Investigation of Gene Splicing Variants Contributing Risk to Psychiatric Disorders

by

Emma Reble

A thesis submitted in conformity with the requirements for the degree of Master of Science

Institute of Medical Science
University of Toronto

© Copyright by Emma Reble 2017
Investigation of Gene Splicing Variants Contributing Risk to Psychiatric Disorders

Emma Reble

Master of Science

Institute of Medical Science
University of Toronto

2017

Abstract

The majority of the variants found through genome-wide association studies (GWAS) to be associated with psychiatric disorders are not found to cause a protein coding change, therefore it is predicted that these variants are causing gene expression changes, including changes to alternative splicing. Datasets of variants associated with alternative splicing changes in the general population have also been curated with the aim of aiding the search for functional GWAS variants. This study uses these datasets to identify variants associated with alternatively spliced isoforms in schizophrenia and bipolar disorder. Alternatively spliced isoforms in four genes (APOPT1, AS3MT, NEK4 and RPGRIP1L) were identified that converge on the mitochondria and cilia, implicating these cellular structures in psychiatric disease and providing avenues of genes and pathways for further study. A method to study alternative splicing changes in psychiatric disorders is presented that would also be a useful tool for studying other complex genetic disorders.
Table of Contents

Table of Contents ........................................................................................................ iii
Acknowledgments ........................................................................................................ vi
Statement of Contributions ....................................................................................... vii
List of Abbreviations ................................................................................................. viii
List of Tables .............................................................................................................. x
List of Figures ............................................................................................................. xi

Chapter 1: Introduction and Background ................................................................ 1
1.1 Psychiatric Disorders and Their Genetic Contributions ..................................... 1
  1.1.1 Schizophrenia .................................................................................................. 1
  1.1.2 Bipolar Disorder ............................................................................................. 3
  1.1.3 Elucidating Genetic Risk for Psychiatric Disorders ....................................... 4
1.2 Gene Splicing ........................................................................................................ 5
  1.2.1 Splicing Mechanisms ..................................................................................... 5
  1.2.2 Types of Alternative Splicing ....................................................................... 7
  1.2.3 Alternative Splicing in Mendelian Disease ................................................... 9
  1.2.4 Alternative Splicing in the Brain ................................................................ 10
  1.2.5 Alternative Splicing in Psychiatric Disorders ........................................... 11
    1.2.5.1 Alternative Splicing in Autism Spectrum Disorder (ASD) ..................... 12
    1.2.5.2 Alternative Splicing in Bipolar Disorder ............................................ 14
    1.2.5.3 Alternative Splicing in Schizophrenia ............................................. 15
1.3 Splicing Predictions and Datasets ..................................................................... 17
  1.3.1 SPIDEX ......................................................................................................... 18
  1.3.2 Altrans .......................................................................................................... 19
  1.3.3 sQTLSekeRe ................................................................................................. 19
  1.3.4 Protein Truncating Variants (PTV) ............................................................. 20
  1.3.5 DLPFC sQTLs .............................................................................................. 21
1.4 Protein Structure and Function Prediction ....................................................... 21
1.5 Hypothesis, Aims and Overview of Thesis ....................................................... 23

Chapter 2: Materials and Methods ....................................................................... 25
2.1 Identification of Schizophrenia and Bipolar Disorder Candidate SNVs .......... 25
2.2 Brain tissue samples ........................................................................................... 25
2.3 SNV Genotyping ................................................................................................. 27
2.4 Expression of Alternative Splicing Isoforms .................................................... 28
  2.4.1 Identification of Splicing Variants ............................................................... 28
  2.4.2 Quantification of Identified Splicing Variants Using ddPCR ....................... 31
2.5 Predicting the Mechanism of the Splice Variants ............................................ 33
Chapter 3: Results.................................................................................................................. 34

3.1 Splicing SNVs associated with Schizophrenia and Bipolar Disorder ........... 35

3.2 Genes studied for alternative splicing................................................................. 42

3.2.1 APOPT1 ........................................................................................................ 42

3.2.1.1 APOPT1 SNVs Associated with Alternative Splicing ......................... 43

3.2.1.2 Identification of Two APOPT1 Transcript Isoforms ............................. 45

3.2.1.3 Quantification of APOPT1 Isoforms and Association to rs4906337, rs10431750
and rs7140568 ......................................................................................... 46

3.2.1.4 Protein Function Prediction of APOPT1 Isoforms .................................. 48

3.2.2 AS3MT ........................................................................................................... 50

3.2.2.1 AS3MT SNV Predicted to Alter Splicing ............................................. 51

3.2.2.2 Identification of Two AS3MT Transcript Isoforms ............................... 53

3.2.2.3 Quantification of AS3MT Isoforms and Association to rs3740400 ........ 54

3.2.2.4 Protein Function Prediction of AS3MT Isoforms ................................. 56

3.2.2.5 Hypothesis of inefficient arsenic metabolism contributing to risk of schizophrenia

3.2.3 BNIP3L ........................................................................................................... 59

3.2.3.1 SNVs Associated with Alternative BNIP3L Transcripts ....................... 59

3.2.3.2 Identification of Two BNIP3L Transcript Isoforms .............................. 60

3.2.3.3 Quantification of BNIP3L Isoforms and Association to rs11779570, rs11786753,
rs60377146 and rs1135938 ........................................................................ 61

3.2.3.4 Protein Function Prediction of BNIP3L Isoforms ................................. 63

3.2.4 CACNA1D .................................................................................................... 65

3.2.5 KIF21B ....................................................................................................... 66

3.2.6 LRRC48 ....................................................................................................... 68

3.2.6.1 LRRC48 SNVs Predicted to Alter Exon 7 and 8 Splicing .................... 69

3.2.6.2 LRRC48 SNV Predicted to Alter Exon 2 Splicing .............................. 71

3.2.6.3 Identification of Three LRRC48 Transcript Isoforms ....................... 72

3.2.6.4 Quantification of LRRC48 Isoforms and Association to rs11870660 .... 73

3.2.6.5 Protein Function Prediction of LRRC48 Isoforms .............................. 76

3.2.7 NEK4 ....................................................................................................... 76

3.2.7.1 NEK4 SNVs Predicted to Alter Exon 5 Splicing ................................. 77

3.2.7.2 NEK4 SNVs Predicted to Alter Exon 4 Splicing ................................. 78

3.2.7.3 Identification of Two NEK4 Transcript Isoforms ............................... 81

3.2.7.4 Quantification of NEK4 Isoforms and Association to SNVs associated with NEK4
exon 4 skipping .......................................................... 83

3.2.7.5 Protein Function Prediction of NEK4 Isoforms ................................. 84

3.2.8 NGEF ....................................................................................................... 85

3.2.9 PRKAG1 .................................................................................................... 87

3.2.10 PPP1R16B ............................................................................................... 90

3.2.11 RPRGIP1L .................................................................................................. 93

3.2.11.1 RPRGIP1L SNVs Associated with Alternative Splicing ................. 93

3.2.11.2 Identification of Two RPRGIP1L Transcript Isoforms ....................... 95

3.2.11.3 Quantification of RPRGIP1L Isoforms and Association to SNVs associated with
RPRGIP1L exon 20 skipping ................................................ 97

3.2.11.4 Protein Function Prediction of RPRGIP1L Isoforms .......................... 99

3.2.12 SNAP91 ................................................................................................. 100

3.2.13 STAB1 .................................................................................................... 103
Chapter 4: Discussion and Conclusions

4.1 Implications

4.1.1 Alternative transcripts converge on the mitochondria and the cilia

4.1.2 Combined Alternative Splicing Model of Psychiatric Risk

4.1.3 Negative Results

4.1.4 Useful method to identify splicing changes in other complex disorders

4.2 Future Directions

4.2.1 Causal SNV confirmation

4.2.2 Experimentally tested protein function

4.2.3 Combined alternative splicing and environmental effects

4.2.4 Splicing Therapies

4.3 Conclusions

Bibliography
Acknowledgments

First and foremost I would like to thank my supervisor, Dr. Cathy Barr, for giving me the opportunity to work in her lab as well as her support and guidance throughout this project. She is incredibly dedicated and passionate about science and genetics in particular and has stimulated and encouraged my own passion for research. I would also like to thank my committee members, Dr. Ana Andreazza, Dr. Jennifer Mitchell and Dr. Jeff Wrana for their feedback and insight on this project.

