC	8	145068180	145068200	0.000727758	Decreased
PLEC	8	145068180	145068200	0.000727758	Decreased
PLEC	8	145068180	145068200	0.000727758	Decreased
PLEC	8	145068180	145068200	0.000727758	Decreased
PLEC	8	145068180	145068200	0.000727758	Decreased
PLEC	8	145068180	145068200	0.000727758	Decreased
PLEC	8	145068180	145068200	0.000727758	Decreased
PLEC	8	145068180	145068200	0.000727758	Increased
RPL23AP53	8	169134	169154	0.000788404	Increased
ZNF596	8	169134	169154	0.000788404	Increased
ZNF596	8	169134	169154	0.000788404	Increased
ZNF596	8	169134	169154	0.000788404	Increased
ZNF596	8	169134	169154	0.000788404	Increased
ZNF596	8	169134	169154	0.000788404	Increased
ZFP41	8	144403919	144403939	0.000909697	Increased
ZFP41	8	144403919	144403939	0.000909697	Increased
ADCK5	8	145567737	145567757	0.000909697	Increased
HEMGN	9	99747860	99747880	0.000485172	Decreased
ZNF483	9	113326342	113326362	0.000849051	Decreased

SOHLH1	9	137732516	137732536	0.000909697	Increased
KCNT1	9	137732516	137732536	0.000909697	Increased
PDCD4-AS1	10	112621108	112621128	0.000242586	Decreased
PDCD4	10	112621108	112621128	0.000242586	Decreased
C10orf111	10	15179245	15179265	0.000424525	Decreased
RPP38	10	15179245	15179265	0.000424525	Decreased
BEND3P3	10	81111811	81111831	0.000485172	Decreased
RSU1	10	16900364	16900384	0.000909697	Decreased
TEX36	10	127362608	127362628	0.000909697	Increased
LOC283038	10	127362608	127362628	0.000909697	Increased
INPP5F	10	121474673	121474693	0.000970344	Increased
ADAM8	10	134941267	134941287	0.000970344	Increased
OR5B17	11	57884023	57884043	0.000060600	Decreased
CAPN5	11	76454714	76454734	0.000485172	Decreased
OR52N2	11	5797216	5797236	0.000545818	Increased
OR52N4	11	5731573	5731593	0.000788404	Increased
EIF3F	11	7964528	7964548	0.000909697	Increased
PANX1	11	93500851	93500871	0.000909697	Increased
PUS3	11	125278893	125278913	0.000909697	Increased
DDX25	11	125278893	125278913	0.000909697	Increased
UBTFL1	11	89457840	89457860	0.000970344	Increased
STAT6	12	55792333	55792353	0.000970344	Increased
STAT6	12	55792333	55792353	0.000970344	Increased
ESD	13	46270273	46270293	0.000242586	Decreased
GAS6-AS2	13	113546285	113546305	0.000909697	Increased
GAS6	13	113546285	113546305	0.000909697	Increased
BTF3P11	13	76399660	76399680	0.000970344	Increased
ADAM21	14	69993044	69993064	0.000060600	Decreased
OR4E2	14	21202211	21202231	0.000545818	Increased
MOK	14	101842189	101842209	0.000909697	Increased

FOXA1	14	37138067	37138087	0.000970344	Increased
TICRR	15	87918941	87918961	0.000121293	Decreased
SNORD115-33	15	23030052	23030072	0.000181939	Increased
SNORD115-34	15	23030052	23030072	0.000181939	Increased
SNORD115-35	15	23030052	23030072	0.000181939	Increased
SNORD115-11	15	23030052	23030072	0.000181939	Increased
SNORD115-29	15	23030052	23030072	0.000181939	Increased
SNORD115-36	15	23030052	23030072	0.000181939	Increased
SNORD115-43	15	23030052	23030072	0.000181939	Increased
SNORD115-37	15	23030052	23030072	0.000181939	Increased
SNORD116-8	15	22868070	22868090	0.000242586	Increased
SNORD116-3	15	22868070	22868090	0.000242586	Increased
SNORD116-9	15	22868070	22868090	0.000242586	Increased
SNORD116-10	15	22868070	22868090	0.000242586	Increased
SNORD116-11	15	22868070	22868090	0.000242586	Increased
BUB1B	15	38295995	38296015	0.000242586	Decreased
PAK6	15	38295995	38296015	0.000242586	Decreased
SNORD115-10	15	22983806	22983826	0.000303232	Increased
SNORD115-12	15	22983806	22983826	0.000303232	Increased
SNORD115-9	15	22983806	22983826	0.000303232	Increased
SNORD115-5	15	22983806	22983826	0.000303232	Increased
SNORD115-10	15	22983806	22983826	0.000303232	Increased
SNORD115-11	15	22983806	22983826	0.000303232	Increased
SNORD115-29	15	22983806	22983826	0.000303232	Increased
SNORD115-36	15	22983806	22983826	0.000303232	Increased
SNORD115-43	15	22983806	22983826	0.000303232	Increased
SNORD115-10	15	22983806	22983826	0.000303232	Increased
SNORD115-12	15	22983806	22983826	0.000303232	Increased
SNORD115-9	15	22983806	22983826	0.000303232	Increased
SNORD115-5	15	22983806	22983826	0.000303232	Increased

DAPK2	15	62126479	62126499	0.000849051	Increased
RBFOX1	16	6762885	6762905	0.000909697	Increased
RBFOX1	16	6762885	6762905	0.000909697	Increased
EXOC3L1	16	65782578	65782598	0.000970344	Increased
E2F4	16	65782578	65782598	0.000970344	Increased
MIR132	17	1904363	1904383	0.000545818	Increased
MIR212	17	1904363	1904383	0.000545818	Increased
HIC1	17	1904363	1904383	0.000545818	Increased
HIC1	17	1904363	1904383	0.000545818	Increased
FUT2	19	53890407	53890427	0.000303232	Decreased
ZNF226	19	49360408	49360428	0.000485172	Decreased
ZNF226	19	49360408	49360428	0.000485172	Decreased
ZNF226	19	49360408	49360428	0.000485172	Increased
SLC17A7	19	54637517	54637537	0.000788404	Increased
S1PR2	19	10203572	10203592	0.000970344	Increased
MIR4322	19	10203572	10203592	0.000970344	Increased
GNAS	20	56879799	56879819	0.000181939	Decreased
GNAS	20	56879799	56879819	0.000181939	Decreased
SYS1-DBNDD2	20	43468281	43468301	0.000909697	Increased
DBNDD2	20	43468281	43468301	0.000909697	Increased
DBNDD2	20	43468281	43468301	0.000909697	Increased
DBNDD2	20	43468281	43468301	0.000909697	Increased
DBNDD2	20	43468281	43468301	0.000909697	Increased
DBNDD2	20	43468281	43468301	0.000909697	Increased
DBNDD2	20	43468281	43468301	0.000909697	Increased
GSTTP1	22	22678163	22678183	0.000970344	Increased
HPS4	22	25208590	25208610	0.000970344	Increased
HPS4	22	25208590	25208610	0.000970344	Increased
SRRD	22	25208590	25208610	0.000970344	Increased
PDGFB	22	37969786	37969806	0.000970344	Increased

PDGFB	22	37969786	37969806	0.000970344	Increased
XIST	X	72974756	72974776	0.000424525	Decreased
BCOR	X	39846274	39846294	0.000545818	Increased
BCOR	X	39846274	39846294	0.000545818	Increased
TIMM8A	X	100491393	100491413	0.000909697	Increased
TIMM8A	X	100491393	100491413	0.000909697	Increased
BTK	X	100491393	100491413	0.000909697	Increased
BTK	X	100491393	100491413	0.000909697	Increased
DGKK	X	50231382	50231402	0.000970344	Increased
AIFM1	X	129100601	129100621	0.000970344	Increased
AIFM1	X	129100601	129100621	0.000970344	Increased

Appendix H - *De novo* sequencing variations in the affected twin of Family 1 that show co-localized differential methylation and sequence variation. (NCBI Build 37/hg19)

Chr	Position	Gene	Reference Allele (Build 37)	Sample Allele	Variation Type	Gene Region	Impact	SIFT Function Prediction	SIFT Score	PolyPhen-2 Function Prediction	dbSNP
2	44095621	*ABCG8*	A	G	SNV	Intronic					13395859
16	20440701	*ACSM5*		ATCA	Insertion	Intronic					397733540
10	1255477	*ADARB2*	G	A	SNV	Intronic					114312758
10	1255488	*ADARB2*	TGCCA	CATCG	Substitution	Intronic					
10	1258701	*ADARB2*	G	A	SNV	Intronic					113359529
10	1258809	*ADARB2*	A	C	SNV	Intronic					11250339
10	1266123	*ADARB2*	GTGG	A	Substitution	Intronic					
10	1266131	*ADARB2*		GTA	Insertion	Intronic					
10	1377512	*ADARB2*	G	A	SNV	Intronic					12250711
10	1415954	*ADARB2*	AT	GC	Substitution	Intronic					386739833
10	1418159	*ADARB2*		GC	Insertion	Intronic					60001440
10	1486550	*ADARB2*	ACTCAC CCATCC		Deletion	Intronic					
10	1487383	*ADARB2*	CCCACCA ACCCATC CATCCA		Deletion	Intronic					371456791
5	324141	*AHRR*		TT	Insertion	Intronic					
5	324152	*AHRR*	G	C	SNV	Intronic					375043354
5	426735	*AHRR*		AGAT	Insertion	Intronic					142240127
5	436388	*AHRR*	T		Deletion	3'UTR					5865330
12	133319367	*ANKLE2*	TACCACC ATCATC	CACATC TTCACC ATCATC ATCCTC ATCAGC ATCACA ACCATC ATCACT	Substitution	Intronic					