I would like to thank the Harvard Brain Tissue Resource, the Douglas Brain Bell Canada Brain Bank and the National Institute of Child Health and Human Development for providing the brain tissue samples. I would also like to acknowledge the support of this research by the Ontario Mental Health Foundation and my scholarship supported by the Hospital for Sick Children Research Training Centre.

Finally, I would especially like to thank all of the Barr lab members for their support and friendship throughout the past two years. Specifically, I would like to thank Karen Wigg for her help with the bioinformatics portion of this project, Yu Feng for all her help and troubleshooting in the lab, Dr. Aidan Dineen for his advice and mentorship and Kaitlyn Price for her encouragement.
Statement of Contributions

Karen Wigg was involved in the bioinformatics analyses including downloading data, making BED files and training of bioinformatics techniques. Karen Wigg and Kaitlyn Price were both involved in the preparation and analysis of the genotyping results. Yu Feng performed the brain sample DNA and RNA extractions and was involved in the training of lab techniques.

The brain tissue samples were generously received from the Harvard Brain Tissue Resource, the Douglas Brain Bell Canada Brain Bank and the National Institute of Child Health and Human Development.

This research was supported by the Ontario Mental Health Foundation and I am supported by a scholarship from the Hospital for Sick Children Research Training Centre.
List of Abbreviations

AChE: Acetylcholinesterase
ANOVA: Analysis Of Variance
AON: Antisense Oligonucleotide
APOPT1: Apoptogenic 1, mitochondrial
AS3MT: Arsenic (+3 Oxidation State) Methyltransferase
ASD: Autism Spectrum Disorder
ATM: Ataxia Telangiectasia Mutated
BD: Bipolar Disorder
BNIP3L: BCL2 Interacting Protein 3 Like
CACNA1D: Calcium Voltage-Gated Channel Subunit Alpha1 D
cDNA: Complimentary Deoxyribonucleic Acid
CNV: Copy Number Variation
ddPCR: Digital Droplet Polymerase Chain Reaction
DLPFC: Dorsolateral Prefrontal Cortex
DNA: Deoxyribonucleic Acid
DRD2: Dopamine Receptor D2
eQTL: expression Quantitative Trait Loci
ESE: Exonic Splicing Enhancer
ESS: Exonic Splicing Silencer
EtBr: Ethidium Bromide
FTDP-17: Frontotemporal Dementia with Parkinsonim-17
GABA: Gamma-Aminobutyric Acid
GAD: Glutamic Acid Decarboxylase
GO: Gene Ontology
GTEX: Genotype Tissue Expression
GWAS: Genome Wide Association Study
hnRNP: Heterogeneous Nuclear Ribonucleoprotein
ISE: Intrinsic Splicing Enhancer
ISS: Intrinsic Splicing Silencer
KIF21B: Kinesin Family Member 21B
LD: Linkage Disequilibrium
LRRC48: Leucine Rich Repeat Containing 48
MAF: Minor Allele Frequency
MAPT: Microtubule-Associated Protein Tau
mRNA: Messenger Ribonucleic Acid
N-DRC: Nexin-Dynein Regulatory Complex
NEK4: Never In Mitosis Gene A Related Kinase 4
NF1: Neurofibromin 1
NGEF: Neuronal Guanine Nucleotide Exchange Factor
NOVA: Neuro-Oncological Ventral Antigen
NMD: Nonsense Mediated Decay
NRG: Neuregulin
PCR: Polymerase Chain Reaction
PPP1R16B: Protein Phosphatase 1 Regulatory Subunit 16 B
PRKAG1: Protein Kinase AMP-activated Non-Catalytic Subunit Gamma 1
PSI: Percent Spliced In
PTBP: Polypyrimidine Tract Binding Protein
PTV: Protein Truncating Variant
RIN: RNA Integrity Number
RNA: Ribonucleic Acid
ROS: Reactive Oxygen Species
RPGRIP1L: Retinitis Pigmentosa GTPase Regulator Interacting Protein 1-Like
SCZ: Schizophrenia
SNAP91: Synaptosome Associated Protein 91
snRNP: Small Nuclear Ribonucleic Protein
SNV: Single Nucleotide Variant
sQTL: splicing Quantitative Trail Loci
SR: Serine/Arginine Rich
SRRM4/nSR100: Serine/Arginine Repetitive Matrix Protein 4
STAB1: Stabilin 1
VNTR: Variable Number Tandem Repeat
List of Tables

Table 1. Demographics of individuals from whom brain sample tissues were collected. 26
Table 2. Schizophrenia and alternative splicing associated SNVs. 36
Table 3. Bipolar disorder and alternative splicing associated SNVs. 38
Table 4. SNVs associated with alternative splicing of APOPT1. 44
Table 5. SNV predicted to alter splicing of AS3MT. 52
Table 6. SNVs associated with alternative BNIP3L transcripts. 60
Table 7. SNV predicted to alter splicing of CACNA1D. 65
Table 8. SNV associated with alternative splicing of KIF21B. 67
Table 9. SNVs predicted to alter splicing of LRRC48. 69
Table 10. SNV predicted to alter splicing of LRRC48. 71
Table 11. SNVs associated with alternative splicing of NEK4. 77
Table 12. SNVs associated with alternative splicing of NEK4. 79
Table 13. Brain tissue sample genotypes of SNVs associated with alternative splicing of NEK4. 81
Table 14. SNV predicted to alter splicing of NGEF. 85
Table 15. SNVs associated with alternative splicing of PRKAG1. 88
Table 16. Brain tissue sample genotypes of SNVs associated with alternative splicing of PRKAG1. 88
Table 17. SNVs associated with alternative splicing of PPP1R16B. 90
Table 18. SNVs associated with alternative splicing of RPGRIP1L. 94
Table 19. Brain tissue sample genotypes of SNVs associated with alternative splicing of RPGRIP1L. 96
Table 20. SNVs associated with alternative splicing of SNAP91. 100
Table 21. Brain tissue sample genotypes of SNVs associated with alternative splicing of SNAP91. 101
Table 22. SNV predicted to alter splicing of STAB1. 103
Table 23. Summary of major findings: alternatively spliced transcripts associated with schizophrenia and bipolar disorder risk alleles. 105
List of Figures

Figure 1. Schematic of spliceosome snRNPs binding to pre-mRNA. .................................................. 6
Figure 2. Events leading to alternative transcripts. .................................................................................. 8
Figure 3. Schematic of the cDNA assay used to identify alternative splicing transcripts with exon skipping ................................................................. 29
Figure 4. Schematic of the cDNA assay used to identify alternative splicing transcripts with intron retention ........................................................................ 30
Figure 5. Schematic of the cDNA assay used to identify alternative splicing transcripts with alternative first exons ..................................................................... 31
Figure 6. Schematic of the probe-based assay used to identify the proportion of each splicing isoform. ........................................................................................................ 32
Figure 7. Digital Droplet PCR output of probe-based exon skipping assay. ........................................... 32
Figure 8. Identification of two APOPT1 transcript isoforms through cDNA PCR amplification. .............................................................. 46
Figure 9. Quantification of APOPT1 transcripts in the DLPFC. ............................................................. 47
Figure 10. Quantification of APOPT1 transcripts in the hippocampus. ................................................. 48
Figure 11. PredictProtein structural and binding site results for full-length and truncated APOPT1 protein sequences. ................................................................................. 49
Figure 12. Arsenic metabolism pathway. ................................................................................................ 51
Figure 13. rs3740400 SpliceAid results. ................................................................................................ 52
Figure 14. Identification of two AS3MT transcript isoforms through cDNA PCR amplification. ........................................................................................................ 54
Figure 15. Quantification of AS3MT transcripts in the DLPFC.............................................................. 55
Figure 16. Quantification of AS3MT transcripts in the hippocampus ................................................... 56
Figure 17. AS3MT Pfam protein predictions. ........................................................................................... 57
Figure 18. Identification of two BNIP3L transcript isoforms through cDNA PCR amplification. ........................................................................................................ 61
Figure 19. Quantification of BNIP3L transcripts in the DLPFC. ............................................................ 62
Figure 20. Quantification of BNIP3L transcripts in the hippocampus .................................................... 63
Figure 21. PredictProtein binding site results for full-length and truncated BNIP3L protein sequences. ........................................................................................................ 64
Figure 22. No identification of exon 25 skipped CACNA1D transcript. ............................................... 66
Figure 23. No identification of exon 28 skipped KIF21B transcript. .................................................... 68
Figure 24. No identification of exon 7 or 8 skipped LRRRC48 transcripts ........................................... 70
Figure 25. Identification of three LRRRC48 transcript isoforms through cDNA PCR amplification. ........................................................................................................ 72
Figure 26. Quantification of LRRRC48 transcripts in the DLPFC.......................................................... 74
Figure 27. Quantification of LRRRC48 transcripts in the hippocampus .............................................. 75
Figure 28. No identification of exon 5 skipped NEK4 transcript. ......................................................... 78
Figure 29. Identification of two NEK4 transcript isoforms through DLPFC cDNA PCR amplification. ........................................................................................................ 82
Figure 30. Quantification of NEK4 transcripts in the DLPFC. .............................................................. 83
Figure 31. Quantification of NEK4 transcripts in the hippocampus ....................................................... 84
Figure 32. NEK4 Pfam protein predictions. ............................................................................................. 85
Figure 33. No identification of exon 2 skipped NGEF transcript. ....................................................... 87
Figure 34. No identification of *PRKAG1* transcript with partial intron 2 retention. ........ 89
Figure 35. No identification of exon 7 skipped *PPP1R16B* transcript. ..................... 92
Figure 36. Identification of two *RPGRIP1L* transcript isoforms through cDNA PCR amplification. ........................................................................................................... 97
Figure 37. Quantification of *RPGRIP1L* transcripts in the DLPFC. ............................. 98
Figure 38. Quantification of *RPGRIP1L* transcripts in the hippocampus. ................ 99
Figure 39. No identification of *SNAP91* transcript with intron 11 retention. .......... 102
Chapter 1

Introduction and Background

1.1 Psychiatric Disorders and Their Genetic Contributions

1.1.1 Schizophrenia
Schizophrenia is a chronic and devastating disorder affecting ~1.1% of the population, with little variance in prevalence worldwide (Regier et al., 1993). The onset of schizophrenia is generally just prior to or during early adulthood. The hallmarks of this disorder are abnormal social behavior and a distortion in the perception of reality (Regier et al., 1993). Symptoms are divided into four main categories: positive, negative, cognitive and mood disturbances. Positive symptoms include delusions and hallucinations; negative symptoms include a flat affect and social and emotional withdrawal; cognitive symptoms include problems with attention, memory, decision-making and executive functioning; and mood disturbances include dysphoria and depression.

The underlying mechanisms of schizophrenia are not well understood, however a strong genetic component is observed with family and twin studies identifying heritability estimates of ~81% (Sullivan, Kendler, & Neale, 2003). Despite known genetic involvement, the specific genes contributing to schizophrenia and their potential impact are still not well known. This lack of etiologic understanding, among other factors, has prevented the discovery of effective and personalized medicine. Current treatments largely target the positive symptoms of schizophrenia and are often accompanied by severe side effects with little impact on the other categories of detrimental symptoms (Wright, 2014). Due to the lack of effective treatment strategies for most patients, two-thirds of individuals with schizophrenia live with detrimental symptoms and 5-13% eventually die by suicide (Pompili et al., 2007; Saha, Chant, Welham, & McGrath, 2005).
Taken together, these factors have caused schizophrenia to be a large social, financial and health burden to patients, families and society (Knapp, Mangalore, & Simon, 2004).

Unlike monogenic disorders where a mutation in a single gene will cause a specific phenotype, the genetic basis of schizophrenia is very complex. Variants in hundreds of genes have been identified in schizophrenia patients, suggesting that multiple genes and variants are likely to be contributing to the schizophrenia phenotype (International Schizophrenia et al., 2009; S. H. Lee et al., 2012; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Debate exists in the field with regards to the relative involvement of rare or common variants in schizophrenia etiology. Two main theories have been suggested about the genetic architecture of schizophrenia. One proposes an extreme heterogeneity or multiple rare variant model in which different rare variants cause the same schizophrenic phenotype in different patients (McClellan & King, 2010). The other theory is a polygenic/multifactorial threshold model in which a predisposition and accumulation of both common and rare risk variants exceeds the threshold of disease for schizophrenia (Gottesman & Shields, 1967; International Schizophrenia et al., 2009). Due to the identification of many common variants associated with schizophrenia, the latter theory is becoming the most widely accepted.

With the goal to identify variants involved in psychiatric disorders, genome-wide association studies (GWAS) have been performed using large samples of psychiatric patients and controls. GWAS examine genetic variants in control and affected groups to search for associations between variants and a disorder that are more commonly seen in patients. These studies have supported the polygenic theories of psychiatric disorders (Sullivan, Daly, & O'Donovan, 2012). GWAS are responsible for identifying new genetic loci associated with psychiatric disorders, such as a breakthrough paper that recently identified 108 regions associated with schizophrenia (Ripke et al., 2013; Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium, 2011; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014).
Although some rare variants of large effect have been identified for schizophrenia, there is more evidence to support the multiple common variants theory with each variant contributing to the risk of schizophrenia (International Schizophrenia et al., 2009). An estimated 23% of variation in schizophrenia risk is thought to be caused by common single nucleotide variants (SNVs), some of which are causal risk variants (S. H. Lee et al., 2012). Therefore, variants associated with schizophrenia are not necessarily found at an extremely low frequency in the general population. Even in the case where a rare, likely causal variant has been identified in schizophrenia patients, it is now thought that common risk variants are still necessary for the disorder to occur. A recent study suggests that common risk variants are contributing to schizophrenia in patients with rare copy number variations (CNVs) (Tansey et al., 2016). Polygenic risk profile scores (RPS) were used to determine the common risk allele burden in schizophrenia cases with a rare schizophrenia-associated CNV, schizophrenia cases with no rare schizophrenia-associated CNV and healthy controls. It was found that both schizophrenia patients with and without a rare schizophrenia-associated CNV had a higher RPS than controls and also that the RPS was not significantly different between schizophrenia patients with and without a schizophrenia-associated CNV (Tansey et al., 2016). This has strong implications for the role of common variants in schizophrenia.

1.1.2 Bipolar Disorder
Bipolar spectrum disorder is a chronic disorder affecting 2.4% of people worldwide (Merikangas et al., 2011). Bipolar disorder (BD) is characterized by disruptive changes in mood and behavior and is categorized into two subtypes: bipolar disorder I (BD-I) and bipolar disorder II (BD-II). Both subtypes involve periods of mania and depression, with BD-I involving manic episodes while BD-II involves periods of hypomania. Hypomanic episodes involve elevated moods, lack of sleep and increased motor drive, while mania encompasses these symptoms but is often more severe with psychotic symptoms present in 75% of patients with manic episodes (Goodwin, Jamison, & Ghaemi, 2007). The risk of suicide is 20 times higher in individuals with bipolar disorder, especially when the disorder is untreated, and 15-20% of people with bipolar disorder die by suicide (Gibbons, Hur, Brown, & Mann, 2009; Pompili et al., 2013; Schaffer et al., 2015). Both
BD-I and BD-II affect social, cognitive and occupational functioning, can lead to hospitalization and cause a decreased quality of life (Martinez-Aran et al., 2007).

The underlying causes of bipolar disorder are not well known, therefore treatments only target acute symptoms as well as long-term management. Antipsychotics, mood stabilizers and anti-depressants are used for acute treatment of mania and depression, however their effects are variable (Grande & Vieta, 2015; Vieta et al., 2010). Lithium is also used as a treatment for bipolar disorder and has generally been found to be the most effective and widely used treatment for both acute symptoms and long-term management (Miura et al., 2014). Lithium is thought to act through glycogen synthase kinase 3, cAMP response element-binding protein and Na(+)-K(+) ATPase signaling, however its exact mechanism of action is still not known (Alda, 2015). Neither the causes of bipolar disorder nor the specific mechanism of action of its treatments are well established therefore more research is required.

A known genetic component has been identified for bipolar disorder with family and twin studies suggesting 77% heritability (Edvardsen et al., 2008). Similar to schizophrenia, bipolar disorder is recognized as a complex genetic disorder with multiple genes and variants that interact with each other and the environment to contribute to the bipolar disorder phenotype (Craddock & Sklar, 2013). Genome-wide association studies have been performed with genotype data from bipolar disorder patients and controls to identify variants that are more prevalent among patients with bipolar disorder and therefore may contribute risk to the disorder (D. T. Chen et al., 2013; Muhleisen et al., 2014). The largest and most recent GWAS of bipolar disorder included genotype data of over 9 million variants from 9,784 people with bipolar disorder and 30,471 controls (Hou et al., 2016). Four genetic loci were replicated that had previously been associated with bipolar disorder and two new genetic loci were identified.