12	133319374	*ANKLE2*		GTCATC ACCATC ACCATC ACCACC ATCATC CTCACT ACCACC	Insertion	Intronic					
12	133319389	*ANKLE2*	A	C	SNV	Intronic					79839935
12	133319393	*ANKLE2*	T	C	SNV	Intronic					
22	18077829	*ATP6V1E1*	G	TT	Substitution	Intronic					
16	88027231	*BANP*	AG		Deletion	Intronic					397782415
7	1155219	*C7orf50*	G	C	SNV	Intronic					6971277
9	140778126	*CACNA1B*	G	C	SNV	Intronic					71510860
9	140783083	*CACNA1B*	G	A	SNV	Intronic					
9	140843825	*CACNA1B*	G		Deletion	Intronic					375703513
9	140843829	*CACNA1B*	GCA	TTGG	Substitution	Intronic					
9	140843835	*CACNA1B*	G	T	SNV	Intronic					
1	6997348	*CAMTA1*	C	A	SNV	Intronic					200150393
1	6997357	*CAMTA1*	T	C	SNV	Intronic					114689433
1	6997367	*CAMTA1*	C	T	SNV	Intronic					202203323
1	6997379	*CAMTA1*	C	G	SNV	Intronic					200484676
1	7000990	*CAMTA1*	AA	GGGGG GTCCTC AGAGG GAGAG GGATCC TCCCG GGGAA G	Substitution	Intronic					
1	7044157	*CAMTA1*	GGTG		Deletion	Intronic					
1	7044158	*CAMTA1*	GTGA	ATAGGT GG	Substitution	Intronic					
1	7513062	*CAMTA1*	G		Deletion	Intronic					
1	7553229	*CAMTA1*	GC	ACGT	Substitution	Intronic					
1	7553329	*CAMTA1*	A	G	SNV	Intronic					62653661

1	7553338	*CAMTA1*	TACA	CG	Substitution	Intronic					386628175
16	89041618	*CBFA2T3*	C	T	SNV	Intronic					4782488
6	47464977	*CD2AP*	TTTCTTT TTTCCT		Deletion	Intronic					
6	47524769	*CD2AP*	G	A	SNV	Intronic					145551659
15	68507331	*CLN6*		CACCT TCCC	Insertion	Intronic					55798476
7	155308003	*CNPY1*	CG	A	Substitution	Intronic					386719929
20	61958464	*COL20A1*	AGGGACT TCCTGTT TACCACT CAGGGA CTTTCTG TTGACTG CC		Deletion	Intronic					
1	34043170	*CSMD2*	C		Deletion	Intronic					11291100
1	34290740	*CSMD2*	T	C	SNV	Intronic					6657112
1	34318453	*CSMD2*	C	T	SNV	Intronic					386427638
1	34346601	*CSMD2*		CCAGTA TGTG	Insertion	Intronic					112560751
1	34349981	*CSMD2*		CC	Insertion	Intronic					61360744
1	34592405	*CSMD2*		ATG	Insertion	Intronic					3044852
10	491882	*DIP2C*	GCG	ACA	Substitution	Intronic					
10	523535	*DIP2C*		GGGAG GAAGGT GGATG GGT	Insertion	Intronic					
10	716205	*DIP2C*	CACGCAT GCATATA		Deletion	Intronic					
6	487331	*EXOC2*	G		Deletion	Intronic					
6	528886	*EXOC2*		G	Insertion	Intronic					
6	170627586	*FAM120B*	G	T	SNV	Exonic; Intronic	Missense			Possibly Damaging	6917485
15	26870736	*GABRB3*		TTGGCT GTGGG GA	Insertion	Intronic					66894857

7	6546285	*GRID2IP*	T	C	SNV	Intronic					13233173
19	1085509	*HMHA1*	T		Deletion	Intronic					149458285
12	121432669	*HNF1A*	TCCA	AT	Substitution	Intronic					
12	121432675	*HNF1A*	C	A	SNV	Intronic					56405042
12	121432680	*HNF1A*	C	T	SNV	Intronic					56184283
1	13184211	*HNRNPCL1/ HNRNPCL2*	GAC	CAT	Substitution	Intronic					
1	13184230	*HNRNPCL1/ HNRNPCL2*	A	C	SNV	Intronic					66601730
1	13184274	*HNRNPCL1/ HNRNPCL2*	TTGT	GCTC	Substitution	5'UTR					386628615
1	13184283	*HNRNPCL1/ HNRNPCL2*	AGAA	GCAG	Substitution	5'UTR					386628616
1	13184291	*HNRNPCL1/ HNRNPCL2*	C	T	SNV	5'UTR					76654751
1	13184308	*HNRNPCL1/ HNRNPCL2*	G	A	SNV	5'UTR					72474536
X	104360584	*IL1RAPL2*	A	G	SNV	Intronic					138629862
X	104651788	*IL1RAPL2*	CTCA		Deletion	Intronic					
18	44553082	*KATNAL2*	C	A	SNV	Intronic					150979709
1	15374352	*KAZN*	C	G	SNV	Intronic					147484885
7	1856066	*MAD1L1*	G	A	SNV	Intronic					186387153
7	1869473	*MAD1L1*	T	A	SNV	Intronic					34373690
7	1910705	*MAD1L1*	T	C	SNV	Intronic					111519111
7	1927507	*MAD1L1*		TGGCTG CCTGG GCGTGT GTGTG	Insertion	Intronic					
7	1927521	*MAD1L1*	G	A	SNV	Intronic					112388595
7	1953083	*MAD1L1*		GCGAA GGGAG TGGGA	Insertion	Intronic					57726885
7	2205259	*MAD1L1*		GT	Insertion	Intronic					59181779
2	2283340	*MYT1L*	TG	GGAAG A	Substitution	Intronic					

18	77195348	*NFATC1*	GT	TTGGTT TGTTTG	Substitution	Intronic					
18	77223426	*NFATC1*	C	T	SNV	Intronic					60884900
17	47582123	*NGFR*	C	TCTT	Substitution	Intronic					375870215
17	47585127	*NGFR*		TATA	Insertion	Intronic					11466139
1	228551859	*OBSCN*		ACAGTG CCCTCC CCAAG GACTGC ACGGG	Insertion	Intronic					148845306
15	28259023	*OCA2*	GT		Deletion	Intronic					369589569
1	29188423	*OPRD1*		CTT	Insertion	Intronic					372713364
2	240985511	*OR6B3*	A	G	SNV	Promoter					10933560
19	45900730	*PPP1R13L*	C	T	SNV	Intronic					4803815
7	623092	*PRKAR1B*		CT	Insertion	Intronic					200178183
7	651882	*PRKAR1B*	C	T	SNV	Intronic					62431453
7	651899	*PRKAR1B*	C	T	SNV	Intronic					62431454
7	696635	*PRKAR1B*	T	C	SNV	Intronic					113961823
7	697065	*PRKAR1B*	AATGGGT GA	GGTG	Substitution	Intronic					
7	157376644	*PTPRN2*	C	T	SNV	Intronic					4716760
7	157413940	*PTPRN2*	T	G	SNV	Intronic					2906925
7	157450528	*PTPRN2*	AC	GCCA	Substitution	Intronic					
7	157738435	*PTPRN2*	A	G	SNV	Intronic					62475406
7	157821939	*PTPRN2*	G	A	SNV	Intronic					9801601
7	157923758	*PTPRN2*	CAC	T	Substitution	Intronic					386720267
7	157923771	*PTPRN2*	AC		Deletion	Intronic					56136013
7	157923844	*PTPRN2*		GAAA	Insertion	Intronic					56332004
7	157923851	*PTPRN2*		AC	Insertion	Intronic					
7	157924071	*PTPRN2*	CACT		Deletion	Intronic					76071394
7	157948224	*PTPRN2*	G	A	SNV	Intronic					
7	157948234	*PTPRN2*	C	A	SNV	Intronic					112891751

7	157957760	*PTPRN2*		C	Insertion	Intronic					
7	157957768	*PTPRN2*	ATT	GTC	Substitution	Intronic					
7	158055480	*PTPRN2*	CA		Deletion	Intronic					
7	158055576	*PTPRN2*		TCACACACTGCACAC	Insertion	Intronic					371888713
7	158055616	*PTPRN2*	T	C	SNV	Intronic					62649324
7	158055710	*PTPRN2*	AGTGCACACAT	TGCACACATAGATTCACACACATTGCACACACACATTCACACACAGTGCAC	Substitution	Intronic					
7	158061414	*PTPRN2*	AGA	GGC	Substitution	Intronic					386720307
7	158062601	*PTPRN2*	GACAGAGAAACAGAGA		Deletion	Intronic					
7	158119352	*PTPRN2*		AC	Insertion	Intronic					
7	158120173	*PTPRN2*	ACAT		Deletion	Intronic					
7	158180204	*PTPRN2*	TGGA	CAGCTG	Substitution	Intronic					
7	158203643	*PTPRN2*	GGTAC	TGCAG	Substitution	Intronic					386720335
7	158203653	*PTPRN2*	C	T	SNV	Intronic					4909216
7	158233753	*PTPRN2*	GGGTGGTAGTGATGG		Deletion	Intronic					
7	158246600	*PTPRN2*	T	C	SNV	Intronic					1612814
7	158288885	*PTPRN2*	TGTCCCCCTTCCCTGCAG		Deletion	Intronic					377678840
16	668837	*RAB40C*	A	T	SNV	Intronic					
13	114795723	*RASA3*	A	C	SNV	Intronic					2274709
13	114795754	*RASA3*	A	G	SNV	Intronic					2274708
17	66675	*RPH3AL*	G	A	SNV	Intronic					8066107

17	78608052	*RPTOR*	A	G	SNV	Intronic					62068346
17	78608075	*RPTOR*	G	A	SNV	Intronic					62641966
17	78608076	*RPTOR*		TACTGTGTGTGTGCG	Insertion	Intronic					
17	78608631	*RPTOR*	TGTGTGCA	CGTGTGCG	Substitution	Intronic					
17	78608663	*RPTOR*	TGTGTGCATAC		Deletion	Intronic					147584971
17	78674429	*RPTOR*	CGGC	T	Substitution	Intronic					
17	78785716	*RPTOR*	A	G	SNV	Intronic					
17	78785728	*RPTOR*	C	T	SNV	Intronic					374712081
9	137286091	*RXRA*	ACGGTGGTGA	TTAATGTTGATGATGGTAGTGATGGTGGTGATGATGGTGGTGATGATGC	Substitution	Intronic					
3	38621396	*SCN5A*	A	C	SNV	Intronic					56887724
X	587163	*SHOX*	C	T	SNV	Intronic					
X	587229	*SHOX*	C	A	SNV	Intronic					
2	220502072	*SLC4A3*	T		Deletion	Intronic					
5	1400943	*SLC6A3*		CTGTCTACACCAGCC	Insertion	Intronic					377330133
5	1421234	*SLC6A3*	T	A	SNV	Intronic					112306339
16	12123802	*SNX29*		T	Insertion	Intronic					35128101
16	12502034	*SNX29*	C	A	SNV	Intronic					76748851
16	12502038	*SNX29*	A	C	SNV	Intronic					72771556
19	7709271	*STXBP2*	A	G	SNV	Intronic					3815754
13	114477564	*TMEM255B*	C	T	SNV	Intronic					367698554
13	114477570	*TMEM255B*	AATG	CC	Substitution	Intronic					386775122

13	114477577	*TMEM255B*	CAC	TAT	Substitution	Intronic					386775123
13	114478005	*TMEM255B*	TGACACA CATTGCA CACACAA CACACCA CACACCA CACA		Deletion	Intronic					371947737
13	114478044	*TMEM255B*		CA	Insertion	Intronic					
13	114498960	*TMEM255B*	G	C	SNV	Intronic					111173929
13	114498969	*TMEM255B*	G	C	SNV	Intronic					111173928
8	140885212	*TRAPPC9*	TTTGGTG GATAA	ATAAGT GGATG G	Substitution	Intronic					
8	140885238	*TRAPPC9*	A	G	SNV	Intronic					
8	141125731	*TRAPPC9*		T	Insertion	Intronic					397689235
8	141206857	*TRAPPC9*	CTCA		Deletion	Intronic					373821045
8	143299632	*TSNARE1*		AGCAG GATCAG GGTTCA GTGTGT G	Insertion	Intronic					
8	143299651	*TSNARE1*	G	T	SNV	Intronic					78394843
8	143304426	*TSNARE1*	T	C	SNV	Intronic					116999715
8	143304426	*TSNARE1*	T	CGTTCA TTC	Substitution	Intronic					
8	143353920	*TSNARE1*	TCCATCC ACCCACC CGTCTAC ACCTTCC TCC	CT	Substitution	Intronic					
8	143353962	*TSNARE1*	CAC	TGT	Substitution	Intronic					
8	143354002	*TSNARE1*	CA	TG	Substitution	Intronic					
2	3308451	*TSSC1*	G	A	SNV	Intronic					71444245
2	3308616	*TSSC1*		GTTCAG CCCTCC ACTCCC GCCCA	Insertion	Intronic					

2	3323527	*TSSC1*	GCA	ACC	Substitution	Intronic					
2	3323534	*TSSC1*	C		Deletion	Intronic					
2	3323541	*TSSC1*	G	A	SNV	Intronic					
2	3340808	*TSSC1*	AAAAGAGAAGGGAAAAGGGTAAAAAGAAG		Deletion	Intronic					
3	75779680	*ZNF717*	A	G	SNV	3'UTR					78431785
3	75779688	*ZNF717*	T	G	SNV	3'UTR					78506989
3	75780725	*ZNF717*		A	Insertion	Intronic					
3	75780783	*ZNF717*	G	A	SNV	Intronic					78199428
3	75780806	*ZNF717*	G	A	SNV	Intronic					62250077
3	75780813	*ZNF717*	C	T	SNV	Intronic					62250078
3	75780841	*ZNF717*	A	T	SNV	Intronic					111646898
3	75780853	*ZNF717*	C	T	SNV	Intronic					113521980
3	75780874	*ZNF717*	T	G	SNV	Intronic					112358097
3	75781049	*ZNF717*	G	T	SNV	Intronic					77738888
3	75781404	*ZNF717*	G	T	SNV	Intronic					73841590
3	75781536	*ZNF717*	C	G	SNV	Intronic					74391052
3	75781564	*ZNF717*	T	C	SNV	Intronic					368766203
3	75781579	*ZNF717*	G	C	SNV	Intronic					79224266
3	75781583	*ZNF717*	C	A	SNV	Intronic					76090413
3	75781665	*ZNF717*	C	T	SNV	Intronic					79319996
3	75782062	*ZNF717*	T	C	SNV	Intronic					149141037
3	75782087	*ZNF717*		A	Insertion	Intronic					
3	75782359	*ZNF717*	T	A	SNV	Intronic					74744994
3	75782372	*ZNF717*	G	C	SNV	Intronic					146049279
3	75782733	*ZNF717*	CT	TC	Substitution	Intronic					
3	75782765	*ZNF717*	GAA	AT	Substitution	Intronic					
3	75784256	*ZNF717*	T	A	SNV	Intronic					74369226

3	75784284	*ZNF717*	A	G	SNV	Intronic					79188089
3	75784777	*ZNF717*	CTGTCTCAATAAAAAAAAAAAAA	ATGCTGTTTCAATTTC	Substitution	Intronic					
3	75784991	*ZNF717*	T	C	SNV	Intronic					78939026
3	75784998	*ZNF717*	C	T	SNV	Intronic					77687685
3	75785130	*ZNF717*	A	G	SNV	Intronic					78906640
3	75785133	*ZNF717*	G	T	SNV	Intronic					78698209
3	75785149	*ZNF717*	C	G	SNV	Intronic					79135712
3	75785318	*ZNF717*	A	G	SNV	Intronic					80053710
3	75785333	*ZNF717*	G	A	SNV	Intronic					79305051
3	75785338	*ZNF717*	TG	CA	Substitution	Intronic					386662578
3	75785351	*ZNF717*	T	C	SNV	Intronic					77346754
3	75785355	*ZNF717*	T	C	SNV	Intronic					75958893
3	75785752	*ZNF717*	GAG		Deletion	Intronic					145023391
3	75785757	*ZNF717*	G	C	SNV	Intronic					80304037
3	75785773	*ZNF717*	G	T	SNV	Intronic					75922518
3	75786892	*ZNF717*	A	C	SNV	Exonic; Intronic	Missense	Damaging	0	Probably Damaging	76826286
3	75786899	*ZNF717*	T	C	SNV	Exonic; Intronic	Synonymous				78828372
3	75786916	*ZNF717*	C	A	SNV	Exonic; Intronic	Stop gain				78906544
3	75786919	*ZNF717*	A	T	SNV	Exonic; Intronic	Missense	Damaging	0	Probably Damaging	77101176
3	75787222	*ZNF717*	AGC	GGT	Substitution	Exonic; Intronic	In-frame				
3	75787240	*ZNF717*	A	G	SNV	Exonic; Intronic	Missense	Damaging	0	Probably Damaging	76111663
3	75787279	*ZNF717*	A	T	SNV	Exonic; Intronic	Missense	Activating	1	Benign	80214832

3	75787288	*ZNF717*	T	C	SNV	Exonic; Intronic	Missense	Tolerated	0.74	Benign	76815006
3	75787298	*ZNF717*	T	C	SNV	Exonic; Intronic	Synonym -ous				143229208
3	75787304	*ZNF717*	AC	TT	Substitution	Exonic; Intronic	In-frame				
3	75787996	*ZNF717*	C	T	SNV	Exonic; Intronic	Missense	Damaging	0.01	Possibly Damaging	78640256
3	75788777	*ZNF717*	C	A	SNV	Intronic					79935262
3	75789592	*ZNF717*	T	C	SNV	Intronic					114166831
3	75789722	*ZNF717*	TTT	CTA	Substitution	Intronic					
3	75791120	*ZNF717*		AT	Insertion	Intronic					
3	75791136	*ZNF717*	T	C	SNV	Intronic					62246579
3	75791370	*ZNF717*	G	T	SNV	Intronic					137896860
3	75791393	*ZNF717*	TG	CA	Substitution	Intronic					
3	75791630	*ZNF717*	G	A	SNV	Intronic					77369095
3	75791639	*ZNF717*	A	C	SNV	Intronic					73843033
3	75791647	*ZNF717*	C	T	SNV	Intronic					73843034
3	75792812	*ZNF717*		GAGGA CAC	Insertion	Intronic					
3	75792961	*ZNF717*	T	C	SNV	Intronic					77809421
3	75792965	*ZNF717*	G	C	SNV	Intronic					143220375
3	75793015	*ZNF717*	G	A	SNV	Intronic					79963736
3	75793038	*ZNF717*	A	G	SNV	Intronic					76503683
3	75793322	*ZNF717*	G	A	SNV	Intronic					79089350
3	75793328	*ZNF717*	CTA		Deletion	Intronic					149321743
3	75793346	*ZNF717*	A	T	SNV	Intronic					75299410
3	75793355	*ZNF717*	G	T	SNV	Intronic					80271408
3	75793786	*ZNF717*	A		Deletion	Intronic					142452329
3	75793801	*ZNF717*	T	C	SNV	Intronic					76017054
3	75793805	*ZNF717*	C	T	SNV	Intronic					111761703