1.1.3 Elucidating Genetic Risk for Psychiatric Disorders
The large number of GWAS and candidate gene studies for psychiatric disorders has identified a high degree of genetic overlap between psychiatric disorders, with five
genetic loci found to be common between schizophrenia, bipolar disorder, major depressive disorder, attention deficit-hyperactivity disorder and autism spectrum disorder (Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2013). A high degree of genetic similarity has especially been found between schizophrenia and bipolar disorder. A recent study aimed at identifying potential genetic associations between bipolar disorder and the risk loci identified in the recent schizophrenia GWAS found that 22 of the schizophrenia-associated index SNVs were also nominally associated with bipolar disorder (Forstner et al., 2017; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Psychiatric disorders clearly have a strong and overlapping genetic contribution, however the exact mechanism by which the associated genes and variants are contributing risk to these disorder is yet to be determined.

The next step in elucidating the contribution of genetic risk to psychiatric disorders is to identify the genes within these regions and further, the mechanism by which these DNA variants are contributing risk. The initial focus of GWAS variants would be to identify non-synonymous, protein-coding changing variants because the effects of these DNA changes on protein structure are more easily predicted. However, many of the GWAS variants identified in psychiatric disorders, and variants identified by GWAS in general, cannot be explained by a coding region SNV. Approximately 93% of GWAS variants are not found in the coding region of the genome (Maurano et al., 2012). For example, within the recently identified 108 loci associated with schizophrenia, only 11 regions contained a potentially causal non-synonymous, protein-coding changing variant (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Therefore more insight is needed into the mechanisms by which these variants identified by GWAS to be associated with psychiatric disorders are related to the risk of the disorder.

1.2 Gene Splicing

1.2.1 Splicing Mechanisms
Splicing of precursor mRNA to remove introns and join exons together is necessary to create a readable transcript to be further translated into a functional RNA or protein. Splicing occurs through the spliceosome complex and requires multiple RNAs and
proteins in order to function properly (Rappsilber, Ryder, Lamond, & Mann, 2002; Zhou, Licklider, Gygi, & Reed, 2002). The spliceosome and the critical consensus splice sites are the basis of the splicing reaction, and are further influenced by intronic and exonic splicing enhancers and inhibitors. The spliceosome consists of the small nuclear ribonucleic proteins (snRNPs) U1 and U2 and the U4, U5 and U6 tri-snRNP (Y. Wang et al., 2015). These snRNPs bind to the critical consensus splice sites: the 5’ splice site or splicing donor, whose consensus sequence is GU, and the 3’ splice site or splicing acceptor, whose consensus sequence is AG. The U1 snRNP binds to the 5’ splice site which initiates splicing and triggers interactions between the U2 snRNP with the branchpoint sequence and the U4, U5 and U6 tri-snRNP with the 5’ and 3’ splice sites (Havens, Duelli, & Hastings, 2013) (Figure 1).

Figure 1. Schematic of spliceosome snRNPs binding to pre-mRNA.

Proper binding of the snRNPs to the 5’ splice site, branchpoint sequence and 3’ splice site is necessary for proper splicing to occur (Havens et al., 2013). The resulting sequence of exons within the mature mRNA transcript after splicing will be determined by the interactions between the cis- and trans-acting elements, which either promote or inhibit interactions between the spliceosome and weak splice sites. Two families of RNA-binding proteins have been identified that act as trans-acting elements and bind to cis-elements of pre-mRNA transcripts to determine the splicing pattern (He & Smith, 2009; Long & Caceres, 2009). Positive cis-acting elements such as exonic splicing enhancers (ESEs) and intronic splicing enhancers (ISEs) interact with positive trans-acting elements, such as serine/arginine-rich (SR) proteins to promote interactions between the
spliceosome and weak splice sites. Negative cis-acting elements such as exonic splicing silencers (ESSs) and intronic splicing silencers (ISSs) interact with negative trans-acting elements such as heterogeneous nuclear ribonucleoproteins (hnRNPs) to inhibit interactions between the spliceosome and weak splice sites and generally promote exon exclusion or skipping events. SR proteins generally antagonize the function of hnRNPs and result in exon inclusion, however some studies have shown that SR proteins regulate exon skipping (Shen & Mattox, 2012). Changes in binding site sequences can affect which splice sites are used and can change the resulting mRNA sequence.

1.2.2 Types of Alternative Splicing
Multiple forms of alternative splicing have been identified that create structurally and functionally different proteins from the same pre-mRNA transcript. More than 90% of all human multi-exon genes are predicted to be affected by alternative splicing of mRNA (Pan, Shai, Lee, Frey, & Blencowe, 2008; E. T. Wang et al., 2008). The most common form of alternative splicing is exon skipping, which accounts for around one third of alternative splicing events in humans (Blencowe, 2006). Other forms of alternative splicing include mutually exclusive exons, intron retention and alternative 5’ or 3’ splice sites. Alternative transcripts may also arise through the use of alternative promoters/first exons and alternative poly A sites/terminal exons (Figure 2).
Alternative splicing may even result in very small changes to the protein sequence, such as the case with NAGNAG acceptors that cause only one codon change depending on which splice site acceptor is used and are found in approximately 30% of human genes (Hiller et al., 2004). These small protein changes have the potential to cause very large consequences and be involved in disease. For example, an SNV in the SFTPA2 gene,
rs1650232, creates an NAGNAG splice acceptor and the addition of a codon into the coding sequence of the surfactant protein A2, and has found to be associated with risk for respiratory diseases (Azad et al., 2013).

This study will focus specifically on cis-acting splice variants, which refer to variants within a gene that affect the splicing of its pre-mRNA transcript. Both exonic and intronic variants may affect splicing in this way. 25% of exonic variants are predicted to alter splicing, including silent exonic variants that have been previously overlooked but are now predicted to cause splicing changes as well as missense exonic variants previously thought to be involved in disease by causing an amino acid change but may in fact contribute to disease by altering splicing (Lim, Ferraris, Filloux, Raphael, & Fairbrother, 2011; Sterne-Weiler, Howard, Mort, Cooper, & Sanford, 2011). Intronic splice site variants are also estimated to account for ~15% of disease variants (Krawczak, Reiss, & Cooper, 1992). Both intronic and exonic variants taken together have been estimated to account for over 30% of all disease causing variants (Lim et al., 2011). In cases where DNA and RNA sequences have been used to identify the mechanism of disease-causing variants, 50% of variants have been found to act by affecting splicing (Ars et al., 2000; Teraoka et al., 1999). Furthermore, a model predicting the effect of disease-causing mutations to either affect splicing or alter the coding sequence found that ~60% of mutations are predicted to be splicing mutations (Lopez-Bigas, Audit, Ouzounis, Parra, & Guigo, 2005). The exact prevalence of disease-causing variants that affect splicing is unknown, however the multiple lines of evidence discussed above suggest that it is likely much higher than previously thought. Along with disease-causing variants that cause splicing changes, common variants that alter the splicing pattern and efficiency may also affect predisposition to disease and disease severity.

### 1.2.3 Alternative Splicing in Mendelian Disease
Disruption of splicing regulation has been identified in many genetic disorders, and has been best characterized in Mendelian diseases. Around 50% of the mutations in the *ATM* and *NF1* genes that cause ataxia-telangiectasia and neurofibromatosis type 1, respectively, are mutations that cause aberrant splicing (Ars et al., 2000; Teraoka et al.,
For these disorders, the sequencing of DNA and mRNA from patients with the disorders has been essential in determining disease-causing splicing mutations (Cartegni, Chew, & Krainer, 2002). In frontotemporal dementia with Parkinsonism linked to chromosome 17 (FTDP-17), cis-acting mutations that affect the splicing of MAPT and the 3:4 ratio of a repeat domain which occurs based on the exclusion or inclusion of exon 10 have been identified as causal (Buee, Bussiere, Buee-Scherrer, Delacourte, & Hof, 2000; V. M. Lee, Goedert, & Trojanowski, 2001). As well, mutations that cause inappropriate splicing through exon skipping or disrupting of splicing enhancers in the dystrophin gene have been found to cause Duchenne muscular dystrophy (Ahn & Kunkel, 1993).

### 1.2.4 Alternative Splicing in the Brain

Alternative splicing occurs more frequently in the brain than in any other tissue (Blencowe, 2006; Yeo, Holste, Kreiman, & Burge, 2004). For example, Neurexin genes, important for synaptic function, have been found to have almost 3000 different isoforms due to alternative splicing (Missler & Sudhof, 1998). Alternative splicing can also be altered by changes to the environment or exposure to stimuli, for example an increase in the AChE (acetylcholinesterase) transcript containing intron 4 has been found due to forced swimming stress in mice and AChE protein inhibition (Kaufer, Friedman, Seidman, & Soreq, 1998).

Alternative splicing is also greatly regulated by tissue type and development specific splicing factors, such as the neuronal splicing factors neuro-oncological ventral antigen 1 (NOVA1) and neuro-oncological ventral antigen 2 (NOVA2), which bind YCAY (Y = C or U) elements (Ule et al., 2005). Mice knockouts show important roles for these genes in motor neuron function and brain development. NOVA2 knockout mice show mis-localization to lower layers of neurons from cortical layers II, III and IV (Yano, Hayakawa-Yano, Mele, & Darnell, 2010). Different polypyrimidine tract binding proteins (PTBPs) are also found to be brain or development specific, with PTBP1 expressed in neural stem and progenitor cells but not in mature neurons, PTBP2 expressed in neurons, and PTBP3 not expressed in neurons, or neural cell types (Spellman et al., 2005). PTBP1 has also been found to play a large role in preventing...
neuronal differentiation, with knockdown of PTBP1 in many cell types, including HeLa cells, human embryonic carcinoma cells, human retinal epithelial cells, primary mouse embryo fibroblasts and mouse neural progenitor cells, resulting in the cells exhibiting characteristics of neuronal cells such as neurite outgrowth and expression of neuronal markers (Xue et al., 2013). Knockdown of PTBP1 also caused upregulation of PTBP2, which is generally targeted for nonsense-mediated decay (NMD) due to exon 10 skipping in the presence of PTBP1 (Xue et al., 2013).

Serine/arginine repetitive matrix protein 4 (SRRM4 or nSR100) is another brain specific splicing factor that has been found to play a role in neuronal fate and differentiation. Knockdown of this protein has been reported to inhibit neuronal differentiation and disrupt cortical layering, neurite outgrowth and axon guidance (Quesnel-Vallieres, Irimia, Cordes, & Blencowe, 2015; Raj et al., 2011). nSR100 has also been found to regulate neural microexons, which are 3 to 27 nucleotides in length (Quesnel-Vallieres et al., 2015).

The RBFOX family of RNA-binding proteins, consisting of RBFOX1, RBFOX2, and RBFOX3, are important regulators of splicing in the brain and are expressed at different developmental stages in different neuronal cell types (McKee et al., 2005). The expression pattern is complex, with different family members expressed in specific neural cell types in early development, differentiation and migration.

The large number of brain-specific splicing regulators and their complex expression patterns suggests that alternative splicing in the brain is highly regulated, and that even small changes to the splicing code and its regulators can lead to dysfunction and disease.

1.2.5 Alternative Splicing in Psychiatric Disorders

Due to the lack of causal and protein-altering variants in psychiatric disorders, the high frequency of intronic variants identified by GWAS in psychiatric disorders and the evidence of splicing abnormalities in other neurological disorders, it is predicted that the vast majority of risk variants underlying psychiatric disorders change gene expression,
specifically transcription and splicing. Unlike Mendelian disorders, the role of splicing in complex disorders, such as neurodevelopmental and psychiatric disorders, are much less well known or understood. However, changes in splicing have been identified in tissues from individuals with schizophrenia, Autism Spectrum Disorder (ASD) and bipolar disorder. These include global changes in alternatively spliced transcripts that have been reported when analyzing the transcriptomes of patients compared to controls, genetic variants associated with quantitative changes in transcript isoforms (splicing quantitative trait loci or sQTLs) that are enriched in loci associated with psychiatric disorders and alterations in splicing networks due to altered splicing trans factors that have been found to be altered in patients with psychiatric disorders. Aberrant splicing has been identified in psychiatric disorders, however there is a lack of information about the mechanism of these splicing disruptions and how they may relate to disease risk and progression (Black & Grabowski, 2003; Clinton, Haroutunian, Davis, & Meador-Woodruff, 2003).

1.2.5.1 Alternative Splicing in Autism Spectrum Disorder (ASD)
Evidence of altered gene expression of splicing regulators and differentially spliced genes has been identified in tissues from patients with ASD. In a study of gene expression in RNA isolated from blood samples from boys with ASD compared to typically developing boys, 53 genes were identified to have differential exon usage (Stamova et al., 2013). Multiple splicing regulators were identified among the 53 differentially spliced genes including SFPQ (splicing factor proline/glutamine rich), SRPK1 (serine/arginine-rich splicing factor kinase 1), SRSF11 (serine/arginine-rich splicing factor 11), SRSF2IP (splicing factor, arginine/serine-rich 2-interacting protein), FUS (fused in sarcoma), and LSM14A (SCD6 homolog A-yeast) (Stamova et al., 2013).

Alterations of gene expression of the RBFOX family of splicing regulators affecting splicing networks have also been implicated in ASD. A reduction in the transcript levels of the RBFOX family of RNA-binding splicing regulators has been identified in brain tissues from patients with ASD, with these tissues also having a misregulation of 5-10% of alternative exons (Voineagu et al., 2011; Weyn-Vanhentenryck et al., 2014). Comparison of the RBFOX targets to the Simons Foundation Autism Research Initiative
database (SFARI) identified 48 RBFOX targets that were previously identified as ASD candidate genes and suggests that the RBFOX family may contribute to ASD through the dysregulation of these 48 target genes (Weyn-Vanhentenryck et al., 2014). The RBFOX genes *RBFOX1* and *RBFOX3* have also previously been implicated in ASD as well as other neural disorders including epilepsy, Rolandic epilepsy, intellectual disabilities, attention deficit-hyperactivity disorder and schizophrenia (Bhalla et al., 2004; Elia et al., 2010; Hamshere et al., 2009; Lal et al., 2013; Martin et al., 2007; Sebat et al., 2007; B. Xu et al., 2008). This further implicates the RBFOX family of proteins and their targets in the pathogenicity of ASD, language, cognition and psychiatric disorders.

Similar to the reduction of the RBFOX family, a reduction in the transcript levels of the neural specific nSR100/SRRM4 protein has been identified in one third of individuals with idiopathic ASD, with tissues from those patients with reduced nSR100/SRRM4 having dysregulation of splicing of 30-40% of brain-specific microexons (Irimia et al., 2014). Mice with expression of 50% of wild-type levels of nSR100/SRRM4 due to a frameshifting deletion of nSR100/SRRM4 on one chromosome were found to have altered synaptic transmission and neuronal excitability, with an increase in glutamatergic synapses and a decrease in GABAergic synapses (Quesnel-Vallieres et al., 2016). These mice also demonstrated ASD behavioural features such as sensitivity to different environmental stimuli, including a decrease in pre-pulse inhibition to the startle response which is a characteristic of ASD, as well as deficits in social behaviour, which was found to be more profound in male mice compared to female mice, also representative of the higher prevalence of ASD in males (Quesnel-Vallieres et al., 2016). nSR100 function was found to be dependent on neuronal activity, with depolarization of primary neurons by KCl treatment causing a decrease in nSR100/SRRM4 protein expression leading to an increase in microexon skipping (Quesnel-Vallieres et al., 2016). This suggests a potential link between increased neuronal activity in ASD with a decrease in nSR100 and an increase in microexon skipping.

An expanded sample of post-mortem cortex tissues from 48 ASD patients and 49 controls confirmed altered differential splicing in the cortex with 1,127 splicing events in 833
genes identified (Parikshak et al., 2016). Most of these events involved exclusion of neuron-specific exons and were enriched in genes previously shown to be regulated by neuronal activity. Differential splicing was correlated with differential expression of brain-specific splicing regulators including RBFOX1, nSR100, and PTBP1.

These studies in total point to a major role of altered expression of splicing regulators in ASD with the consequence of altered exon usage in the brain resulting in expression changes in neuronal genes.