3	75794043	*ZNF717*	T		Deletion	Intronic					143048594
3	75794056	*ZNF717*	C	T	SNV	Intronic					79549753
3	75794068	*ZNF717*	A	G	SNV	Intronic					76890775
3	75794078	*ZNF717*	G	A	SNV	Intronic					75860137
3	75794081	*ZNF717*	C	T	SNV	Intronic					76031595
3	75794370	*ZNF717*	T	G	SNV	Intronic					114339799
3	75797741	*ZNF717*	TG	CA	Substitution	Intronic					
3	75824236	*ZNF717*	A	C	SNV	Intronic					372238767
3	75824259	*ZNF717*	C	T	SNV	Intronic					367756272
3	75824378	ZNF717	C	G	SNV	Intronic					62268159
Structural Variants and CNVs											
Chr	**Start**	**End**	**Size**	**Type**	**Gene(s)**						
1	825765	5726936	4901171	Tandem-Duplication	*ACAP3;ACTRT2;AGRN;AJAP1;ARHGEF16;ATAD3A;ATAD3B;ATAD3C;AURKAIP1;B3GALT6;C1orf159;C1orf170;C1orf174;C1orf222;C1orf70;C1orf86;C1orf93;CALML6;CCDC27;CCNL2;CDC2L1;CDC2L2;CPSF3L;DFFB;DVL1;FAM132A;FLJ14100;FLJ39609;FLJ42875;GABRD;GLTPD1;GNB1;HES4;HES5;ISG15;KIAA0495;KIAA0562;KIAA1751;KLHL17;LOC100128003;LOC100128838;LOC100129381;LOC100129534;LOC100131742;LOC100132814;LOC100133612;LOC100287685;LOC100287750;LOC100287848;LOC100287898;LOC100288202;LOC100288271;LOC100288313;LOC100288379;LOC100288479;LOC115110;LOC148413;LOC284661;LOC388588;LOC401934;LOC441869;LOC643988;LOC728661;LOC728716;LRRC47;MEGF6;MIB2;MMEL1;MMP23A;MMP23B;MORN1;MRPL20;MXRA8;NADK;NoC2L;PANK4;PEX10;PLCH2;PLEKHN1;PRDM16;PRKCZ;PUSL1;RER1;SAMD11;SCNN1D;SDF4;SKI;SLC35E2;SSU72;TAS1R3;TMEM52;TMEM88B;TNFRSF14;TNFRSF18;TNFRSF4;TP73;TPRG1L;TTLL10;UBE2J2;VWA1;WDR8*						
2	1801743	1802067	324	Deletion	*MYT1L*						
7	606117	606324	207	Tandem-Duplication	*PRKAR1B*						
15	25468396	25470000	1604	Tandem-Duplication	*SNORD115-29*						

Appendix I - Observed DNA methylation changes in 58 genes that showed co-localized *de novo* sequence variations in Family 1. (NCBI Build 36/hg18)

Chr	Gene Symbol	Start	End	p-value (region)	In Mom	In Dad	Imprinting Database 1 (geneimprint.com/)	Imprinting Database 2 (igc.otago.ac.nz/)
2	*ABCG8*	43919524	43919544	0.000606465			Predicted, Maternal	Not found
16	*ACSM5*	20325935	20325955	0.000667111			No	Not found
10	*ADARB2*	1732764	1732784	0.000242586			No	Not found
5	*AHRR*	356365	356385	0.000667111			No	Not found
5	*AHRR*	356365	356385	0.000667111			No	Not found
12	*ANKLE2*	131819474	131819494	0.000242586	Yes	Yes	No	Not found
22	*ATP6V1E1*	16492493	16492513	0.000909697			No	Not found
22	*ATP6V1E1*	16492493	16492513	0.000909697			No	Not found
22	*ATP6V1E1*	16492493	16492513	0.000909697			No	Not found
16	*BANP*	86591053	86591073	0.000121293			No	Not found
16	*BANP*	86591053	86591073	0.000121293			No	Not found
16	*BANP*	86591053	86591073	0.000121293			No	Not found
16	*BANP*	86591053	86591073	0.000121293			No	Not found
16	*BANP*	86591053	86591073	0.000121293			No	Not found
16	*BANP*	86591053	86591073	0.000121293			No	Not found
16	*BANP*	86591053	86591073	0.000121293			No	Not found
7	*C7orf50*	1122890	1122910	0.000121293			No	Not found
7	*C7orf50*	1122890	1122910	0.000121293			No	Not found
7	*C7orf50*	1122890	1122910	0.000121293			No	Not found
9	*CACNA1B*	139936085	139936105	0.000849051		Yes	No	Not found
9	*CACNA1B*	139936085	139936105	0.000849051		Yes	No	Not found
1	*CAMTA1*	7687801	7687821	0.000363879			No	Not found
16	*CBFA2T3*	87489347	87489367	0.000667111			No	Not found
16	*CBFA2T3*	87489347	87489367	0.000667111			No	Not found

6	*CD2AP*	47552558	47552578	0.000667111			No	Not found
15	*CLN6*	66286407	66286427	0.000667111			No	Not found
7	*CNPY1*	154995456	154995476	0.000242586			No	Not found
20	*COL20A1*	61394057	61394077	0.000667111			No	Not found
1	*CSMD2*	34403910	34403930	0.000788404			No	Not found
1	*CSMD2*	34403910	34403930	0.000788404			No	Not found
1	*DFFB*	3763670	3763690	0.000606465			No	Not found
1	*DFFB*	3763670	3763690	0.000606465			No	Not found
1	*DFFB*	3763670	3763690	0.000606465			No	Not found
10	*DIP2C*	620619	620639	0.000545818			No	Not found
10	*DIP2C*	684962	684982	0.000849051			No	Not found
10	*DIP2C*	382685	382705	0.000363879		Yes	No	Not found
1	*DVL1*	1275260	1275280	0.000909697			Predicted, Maternal	Not found
6	*EXOC2*	465198	465218	0.000849051			No	Not found
6	*EXOC2*	465198	465218	0.000849051			No	Not found
6	*FAM120B*	170528419	170528439	0.000121293			No	Not found
6	*FAM120B*	170528419	170528439	0.000121293			No	Not found
6	*FAM120B*	170528419	170528439	0.000121293			No	Not found
6	*FAM120B*	170528419	170528439	0.000121293			No	Not found
15	*GABRB3*	24425395	24425415	0.000788404			Conflicting Data, Paternal	Conflicting Data
15	*GABRB3*	24425395	24425415	0.000788404			Conflicting Data, Paternal	Conflicting Data
15	*GABRB3*	24425395	24425415	0.000788404			Conflicting Data, Paternal	Conflicting Data
15	*GABRB3*	24425395	24425415	0.000788404			Conflicting Data, Paternal	Conflicting Data
15	*GABRB3*	24425395	24425415	0.000788404			Conflicting Data, Paternal	Conflicting Data
15	*GABRB3*	24425395	24425415	0.000788404			Conflicting Data, Paternal	Conflicting Data
7	*GRID2IP*	6510144	6510164	0.000970344			No	Not found
19	*HMHA1*	1017228	1017248	0.000909697			No	Not found
19	*HMHA1*	1017228	1017248	0.000909697			No	Not found
19	*HMHA1*	1017228	1017248	0.000909697			No	Not found
12	*HNF1A*	119900006	119900026	0.000303232			No	Not found

1	*HNRNPCP5*	13107459	13107479	0.000849051			No	Not found
X	*IL1RAPL2*	104352919	104352939	0.000849051			No	Not found
18	*KATNAL2*	42804957	42804977	0.000242586			No	Not found
18	*KATNAL2*	42810872	42810892	0.000545818			No	Not found
18	*KATNAL2*	42798912	42798932	0.000303232			No	Not found
1	*KAZN*	15122441	15122461	0.000121293			No	Not found
1	*KAZN*	15122441	15122461	0.000121293			No	Not found
1	*KAZN*	15122441	15122461	0.000121293			No	Not found
1	*KIAA1751*	1901884	1901904	0.000849051			No	Not found
7	*MAD1L1*	1877342	1877362	0.000545818			No	Not found
7	*MAD1L1*	2056253	2056273	0.000121293			No	Not found
7	*MAD1L1*	1877342	1877362	0.000545818			No	Not found
7	*MAD1L1*	2056253	2056273	0.000121293			No	Not found
7	*MAD1L1*	1877342	1877362	0.000545818			No	Not found
7	*MAD1L1*	2056253	2056273	0.000121293			No	Not found
2	*MYT1L*	1920311	1920331	0.000242586			No	Not found
18	*NFATC1*	75281489	75281509	0.000667111			No	Not found
18	*NFATC1*	75281489	75281509	0.000667111			No	Not found
18	*NFATC1*	75281489	75281509	0.000667111			No	Not found
18	*NFATC1*	75281489	75281509	0.000667111			No	Not found
18	*NFATC1*	75281489	75281509	0.000667111			No	Not found
18	*NFATC1*	75281489	75281509	0.000667111			No	Not found
18	*NFATC1*	75281489	75281509	0.000667111			No	Not found
18	*NFATC1*	75281489	75281509	0.000667111			No	Not found
18	*NFATC1*	75281489	75281509	0.000667111			No	Not found
18	*NFATC1*	75281489	75281509	0.000667111			No	Not found
17	*NGFR*	44927754	44927774	0.000667111			No	Not found
1	*OBSCN*	226570772	226570792	0.000909697			Predicted, Paternal	Not found
1	*OBSCN*	226467367	226467387	0.000788404			Predicted, Paternal	Not found
1	*OBSCN*	226570772	226570792	0.000909697			Predicted, Paternal	Not found