1.2.5.2 Alternative Splicing in Bipolar Disorder

Splice variants have been identified for a number of risk genes for bipolar disorder. For example, associated genetic variants in the ankyrin-3 gene (ANK3) have been identified by GWAS of bipolar disorder, and associated markers in this gene have been consistently replicated (Ferreira et al., 2008; Muhleisen et al., 2014). Association with schizophrenia has also been reported, however with less significant results (Nie et al., 2015). One of these variants, rs41283526, was recently identified to be located in a splice site of a small 54 nucleotide exon of ANK3, with the minor allele of this variant creating a loss-of-function site and lack of splicing of this small exon into the ANK3 transcript (Hughes et al., 2016). This minor allele and exon skipping event are thought to have protective effects against bipolar disorder and schizophrenia (Hughes et al., 2016).

Neuregulin 3 (NRG3) has also been implicated in schizophrenia phenotypes, including delusion severity in schizophrenia, and bipolar disorder (Kao et al., 2010; Meier et al., 2013). NRG3 has many different transcript isoforms that can be divided into classes I-IV based on exon homology. Paterson et al. (2017) looked at the expression of each of the NRG3 classes in the brain throughout development as well as in association with bipolar disorder. In the dorsolateral prefrontal cortex (DLPFC), NRG3 class I was most highly expressed during the neonatal and infancy periods, classes II and IV were most highly expressed during the fetal and neonatal periods and class III expression was consistently expressed throughout development (Paterson et al., 2017). NRG3 classes I and II were more highly expressed in bipolar disorder patients than controls, and an increase in NRG3
classes II and III expression was found to be associated with the T allele of rs10748842 and the G allele of rs6584400, both of which were associated with bipolar disorder and schizophrenia (Kao et al., 2010; Paterson et al., 2017).

1.2.5.3 Alternative Splicing in Schizophrenia

Global changes in alternative splicing have been identified in schizophrenia patients compared to controls. One study examined the level of exon expression in blood and brain samples from schizophrenia patients and controls and then compared the genes of exons with differential expression between in schizophrenia patients and controls in both blood and brain tissue samples to genotype data from these samples and determined associations between SNVs and differential exon expression in four genes between schizophrenia patients and controls (Oldmeadow et al., 2014). Previous studies have identified changes in gene expression based on schizophrenia risk alleles that affect splicing, such as a SNV in Neuregulin 1 (NRG1) that was associated with lower levels of the NRG1-IVNV isoform in the cortex and a SNV that caused lower levels of a DLG1 isoform in early-onset schizophrenia patients (Paterson, Wang, Kleinman, & Law, 2014; Uezato et al., 2015). Kao et al. (2010) found that NRG3 classes I and IV were more highly expressed in the DLPFC of schizophrenia patients compared to controls. Transcript isoforms of ERBB4 that include exon 16 and exon 26 have been reported to be increased in the DLPFC of schizophrenia patients compared to controls, with different schizophrenia risk SNVs associated with each of these changes (Law, Kleinman, Weinberger, & Weickert, 2007). Another study identified over 1000 genes that show differential splicing and gene expression in schizophrenia patients compared to controls (Wu et al., 2012).

For a few associated variants, molecular studies have sought to identify the functional mechanism by which variants alter splicing. The dopamine receptor D2 (DRD2) gene has long been studied as a risk gene for psychiatric disorders and large GWAS have identified significant markers in the proximity of the DRD2 gene associated with schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). A DRD2 rs1076560(T) variant located in intron 6 was recently found to be
associated with alternative splicing of exon 6 (short isoform) and schizophrenia risk in the Han Chinese population (Cohen et al., 2016). It was also determined that this variant exerts decreased binding affinity for an alternative splicing regulator, Zinc Finger RAN-Binding Domain Containing 2 (ZRBAN2), using oligonucleotides with the variants, and was associated with a decreased ratio of short:long DRD2 isoforms in cell culture in an minigene assay with co-expression of ZRBAN2. Further, altered ratios of DRD2 isoforms were observed in brain tissues samples with the GG homozygous genotype group of rs1076560 associated with a significant increase in the short:total DRD2 mRNA expression ratio relative to those with the GT or TT genotypes (Cohen et al., 2016).

The molecular mechanism behind the second most highly schizophrenia-associated region identified by the recent schizophrenia GWAS, the 10q24.32 region, has also been extensively studied (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Recently, a group investigating this region identified a novel AS3MT transcript with exon 2 and 3 skipping that is associated with schizophrenia and found most highly expressed in the brain (Li et al., 2016). This novel transcript was found to be associated with the three repeat major allele of a variable number tandem repeat (VNTR) in the 5’UTR of AS3MT. However, the mechanism by which the three repeat VNTR could be causing this double exon skipping is not known, therefore it could be due to another variant in linkage disequilibrium (LD) with this VNTR that is causing this alternative splicing. Thus, this alternatively spliced transcript is implicated in schizophrenia but the variant causing this splicing change is still unknown.

The synthesis of the inhibitory neurotransmitter γ-aminobutyric acid (GABA) is regulated by glutamic acid decarboxylase (GAD) genes. Alternative transcripts of the GAD2 gene have been identified in the human brain, including a full length GAD2 transcript coding for the 65kDa GAD enzyme and a truncated GAD2 transcript that lacks an in-frame stop codon and may be subject to NMD (Davis et al., 2016). The transcripts were shown to be differentially expressed across the lifespan with the truncated transcript being a fetal predominant transcript. The full length GAD2 transcript was found to have decreased expression in DLPFC brain tissues from schizophrenia and bipolar disorder patients.
compared to controls, while the truncated \textit{GAD2} transcript was found to be increased in bipolar disorder patients and decreased in schizophrenia patients’ tissues compared to controls (Davis et al., 2016). No differences were observed in the DLPFC from depression samples (Davis et al., 2016). Expression of the full length transcript was also correlated with nicotine exposure and suicide completion in schizophrenia samples, potentially complicating correlations with disease (Davis et al., 2016).

1.3 Splicing Predictions and Datasets
Due to the identification of cis- and trans-acting factors that contribute to splicing, multiple tools have been developed with the aim of predicting splicing patterns and differential splice sites. Two main methods exist for identifying sQTLs. The first method involves computational methods, in which a model is trained based on a framework of known splicing motifs and is designed so that the model will be able to be given an input, have splicing features extracted and determine how that sequence will splice (Z. Wang & Burge, 2008). For these methods, known splicing features, including splice sites, branchpoint sequences, ISEs, ESEs, ISSs and ESSs are used by the model to predict splicing changes. However, the predicted splicing changes may not necessarily have been observed in the human population.

The second method involves identifying quantitative changes in gene structure that are correlated with genetic variation. Large, genotyped, RNA-seq datasets are now providing the correlation of quantitative changes in alternative splicing and genetic variation. The most ambitious of these is the Genotype Tissue Expression (GTEx) project (https://gtexportal.org/home/) that is creating a database of RNAseq and genotype data, using tissues from hundreds of individuals and multiple tissues. The project has developed analysis methods, and is a resource to study the relationship of genetic variation with gene regulation to identify associations between genetic variants and quantitative changes in gene expression (expression quantitative trait loci or eQTLs) and associations between genetic variants and quantitative changes in transcript isoforms (splicing quantitative trait loci or sQTLs) (GTEx Consortium, 2015). The project arose from the need to understand the relationship between gene expression and genetic
variants, given that the majority of GWAS variants are not found to cause a protein codon change. The GTEx project aimed to create these datasets and methods to identify eQTLs and sQTLs that are associated with complex genetic disorders, and that therefore may be contributing to risk for these disorders through their change in gene expression or alternative transcript usage. Multiple pipelines have been created using the GTEx data to identify sQTLs, including Altrans, sQTLSeekeR and protein truncating variants (PTVs).

The CommonMind Consortium is another effort to understand gene expression changes associated with genetic variation specifically in the brain. Data from over 1,000 post-mortem brain samples from individuals with schizophrenia, bipolar disorder and no known psychiatric disorder has been collected by the CommonMind Consortium (https://www.synapse.org/#!Synapse:syn2759792/wiki/69613). This includes data from multiple brain regions. Genotype and RNAseq data from the DLPFC of the CommonMind Consortium samples has been used to identify sQTLs (Takata, Matsumoto, & Kato, 2017).

Specific models using these methods and their datasets are discussed here and their predictions will be used to identify SNVs affecting splicing associated with schizophrenia and bipolar disorder.