1	*OBSCN*	226467367	226467387	0.000788404			Predicted, Paternal	Not found
1	*OBSCN*	226570772	226570792	0.000909697			Predicted, Paternal	Not found
1	*OBSCN*	226467367	226467387	0.000788404			Predicted, Paternal	Not found
15	*OCA2*	25689155	25689175	0.000363879			No	Not found
1	*OPRD1*	29062206	29062226	0.000667111			No	Not found
2	*OR6B3*	240631724	240631744	0.000424525			No	Not found
19	*PPP1R13L*	50591427	50591447	0.000909697			No	Not found
19	*PPP1R13L*	50591427	50591447	0.000909697			No	Not found
7	*PRKAR1B*	717528	717548	0.000667111			No	Not found
7	*PRKAR1B*	717528	717548	0.000667111			No	Not found
7	*PRKAR1B*	717528	717548	0.000667111			No	Not found
7	*PRKAR1B*	717528	717548	0.000667111			No	Not found
7	*PRKAR1B*	717528	717548	0.000667111			No	Not found
7	*PRKAR1B*	717528	717548	0.000667111			No	Not found
7	*PTPRN2*	157352628	157352648	0.000909697			No	Not found
7	*PTPRN2*	157941351	157941371	0.000181939			No	Not found
7	*PTPRN2*	157352628	157352648	0.000909697			No	Not found
7	*PTPRN2*	157941351	157941371	0.000181939			No	Not found
7	*PTPRN2*	157352628	157352648	0.000909697			No	Not found
7	*PTPRN2*	157941351	157941371	0.000181939			No	Not found
16	*RAB40C*	604302	604322	0.000424525			No	Not found
16	*RAB40C*	604302	604322	0.000424525			No	Not found
16	*RAB40C*	604302	604322	0.000424525			No	Not found
16	*RAB40C*	604302	604322	0.000424525			No	Not found
16	*RAB40C*	604302	604322	0.000424525			No	Not found
13	*RASA3*	113824201	113824221	0.000485172			No	Not found
13	*RASA3*	113856967	113856987	0.000242586		Yes	No	Not found
17	*RPH3AL*	206061	206081	0.000545818			No	Not found
17	*RPH3AL*	206061	206081	0.000545818			No	Not found
17	*RPH3AL*	206061	206081	0.000545818			No	Not found

17	*RPH3AL*	206061	206081	0.000545818			No	Not found
17	*RPTOR*	76132310	76132330	0.000909697			No	Not found
17	*RPTOR*	76132310	76132330	0.000909697			No	Not found
9	*RXRA*	136440826	136440846	0.000667111			No	Not found
3	*SCN5A*	38667000	38667020	0.000849051			No	Not found
3	*SCN5A*	38667000	38667020	0.000849051			No	Not found
3	*SCN5A*	38667000	38667020	0.000849051			No	Not found
3	*SCN5A*	38667000	38667020	0.000849051			No	Not found
3	*SCN5A*	38667000	38667020	0.000849051			No	Not found
3	*SCN5A*	38667000	38667020	0.000849051			No	Not found
X	*SHOX*	523098	523118	0.000909697			No	Not found
X	*SHOX*	523098	523118	0.000909697			No	Not found
2	*SLC4A3*	220199610	220199630	0.000363879			No	Not found
2	*SLC4A3*	220199610	220199630	0.000363879			No	Not found
2	*SLC4A3*	220199610	220199630	0.000363879			No	Not found
5	*SLC6A3*	1469888	1469908	0.000849051			No	Not found
15	*SNoRD115-29*	23030052	23030072	0.000545818	Yes	Yes	No	Imprinted
16	*SNX29*	12118466	12118486	0.000363879			No	Not found
19	*STXBP2*	7618186	7618206	0.000121293			No	Not found
19	*STXBP2*	7618186	7618206	0.000121293			No	Not found
19	*STXBP2*	7618186	7618206	0.000121293			No	Not found
19	*STXBP2*	7618186	7618206	0.000121293			No	Not found
13	*TMEM255B*	113606174	113606194	0.000606465			No	Not found
8	*TRAPPC9*	141428560	141428580	0.000909697			No	Not found
8	*TRAPPC9*	141428560	141428580	0.000909697			No	Not found
8	*TSNARE1*	143405544	143405564	0.000485172	Yes	Yes	No	Not found
2	*TSSC1*	3270302	3270322	0.000121293			No	Not found
3	*ZNF717*	75917850	75917870	0.000363879			No	Not found
3	*ZNF717*	75917850	75917870	0.000363879			No	Not found

3	*ZNF717*	75917850	75917870	0.000363879			No	Not found
3	*ZNF717*	75917850	75917870	0.000363879			No	Not found

Appendix J - *De novo* sequencing variations in the affected twin of Family 2 that show co-localized differential methylation and sequence variation. (NCBI Build 37/hg19)

Chr	Position	Gene	Ref Allele (Build 37)	Sample Allele	Variation Type	Gene Region	Impact	SIFT Function Prediction	SIFT Score	PolyPhen-2 Function Prediction	dbSNP
1	33475879	*AK2*	C	T	SNV	ncRNA; 3'UTR					61800917
1	33475967	*AK2*	G	A	SNV	ncRNA; 3'UTR					74066437
1	33475977	*AK2*	GGA	AGG	Substitution	ncRNA; 3'UTR					386630167
1	33475982	*AK2*	C	A	SNV	ncRNA; 3'UTR					114902220
1	33476292	*AK2*	A	G	SNV	ncRNA; 3'UTR					74066439
X	39986111	*BCOR*		A	Insertion	Intronic					
9	138666266	*KCNT1*	CTGT		Deletion	Intronic					
3	195510827	*MUC4*	C	T	SNV	Intronic; Exonic	Missense			Benign	413807
3	195511780	*MUC4*	G	A	SNV	Intronic; Exonic	Missense			Benign	391928
8	145039599	*PLEC*		CTGGTC TGCCAT CATGCC AGGCC AC	Insertion	Intronic					371713005
1	3098038	*PRDM16*	GC		Deletion	Intronic					
1	3098052	*PRDM16*	A	G	SNV	Intronic					77999053
1	3247836	*PRDM16*	A	C	SNV	Intronic					11577229
1	3326640	*PRDM16*	A	G	SNV	Intronic					4648494
7	157413778	*PTPRN2*	CG	TA	Substitution	Intronic					
7	157413789	*PTPRN2*	TGT	CAC	Substitution	Intronic					

7	157413803	*PTPRN2*	C	T	SNV	Intronic					
7	157564480	*PTPRN2*		TCAC ACAC	Insertion	Intronic					55838542
7	157564526	*PTPRN2*	GACACA		Deletion	Intronic					
7	157679515	*PTPRN2*	T	C	SNV	Intronic					77012117
7	157679518	*PTPRN2*	A	G	SNV	Intronic					76659620
7	157841832	*PTPRN2*		CCA	Insertion	Intronic					
7	157923238	*PTPRN2*	A	T	SNV	Intronic					4909267
7	157957688	*PTPRN2*	GA	TG	Substitution	Intronic					67901051
7	158055099	*PTPRN2*	CA		Deletion	Intronic					
7	158061226	*PTPRN2*	A	G	SNV	Intronic					2004588
7	158082384	*PTPRN2*	G		Deletion	Intronic					
7	158105002	*PTPRN2*		TA	Insertion	Intronic					10654725
7	158105117	*PTPRN2*	CACA		Deletion	Intronic					
7	158112250	*PTPRN2*	G	A	SNV	Intronic					56188120
7	158212055	*PTPRN2*	G	A	SNV	Intronic					62478442
7	158233753	*PTPRN2*	GGGTG GTAGTG ATGG		Deletion	Intronic					
16	6113522	*RBFOX1*	C		Deletion	Intronic					71142671
16	6423347	*RBFOX1*	GGC GGA	AGT GGC	Substitution	Intronic					386788248
16	6423384	*RBFOX1*	AG	GT	Substitution	Intronic					
16	7251467	*RBFOX1*	T		Deletion	Intronic					
16	7563288	*RBFOX1*	TG	A	Substitution	Intronic					386788490
16	7563293	*RBFOX1*	G		Deletion	Intronic					201299763
16	7592893	*RBFOX1*	A	G	SNV	Intronic					8053938
16	7748187	*RBFOX1*	T	G	SNV	Intronic					12932981
10	16831467	*RSU1*	T	A	SNV	Intronic					10904813
3	9457871	*SETD5*	TCT		Deletion	Intronic					
4	186509746	*SORBS2*		TCCTCC TCCTCC T	Insertion	Intronic					

4	186717589	*SORBS2*	A	G	SNV	Intronic					13126725
5	166854854	*TENM2*		G	Insertion	Intronic					
5	167398378	*TENM2*		AAG	Insertion	Intronic					
5	167398387	*TENM2*	G	A	SNV	Intronic					71603845
5	167398403	*TENM2*	T	G	SNV	Intronic					13173408
11	89819403	*UBTFL1*	G	T	SNV	Exonic	Missense	Damaging	0	Benign	
Structural Variants and CNVs											
Chr	**Start**	**End**	**Size**	**Type**	**Gene(s)**						
7	158029478	158030452	974	Deletion	*PTPRN2*						

Appendix K - Observed DNA methylation changes in 13 genes that showed co-localized *de novo* sequence variations in Family 2. (NCBI Build 36/hg18)

Chr	Gene Symbol	Start	End	p-value (region)	In Mom	In Dad	Imprinting Database 1 (geneimprint.com/)	Imprinting Database 2 (igc.otago.ac.nz/)
1	*AK2*	33275973	33275993	0.000485172			No	Not found
1	*AK2*	33275973	33275993	0.000485172			No	Not found
1	*AK2*	33275973	33275993	0.000485172			No	Not found
1	*AK2*	33275973	33275993	0.000485172			No	Not found
1	*AK2*	33275973	33275993	0.000485172			No	Not found
X	*BCOR*	39846274	39846294	0.000545818	Yes	Yes	No	Not found
X	*BCOR*	39846274	39846294	0.000545818	Yes	Yes	No	Not found
X	*BCOR*	39846274	39846294	0.000545818	Yes	Yes	No	Not found
X	*BCOR*	39846274	39846294	0.000545818	Yes	Yes	No	Not found
9	*KCNT1*	137732516	137732536	0.000909697			No	Not found
9	*KCNT1*	137732516	137732536	0.000909697			No	Not found
3	*MUC4*	196975464	196975484	0.000909697			No	Not found
3	*MUC4*	196975464	196975484	0.000909697			No	Not found
3	*MUC4*	196975464	196975484	0.000909697			No	Not found
8	*PLEC*	145068180	145068200	0.000727758	Yes	Yes	No	Not found
8	*PLEC*	145068180	145068200	0.000727758	Yes	Yes	No	Not found
8	*PLEC*	145068180	145068200	0.000727758	Yes	Yes	No	Not found
8	*PLEC*	145068180	145068200	0.000727758	Yes	Yes	No	Not found
8	*PLEC*	145068180	145068200	0.000727758	Yes	Yes	No	Not found
8	*PLEC*	145068180	145068200	0.000727758	Yes	Yes	No	Not found
8	*PLEC*	145068180	145068200	0.000727758	Yes	Yes	No	Not found
8	*PLEC*	145068180	145068200	0.000727758	Yes	Yes	No	Not found
1	*PRDM16*	3061919	3061939	0.000788404			Predicted, Paternal	Not found
1	*PRDM16*	3061919	3061939	0.000788404			Predicted, Paternal	Not found

7	*PTPRN2*	157141154	157141174	0.000667111			No	Not found
7	*PTPRN2*	157141154	157141174	0.000667111			No	Not found
7	*PTPRN2*	157141154	157141174	0.000667111			No	Not found
16	*RBFOX1*	6762885	6762905	0.000909697			No	Not found
16	*RBFOX1*	6762885	6762905	0.000909697			No	Not found
16	*RBFOX1*	6762885	6762905	0.000909697			No	Not found
16	*RBFOX1*	6762885	6762905	0.000909697			No	Not found
16	*RBFOX1*	6762885	6762905	0.000909697			No	Not found
16	*RBFOX1*	6762885	6762905	0.000909697			No	Not found
10	*RSU1*	16900364	16900384	0.000909697			No	Not found
10	*RSU1*	16900364	16900384	0.000909697			No	Not found
3	*SETD5*	9414278	9414298	0.000788404			No	Not found
4	*SORBS2*	187115769	187115789	0.000667111			No	Not found
4	*SORBS2*	187115769	187115789	0.000667111			No	Not found
5	*TENM2*	166643495	166643515	0.000909697			No	Not found
11	*UBTFL1*	89457840	89457860	0.000970344			No	Not found

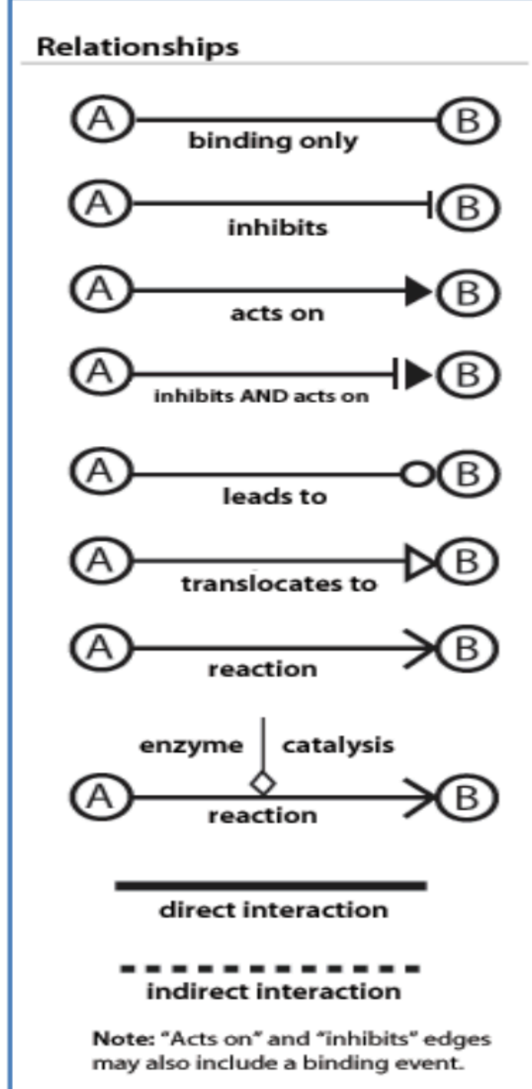
Appendix L - Ingenuity Pathway Analysis Legend

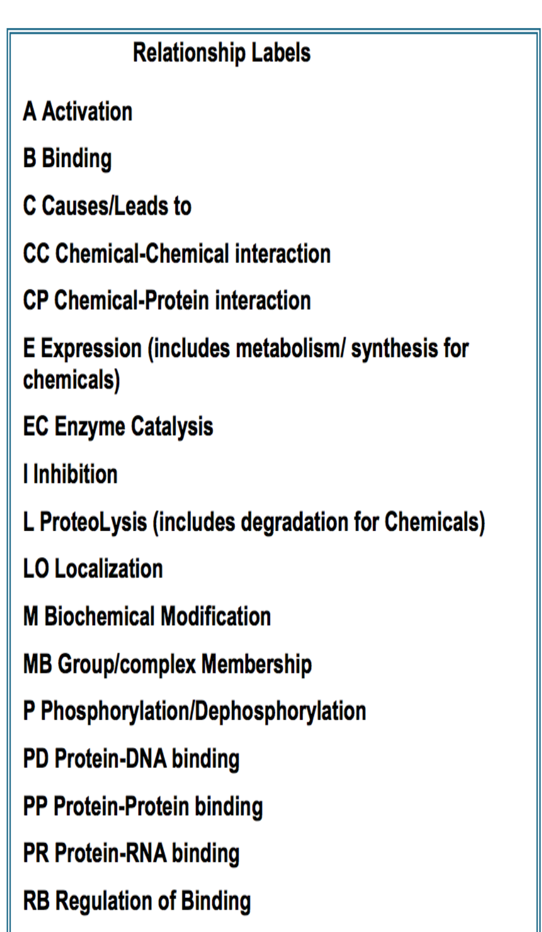
Network Shapes

- Cytokine
- Growth Factor
- Chemical / Drug/ Toxicant
- Enzyme
- G-protein Coupled Receptor
- Ion Channel
- Kinase
- Ligand-dependent Nuclear Receptor
- Peptidase
- Phosphatase
- Transcription Regulator
- Translation Regulator
- Transmembrane Receptor
- Transporter
- microRNA
- Complex / Group
- Other

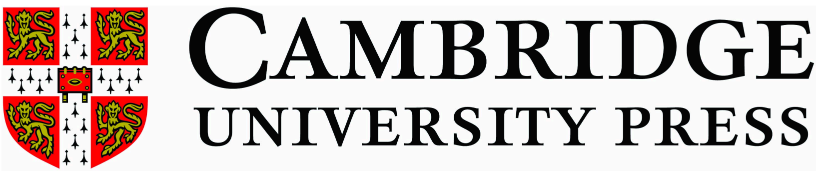
Path Designer Shapes

- Cytokine / Growth Factor
- Drug
- Chemical / Toxicant
- Enzyme
- G-protein Coupled Receptor
- Ion Channel
- Kinase
- Ligand-dependent Nuclear Receptor
- Peptidase
- Phosphatase
- Transcription Regulator
- Translation Regulator
- Transmembrane Receptor
- Transporter
- microRNA
- Complex / Group
- Other


Relationships
A
B
binding only
inhibits
acts on
inhibits AND acts on
leads to
translocates to
reaction
enzyme
catalysis
reaction
direct interaction
indirect interaction
Note: "Acts on" and "inhibits" edges may also include a binding event.