### 1.3.1 SPIDEX

Xiong et al. (2015) created a model that determines the splicing code and have identified 20,813 out of 658,420 SNVs tested that influence splicing. Their model can predict the splicing regulation of any triplet exon sequence. First, they used 10,689 exons to train the model to determine the percent spliced in (PSI) of the central exon in a triplet of exons. The ability of the model to detect differences in splicing based on different genotypes was also tested and it was found that the model was correct in predicting the direction of the PSI change in 73% of pairs of individuals with differing SNVs.

To determine the affect of both common and rare SNVs on splicing regulation, the model was used to compute the PSI of the sequence with and without the SNV for 16 tissues.
The tissue with the largest difference in PSI was then studied for the potential effect. This method identified multiple exonic (missense, nonsense and synonymous) SNVs, intronic SNVs close to splice sites and even intronic SNVs that are over 30 nucleotides away from splice sites but that are still predicted to cause splicing changes. These intronic SNVs more than 30 nucleotides from a splice site were predicted to disrupt splicing regulation 9 times as much as common SNVs in the same region (Xiong et al., 2015).

1.3.2 Altrans
Altrans identifies sQTLs by identifying SNVs that may affect splicing through their association with differences in exon junction expression levels (Ongen & Dermitzakis, 2015). Altrans analyzes the +/-1 Mb region around the transcript start site (TSS) to identify novel and annotated splicing events that result from changes in splice junction usage (GTEx Consortium, 2015).

A study performed by the GTEx Consortium using the Altrans method and RNA sequencing data from nine tissues (subcutaneous adipose, tibial artery, left ventricle of the heart, lung, skeletal muscle, tibial nerve, sun-exposed skin, thyroid and whole blood) identified ~1900 genes with sQTLs per tissue (GTEx Consortium, 2015). 80% of the sQTLs identified by Altrans are exon-skipping events (GTEx Consortium, 2015).

1.3.3 sQTLSeekeR
sQTLSeekeR identifies sQTLs by identifying SNVs that may affect splicing through their association with differences in the relative abundance of transcript isoforms (Monlong, Calvo, Ferreira, & Guigo, 2014). It is implemented in an R package that computes the variability in splicing ratios of a gene using a distance-based approach and computes the genotype variability of the splicing ratios to test the association between a SNV and a gene (Monlong et al., 2014). The transcripts with the highest expression levels are identified for the SNVs predicted to effect splicing. sQTLSeekeR analyzes annotated transcript isoforms, analyzes SNVs within the body of a gene (+/- 5kb) and can detect any variation in the relative abundance of transcript isoforms (Monlong et al., 2014). It is
predicted that changes in the relative abundance of transcript isoforms is likely to be most detrimental for genes with many splice isoforms for which a small change in the ratio of isoforms may cause a significant disruption (Monlong et al., 2014).

A study performed by the GTEx Consortium using the sQTLSeeker method and RNA sequencing data from nine tissues (subcutaneous adipose, tibial artery, left ventricle of the heart, lung, skeletal muscle, tibial nerve, sun-exposed skin, thyroid and whole blood) identified ~250 genes with sQTLs per tissue (GTEx Consortium, 2015). 60% of these sQTLs detected using sQTLSeeker were complex splice events and were mostly mutually exclusive exons (GTEx Consortium, 2015).

1.3.4 Protein Truncating Variants (PTV)
Many protein truncating variants (PTVs), which cause a shorter than normal protein product, have been found to be disease causing. Rivas et al. (2015) developed a predictive model to determine the effects of PTVs on transcripts. They tested their model using DNA and mRNA sequences from the GTEx study and identified 16,286 variants that may cause protein truncation (Rivas et al., 2015). In their study, it was found that common PTVs caused a transcript to be more weakly expressed and expressed in a tissue-specific manner than transcripts that did not contain common PTVs.

As well as resulting in truncated proteins, transcripts with PTVs have also been predicted to be degraded through nonsense-mediated mRNA decay (NMD) (Chang, Imam, & Wilkinson, 2007). NMD was generally thought to function to remove nonfunctional truncated proteins that could accumulate toxically (Neu-Yilik, Gehring, Hentze, & Kulozik, 2004). It has now been shown that alternative splicing naturally creates mRNA that will be targeted for NMD as a form of post-transcriptional regulation (Hillman, Green, & Brenner, 2004; Lewis, Green, & Brenner, 2003; Neu-Yilik et al., 2004). In the Rivas et al. (2015) study, it was found that NMD occurred more often for transcripts with rare PTVs than common PTVs, with NMD occurring 54.3% and 35.7% of the time for rare and common PTVs respectively. Tissue specific differences in NMD due to PTVs
were also identified (Rivas et al., 2015). Therefore common PTVs predicted by Rivas et al. (2015) may be detected through cDNA expression.

### 1.3.5 DLPFC sQTLs
Takata, Matsumoto & Kato (2017) used CommonMind Consortium genotype and RNAseq data from the DLPFC of 206 control samples to identify DLPFC sQTLs. The vast-tools software package was used on the RNAseq data to identify 102,469 alternative splicing events including exon skipping, intron retention and alternative splice site usage. These alternative splicing events were then correlated to genotype and 8,966 sQTLs in 1,341 genes were identified. These sQTLs were found to be enriched in exonic and H3K4me3 regions as well as among GWAS disease-associated regions (Takata et al., 2017).

### 1.4 Protein Structure and Function Prediction
Identifying protein structure and function of novel proteins arising from alternatively spliced transcripts has become a large and necessary challenge as the increase in high-throughput genomic sequencing has identified more predicted protein sequences. Around 40% of proteins in the NCBI database do not have an annotated function (Bernardes & Pedreira, 2013). The first protein prediction approaches involved comparative or homology modeling, in which structure and function of novel proteins is based on sequence similarity to previously annotated proteins. These methods use sequence similarity between an inputted or target protein sequence and protein sequences already found and annotated in the Protein Data Bank, for which information already exists about these proteins (Q. Xu et al., 2008). If the inputted or target sequence aligns to homologous proteins in the Protein Data Bank, structural information can be inferred about the target protein based on what is known about the homologous proteins, which can further lead to information about function and localization of the target protein. Comparative modeling only aligns sequences of homologous proteins and relies on sequence similarity of at least 30% (Khor, Tye, Lim, & Choong, 2015). This type of method only works if the protein is conserved and if there is already information about this protein or a similar protein in protein databases, however it is not ideal for novel
proteins and will likely provide little information about their domains, structure and potential function.

Thus, more sophisticated computational approaches have been developed that use i) local sequence similarities or motif-based methods, ii) 3D structure-based methods and iii) genomic context- and network-based methods.

Motif and domain-based methods compare local sequence similarities of a target protein to identify potential known domains or motifs, which can help provide information about function. Domains are sequences that are known to fold into a certain structure and are often functionally annotated, therefore identifying known domains in the target sequence can provide information about the function of the protein. Motifs are shorter sequences that can provide information about catalytic sites, binding sites and potentially localization sequences.

3D protein structure-based prediction methods, namely protein threading and ab initio methods, have become more popular as they can be used for protein sequences for which homology based methods do not provide much information. Protein threading methods compare an inputted protein sequence to structural information found in protein templates in the Protein Data Bank, such as similar fold or structural motifs (Q. Xu et al., 2008). Protein threading uses secondary structure information, sequence profile based on homologous proteins, pairwise interactions between nearby residues and solvent accessibility. The main goal of protein threading is to correctly align structural similarities of the target protein to the template protein. PSI-BLAST and ClustalW are often used to generate sequence profiles of homologous proteins similar to the target protein sequence.

Genomic context and network-based methods use high-throughput gene expression and protein interaction data sets. Genes that are co-expressed may have similar functions or be involved in similar pathways, therefore identifying proteins of known function that are expressed with a protein of unknown function can help to infer the unknown protein’s
function. As well, identifying proteins that a protein of unknown function interacts with can be important in identifying its potential function. These methods assume that proteins that are co-expressed or interact with each other are likely to function similarly.

I-TASSER, a protein prediction approach developed by Zhang (2009), uses a combination of these protein function prediction methods to provide an in-depth analysis of the potential function of an input amino acid sequence (Y. Zhang, 2009). The I-TASSER pipeline consists of i) identification of protein threading templates, ii) assembly of simulated structures, iii) selection and refinement of model and iv) functional annotation based on identified structures (Yang et al., 2015). The first step uses the LOMETS protein threading algorithm to identify target-template alignments with similar folds or structures from the Protein Data Bank library. The next step involves reassembly of the Protein Data Bank templates into structural models using Monte Carlo simulations and ab initio modeling. Finally, for functional annotation, COACH matches the predicted structural models to the enzyme classification, gene ontology (GO) and ligand-binding site libraries.