Relationship Labels
A Activation
B Binding
C Causes/Leads to
CC Chemical-Chemical interaction
CP Chemical-Protein interaction
E Expression (includes metabolism/ synthesis for chemicals)
EC Enzyme Catalysis
I Inhibition
L ProteoLysis (includes degradation for Chemicals)
LO Localization
M Biochemical Modification
MB Group/complex Membership
P Phosphorylation/Dephosphorylation
PD Protein-DNA binding
PP Protein-Protein binding
PR Protein-RNA binding
RB Regulation of Binding
RE Reaction
RR RNA-RNA Binding
T Transcription
TR Translocation

Appendix M - Journal Copyright Approval



Dear Christina Castellani,

Thank you for your email enquiry, you may include your article in your forthcoming PhD thesis, we just ask that you give full copyright acknowledgement to the original source. There is further information for our authors on our website at the following link.

http://www.cambridge.org/about-us/rights-permissions/permissions/permissions-requests-our-authors/

Article: Christina A. Castellani, Zain Awamleh, Melkaye G. Melka, Richard L. O'Reilly and Shiva M. Singh (2014). Copy Number Variation Distribution in Six Monozygotic Twin Pairs Discordant for Schizophrenia. Twin Research and Human Genetics, 17, pp 108-120. doi:10.1017/thg.2014.6.

Yours sincerely,

Claire Taylor
Senior Publishing Assistant
Cambridge University Press
University Printing House

Note: All other publications included in this thesis are open access publications that are reproducible with attribution to the original source.

Christina A. CASTELLANI
Curriculum Vitae

RESEARCH INTERESTS

Whole Genome Sequencing, Copy Number Variation, Epigenetics, Neurodevelopment, Structural Variants, Complex Disease

CURRENT POSITION

Ph.D. Candidate, Molecular Genetics
Supervisor: Dr. Shiva M. Singh, The University of Western Ontario

Comprehensive Examination – 2012, *Pass With Distinction*
Comprehensive Topics: Genomic Rearrangement in Human Brain, Genome Sequencing Techniques, and Algorithms for Large-scale Data Analysis

DEGREES EARNED

M.Sc., The University of Western Ontario, 2009

Thesis Entitled: Gene Discovery in Schizophrenia through Identification of Copy Number Variations (CNVs) in Discordant Monozygotic (MZ) Twins

B.Sc. (Honours), The University of Western Ontario, 2007

Honors Specialization in Genetics

GRADUATE COURSES

The Biology of Ageing; Statistics for Neuroscience;
Perspectives on Genetic Variation; Epigenetics;
Perspectives on Mutation; Cell Biology of Embryonic Stem Cells.

TEACHING EXPERIENCE

2015: Lead TA, Introductory Genetics, Biology 2581b
Preparing tutorial materials, running Teaching Assistant meetings, managing and organizing TAs, writing quiz/exam questions, and facilitating demo tutorials.

2010-2014: Genome Organization, Mutagenesis and Repair, Biology 3594a

2008- 2011: Introductory Genetics, Biology 2581b

2007: Human Genetics, Biology 3592a

ACADEMIC HONORS AND AWARDS

2015 CIHR (Canadian Institutes of Health Research) Postdoctoral Research Fellowship ($150,000)

2014 Academic Achievement Scholarship (GTA Union) ($500)

2014 Society of Biological Psychiatry (SOBP) Trainee Award ($2,000)

2013 CIHR Travel Award for attendance at Sequencing Workshop ($2,000)

2013 Graduate Thesis Research Award, Faculty of Science, Western University ($500)

2012 CIHR (Canadian Institutes of Health Research) Canada Graduate Scholarship (CGS) Doctoral Research Award (DRA) ($105,000)

2012 Ontario Mental Health Foundation (OMHF) Studentship ($48,000) *(Declined)*

2012 Ontario Graduate Scholarship (OGS) ($15,000) *(Declined)*

2012 3rd Biennial Schizophrenia International Research Society Conference Travel Award ($1,500)

2011 Society of Graduate Students (SOGS) Teaching Award, Western University ($500)

2011 Great Ideas In Innovative Teaching Award, Teaching Support Center, Western University ($50)

2011 American Society of Human Genetics (ASHG) International Congress Trainee Award – ASHG/Charles J. Epstein Trainee Award for Excellence in Human Genetics Research ($400)

2011 World Congress on Psychiatric Genetics (WCPG), Trainee Travel Award ($1,500)

2011 Ontario Graduate Scholarship (OGS) ($15,000)

2011 David E. Laudenbach Scholarship for excellence in balance between research and leadership activities, Department of Biology, Western University ($400)

2011 Canadian Society of Zoologists (CSZ) Invited Travel Award

2010 Vanier Canadian Graduate Scholarship Institutional Nominee

2010 Department of Psychiatry Academic Research Day Best Oral Presentation ($500)

2008 Ontario Graduate Scholarship in Science and Technology ($15,000)

2007 Western Graduate Scholarship, Faculty of Science, Western University ($1,500)

2007 Excellence in Leadership Award, Western University

2003 UWO National Scholarship Entrance Program, Faculty of Science, Western University ($24,000)

2003 WHSC Scholarship for Youth Advocacy ($1,000)

2002 Western's Initiative for Scholarly Excellence, Western University

PROFESSIONAL DEVELOPMENT

- Participant: **Informatics for High-throughput Sequencing Data Workshop** offered through the Canadian Bioinformatics Workshop Program. Toronto, Ontario. June 2013.
- Participant: **Partek Genomics Suite Training, Ingenuity Pathway Analysis Training, GeneMANIA Training**. London, Ontario. 2011.
- Participant: **Complete Genomics User Group Meeting and Workshop** offered through Complete Genomics. San Francisco, California. October 2010.

PUBLICATIONS

Submitted Manuscripts

Castellani, CA., Melka, MG., Gui, JL., O'Reilly, RL., Singh, SM. Integration of DNA Sequence and DNA Methylation Changes in Two Monozygotic Twin Pairs Discordant for Schizophrenia. ***Schizophrenia Bulletin:*** *Under Review.*

Castellani, CA., Melka, MG., Gui, JL., Gallo, A., O'Reilly, RL., Singh, SM. Origin and Timing of *de novo* Mutations Identified by Whole-Genome Sequencing of Monozygotic Twins Discordant for Schizophrenia. ***Nature Scientific Reports:*** *Under Review.*

Diehl, EJ., Laufer, BI., **Castellani, CA.,** Alberry, BL., Singh, SM. Long-term Changes in Hippocampal Histone Methylation in a Mouse Model of Fetal Alcohol Spectrum Disorder (FASD). ***Molecular Brain:*** *Under Review.*

Published Manuscripts

Castellani, CA., Laufer, BI., Melka, MG., Diehl, EJ., O'Reilly, RL., Singh, SM. Shared Differences in DNA Methylation Between Unrelated Monozygotic Twins May Explain Their Discordance for Schizophrenia. ***BMC Medical Genomics:*** 2015 May 6; 8(1):17. *Highly Accessed

Melka, MG., **Castellani, CA.,** O'Reilly, RL., Singh, SM. Insights into the Origin of DNA Methylation Differences Between Monozygotic Twins Discordant for Schizophrenia. ***Journal of Molecular Psychiatry:*** 2015.

Laufer, BI., Kapalanga, J., **Castellani, CA.,** Diehl, EJ., Singh, SM. Associative DNA Methylation Changes in Children with Prenatal Alcohol Exposure. ***Epigenomics:*** 2015 Jul 16: 1-16.

Castellani, CA., Melka, MG., Diehl, EJ., Laufer, BI., O'Reilly, RL., Singh, SM. DNA Methylation in Psychosis: Insights into etiology and Treatment. ***Epigenomics:*** 2015 Feb 7(1):67.

Melka, MG., **Castellani, CA.,** Rajakumar, N., O'Reilly, RL., Singh, SM. Olanzapine-Induced Methylation Alters Cadherin Gene Families and Associated Pathways Implicated in Psychosis. ***BMC Neuroscience****:* 2014 Sep 29:15:112.

Castellani CA., Melka MG., Wishart A., Locke MEO., Awamleh Z., O'Reilly RL., Singh SM. Biological Relevance of CNV Calling Methods Using Familial Relatedness Including Monozygotic Twins. ***BMC Bioinformatics:*** 2014 Apr 21:15(1):114. *Highly accessed

Castellani CA., Awamleh Z., Melka MG., O'Reilly RL., Singh SM. Copy Number Variation Distribution in Six Monozygotic Twin Pairs Discordant for Schizophrenia. ***Twin Research and Human Genetics:*** 2014 Feb 20:1-13.

Melka MG., Laufer BI., McDonald P., **Castellani CA.**, Rajakumar N., O'Reilly RL., Singh SM. The Effects of Olanzapine on Genome-Wide DNA Methylation in the Hippocampus and Cerebellum. ***Clinical Epigenetics:*** 2014 Jan 2;6(1):1. *Highly accessed

Melka MG., **Castellani CA.**, Laufer BL., Rajakumar R., O'Reilly RL., Singh SM. Olanzapine Induced DNA Methylation Changes Support the Dopamine Hypothesis of Psychosis. ***Journal of Molecular Psychiatry:*** 2013 *1:19.*

Kumar HB K., **<u>Castellani CA.</u>**, Maiti S., O'Reilly RL., and Singh SM. Search for Missing Schizophrenia Genes Will Require a New Developmental Neurogenomic Perspective. ***Journal of Human Genetics:*** 2013 Aug; 92(2):335-40.

Kumar K., Maiti S., **<u>Castellani CA.</u>**, O'Reilly R., and Singh SM. A Novel Deletion Cluster at 13q14.2-q21.33 in an 80 Year Old Man with Late Onset Leukemia: Clinical and Molecular Findings. ***Ind. J Hum Genetics:*** *2013 19(1):96-100.*

Abbs B…. **<u>Castellani CA</u>**, et al. The 3rd **Schizophrenia** International Research Society Conference, 14-18 April 2012, Florence, Italy: Summaries of Oral Sessions. ***Schizophrenia Research:*** *2012 141(1):e1-e24.*

Maiti S., Kumar K., **<u>Castellani CA.</u>**, O'Reilly R., and Singh SM (2011). Ontogenetic *de novo* Copy Number Variations (CNVs) as a Source of Genetic Individuality: Studies on Two Families with MZD Twins for Schizophrenia. ***PLoS One****:* 2011 6(3):e17125.

Singh, SM., **<u>Castellani, CA.,</u>** O'Reilly, RL. Autism Meets Schizophrenia via Cadherin Pathway. ***Schizophrenia Research***: 2010 116 (2-3): 293-294.

Singh, SM, **<u>Castellani, CA.,</u>** O'Reilly, RL. Copy Number Variation Showers in Schizophrenia: An Emerging Hypothesis. ***Molecular Psychiatry***: 2009 14: 356-358.

<u>DEPARTMENTAL AND UNIVERSITY SERVICE</u>

2013-Present Graduate Student Member, Faculty task force on academic integrity, ***Faculty of Science***

2012-Present Graduate Member, University Committee on Research Ethics, ***University Senate***

2007-2014 Residence Manager, Coordination of Residence Life Program, ***Huron University College***

2010-2013 Member, Society of Biology Graduate Student, UWO, ***Department of Biology***

2012-2013 Research Committee Representative, ***Department of Biology***

2010-2012 Biology Graduate Education Committee (BGEC) Representative, ***Department of Biology***

2003-2013 University Biology (Introductory Genetics, Genetic Engineering, Biochemistry), ***Private Tutor***

2004-2007 Intensive Care Unit (ICU) Volunteer, ***University Hospital***

2004-2007 Residence Don, ***Department of Housing and Ancillary Services***

PRESENTATIONS

Invited Oral Presentations

- **C. Castellani,** R. O'Reilly, S. Singh. Complete Genome Sequence Based Family Analysis of Monozygotic Twins Discordant For Schizophrenia. *Society of Biological Psychiatry Annual Meeting.* New York, New York. May 9 2014.

- **C. Castellani**, R. O'Reilly, S. Singh. Complete Genome Sequence Based Genetic Analysis of Monozygotic Twins Discordant for Schizophrenia. *XXth World Congress of Psychiatric Genetics (WCPG).* Hamburg, Germany. October 16 2012.

- **C. Castellani** and S. Singh. Complete Genome Sequence Based Genetic Analysis of Monozygotic Twins Discordant for Schizophrenia*. 3rd Biennial Schizophrenia International Research Society Conference.* Florence, Italy. April 17 2012.

- **C. Castellani** and S. Singh. Are Monozygotic Twins Identical in Genetic and Epigenetic Features? *Canadian Society of Zoologists Annual Meeting.* Invited Speaker – Cell and Molecular Development Satellite Symposium, University of Ottawa, Ottawa, ON. May 18 2011. *Session: New perspectives on adaptation to environmental variation via epigenetic mechanisms.*

- **C. Castellani,** S. Maiti, K. Kumar, R. O'Reilly and S. Singh. Incidence and Inheritance of Copy Number Variations in Monozygotic Twins Discordant for Schizophrenia. *Genetics Society of Canada Annual Conference.* McMaster University, Hamilton, ON. June 19 2010.

- **C. Castellani,** S. Maiti, K. Kumar, R. O'Reilly and S. Singh. Incidence and Inheritance of Copy Number Variations in Monozygotic Twins Discordant for Schizophrenia. *Psychiatry Academic Research Day.* June 2008, June 2010, June 2011.

- R. O'Reilly, S. Singh and **C. Castellani**. Copy Number Variations in Monozygotic Twins and the Search for Genes Involved in Complex Disorders: Lessons From Schizophrenia. *Canadian Psychiatric Association 59th Annual Conference*, St. John's Newfoundland, August 27-29, 2009.

Guest Lecture

- **C. Castellani.** Biology 3595a. Advanced Genetics. Topic: *Discordant Monozygotic Twins. September 2014.*

- **C. Castellani**. Biology 4592. Investigative Techniques in Genetics. Topic: *Whole Genome Sequencing. January 2011.*

- **C. Castellani.** Biology 3594a, Genome Organization, Mutagenesis and Repair. Topic: *Copy Number Variants, DNA Microarrays, Monozygotic Twins as a genetic tool, Next-Generation Human Genome Sequencing. October 2010, 2011.*

Invited Poster Presentations

- E. Diehl, B. Laufer, **C. Castellani** and S. Singh. Oxidative Stress Pathways Implicated in Comprehensive Epigenetic and Transcriptomic Assessment of Hippocampus from Fetal Ethanol-Exposed Mice. *Neuroepigenetics.* Santa Fe, NM. February 23 2015.

- **C. Castellani,** J. Gui, M. Melka, R. O'Reilly, S. Singh. Differences in Genetic Features May Explain the Discordance of Monozygotic Twins for Schizophrenia. *American Society of Human Genetics (ASHG) Annual Meeting.* San Diego, CA. October 20 2014.

- B. Laufer, J. Kapalanga., E. Diehl, **C. Castellani,** M. Kleiber, A. Chokroborty-Hoque, B. Alberry, K. Mantha, L. Yan, S. Singh. A NeuroEpigenomic Translation of the Fetal Alcohol Exposure Spectrum. *American Society of Human Genetics (ASHG) Annual Meeting.* San Diego, CA. October 21 2014.

- **C. Castellani,** B. Laufer, M. Melka, R. O'Reilly, S. Singh. Differences in DNA Methylation Between Monozygotic Twins May Explain Their Discordance for Schizophrenia. *2nd Canadian Conference on Epigenetics: Epigenetics Eh!* London, ON. June 25 2014.

- M. Melka, **C. Castellani**, B. Laufer, R. O'Reilly, N. Rajakumar, S. Singh. Efficacy of Antipsychotic Olanzapine is realized via Changes in DNA Methylation in Vivo. *2nd Canadian Conference on Epigenetics: Epigenetics Eh!* London, ON. June 25 2014.

- **C. Castellani**, R. O'Reilly, S. Singh. Complete Genome Sequence Based Family Analysis of Monozygotic Twins Discordant for Schizophrenia. *World Congress of Psychiatric Genetics (WCPG) Annual Meeting.* Boston, MA. October 21 2013.

- **C. Castellani**, R. O'Reilly, S. Singh. Complete Genome Sequence Based Genetic Analysis of Monozygotic Twins Discordant for Schizophrenia. *American Society of Human Genetics (ASHG) Annual Meeting.* San Francisco, CA. November 9 2012.

- **C. Castellani**, S. Maiti, R. O'Reilly and S. Singh. Complete genome sequence based genetic analysis of monozygotic twins discordant (MZD) for schizophrenia. *12th International Congress of Human Genetics.* Montreal, QU. October 13 2011.

- S. Maiti, K. Kumar, **C. Castellani,** R. O'Reilly and S. Singh. Genetic exploration of genome wide variation in monozygotic twins discordant for schizophrenia. *12th International Congress of Human Genetics.* Montreal, QU. October 13 2011.

- R. O'Reilly, **C. Castellani,** S. Maiti and S. Singh. *De novo* CNVs affecting multiple genes/pathways may explain discordance of monozygotic twins for schizophrenia. *12th International Congress of Human Genetics.* Montreal, QU. October 13 2011.

- **C. Castellani,** R. O'Reilly, S. Maiti and S. Singh. *De novo* CNVs affecting multiple genes/pathways may explain discordance of monozygotic twins for schizophrenia. *World Congress of Psychiatric Genetics (WCPG).* Washington, DC. September 11 2011.

- **C. Castellani,** R. O'Reilly, S. Scherer, and S. Singh. Incidence and Inheritance of Copy Number Variations in Monozygotic Twins Discordant for Schizophrenia. *World Discoveries Research Day.* London, ON. 2009.

RESEARCH MENTORSHIP

- Emma Reble. September 2014 – April 2015. Thesis Student. Thesis entitled: **Identifying Somatic Mutations in Monozygotic Twins Discordant for Schizophrenia using VarScan.**

- Jane Liqian Gui. September 2013 – August 2014. Thesis Student and Summer Research Project. Thesis entitled: **Investigation of *de novo* Copy Number Variants (CNVs) and Structural Variants (SVs) Between and Within Generations using Whole Genome Sequences of Parents and Monozygotic Twins.**

- Zain Awamleh. September 2012 – July 2013. Thesis Student and Summer Research Project. Thesis entitled: **Assessing Genomic Variants in Monozygotic Twins Discordant for Schizophrenia using Multiple Software Programs.**

CURRENT PROFESSIONAL AFFILIATIONS

- Canadian Society for Molecular Biosciences (CSMB), Member
- Society of Biological Psychiatry (SOBP), Member
- American Society of Human Genetics (ASHG), Member