PredictProtein is a protein structure and function prediction server that aligns a target protein sequence to over 30 tools including homology-based, structural and functional databases in order to predict structure and function (Yachdav et al., 2014). PredictProtein predicts secondary structure, solvent accessibility features, subcellular localization, protein-protein interactions, binding sites and GO terms based on an input amino acid sequence.

1.5 Hypothesis, Aims and Overview of Thesis
Many genetic variants have been identified that are associated with schizophrenia and bipolar disorder, however the actual functional role of these variants is not known in many cases. We hypothesize that some of these variants will alter splicing, which will in turn alter gene expression, contributing to the risk for schizophrenia and bipolar disorder. In order to test this hypothesis, two main aims are achieved in this study: i) the identification of schizophrenia and bipolar disorder-associated SNVs that are
bioinformatically predicted to alter splicing and ii) an experimentally tested genotype-phenotype correlation between schizophrenia and bipolar disorder-associated SNVs and altered levels of splicing isoforms due to this specific genotype. Seven datasets of SNVs were mined to prioritize variants for testing, including five datasets that contain SNVs with predictions of altered splicing (GTEx Consortium, 2015; Rivas et al., 2015; Takata et al., 2017; Xiong et al., 2015), a sixth dataset containing SNVs associated with schizophrenia identified through the Psychiatric Genetics Consortium (PGC) (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) and a seventh dataset containing SNVs associated with bipolar disorder (Hou et al., 2016). The relationship between SNVs associated with schizophrenia or bipolar disorder and splicing was tested using DNA and RNA from brain tissue samples from 92 individuals, including schizophrenia and bipolar disorder patients and controls. These individuals were then genotyped for the candidate SNVs and the relationship of differential splicing based on genotype using the cDNA from these tissues was determined. Based on the genotype and cDNA expression results, the effect of these SNVs on gene expression was determined using bioinformatics and functional implications of specific SNVs associated with schizophrenia and bipolar disorder were identified.
Chapter 2

Materials and Methods

2.1 Identification of Schizophrenia and Bipolar Disorder Candidate SNVs
One dataset from SPIDEX, three datasets from GTEx: Altrans, sQTLSeekeR and PTVs, and one dataset of DLPFC sQTLs were mined to identify which of the SNVs from these datasets were also GWAS or near GWAS significantly associated with schizophrenia or bipolar disorder from the schizophrenia and bipolar disorder GWASs. This identified a set of candidate SNVs associated with both a splicing change and schizophrenia or bipolar disorder.

Candidate SNVs were then further filtered based on the gene in which they were predicted to cause a splicing change. Inclusion criteria for these genes were based on expression in the adult hippocampus or DLPFC brain regions, previously reported association to schizophrenia or bipolar disorder and a function that may relate to schizophrenia or bipolar disorder. Further, genes were ruled out if a clear candidate was already identified in their associated region.

2.2 Brain tissue samples
Brain tissue samples from 8 schizophrenia patients, 5 bipolar patients and 89 controls were obtained from the Harvard Brain Tissue Resource, the Douglas Brain Bell Canada Brain Bank and the National Institute of Child Health and Human Development (Table 1).
Table 1. Demographics of individuals from whom brain sample tissues were collected.

<table>
<thead>
<tr>
<th>Disease State:</th>
<th>Schizophrenia</th>
<th>Bipolar Disorder</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td># of samples:</td>
<td>8</td>
<td>5</td>
<td>89</td>
</tr>
<tr>
<td>Gender:</td>
<td>8 male, 0 female</td>
<td>4 male, 1 female</td>
<td>57 male, 32 female</td>
</tr>
<tr>
<td>Mean age:</td>
<td>39.75</td>
<td>47</td>
<td>41.22</td>
</tr>
<tr>
<td>Ethnicity:</td>
<td>8 Caucasian</td>
<td>5 Caucasian</td>
<td>78 Caucasian, 10 African American, 1 Mixed</td>
</tr>
</tbody>
</table>

The control postmortem brain tissue samples were from healthy individuals less than 65 years of age with no neurological or psychiatric disorder. The post mortem interval for these tissues was less than 24 hours. Exclusion criteria for the control brain tissue samples included individuals where neurological disease may have influenced the tissue quality such as stroke, prolonged hypoxic death, Alzheimer’s disease or other psychiatric or neurological disorder. The gender, age, ethnicity and cause of death were provided for each individual. The aim of this study was to identify alternative splicing changes associated with psychiatric disorder risk alleles. Because the risk SNVs analyzed in this study are common in the population, with the lowest minor allele frequency of any SNV in this study being 0.0755, brain samples from controls would still carry the risk alleles, and therefore were used to study the relationship between these variants and alternative splicing. Further, as the purpose of this study is to investigate the effect of schizophrenia and bipolar disorder risk variants on alternative splicing, all analyses are performed with samples separated by genotype regardless of disease status.

Two brain regions relevant to schizophrenia and bipolar disorder were collected: dorsolateral prefrontal cortex (DLPFC) Brodmann area 9 (BA9) and hippocampus. These regions were chosen as they have previously been found to be dysregulated and have physical differences between control individuals and individuals with schizophrenia and bipolar disorder, including deficits in dopamine release and dendritic spine density in the
prefrontal cortex (Konopaske, Lange, Coyle, & Benes, 2014) and hippocampal volume differences (McClure et al., 2013).

2.3 SNV Genotyping
DNA was extracted from 50mg of each brain tissue. The brain tissues were homogenized in a plastic disposable pestle in a 2ml tube. 1ml of DNA lysis buffer (10mM Tris-HCl pH 8.2, 400mM NaCl and 2mM EDTA) was added followed by further homogenizing using a 1ml syringe and disposable 0.9 gauge needle. 5ul of RNase A (100ug/ml) was added and then incubated at 37°C for 1 hour. 100ul of 10% SDS and 10ul of proteinase K (20mg/ml) were added and then incubated at 50°C overnight while rotating. 1ml of 25:24:1 phenol/chloroform/isoamyl alcohol was added, centrifuged at 13,000 rpm for 10 minutes and the upper phase was transferred to new Eppendorf tubes. 50ul of 3M NaOAC and 1ml of 100% ethanol were added, incubated at 20°C for 1 hour and centrifuged at 13,000rpm for 10 minutes. The supernatant was removed and the pellet was washed once with 70% ethanol and then soaked in 70% ethanol to remove the precipitation that stuck to the wall of the tube. The column was centrifuged at 13,000rpm for 5 minutes, the ethanol was removed and the pellet was dried at room temperature for 10 minutes. 150ul of TE buffer was then added to dissolve the DNA on a shaker overnight.

Extracted DNA from all 102 brain samples were run on a 2% agarose gel to ensure the quality of the DNA and that there was no apparent degradation. 92 samples were high quality and were sent for genotyping. Array-based genotyping was performed using the OmniExpress Platform, which genotyped 733,202 SNVs. This captured 91% of HapMap markers with minor allele frequency (MAF) >5% in the Caucasian European (CEU) population with linkage disequilibrium (LD) coverage of $r^2 > .80$. PLINK was used for quality control and analyses (http://pngu.mgh.harvard.edu/purcell/plink/) (Abecasis, Cardon, & Cookson, 2000). Genotypes were checked for Hardy-Weinberg Equilibrium and markers with MAF<0.05 and p<0.0001 were removed. Samples with higher than 5% missing genotypes and SNVs with call rates less than 97% were removed. Ethnicity of
the samples was checked using principal component analysis. Genotypes were imputed using the Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html) using 1000 Genomes Phases 1 and 3 (Das et al., 2016).

2.4 Expression of Alternative Splicing Isoforms

2.4.1 Identification of Splicing Variants
The alternative transcript event associated with each disease-associated SNV was determined (ex. exon-skipping, intron retention, alternative first exon usage) based on information from the respective splicing dataset from which the SNV was identified. Transcript isoform expression was then tested based on these predictions using DLPFC and hippocampal cDNA from the brain tissue samples. RNA was extracted from the brain tissues using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). Samples were placed in a 2ml tube on dry ice, homogenized and then 1ml of TRIzol reagent was added and mixed. The brain tissue and TRIzol were further homogenized using a 1ml syringe and 0.9 gauge needle and then transferred to a 1ml Eppendorf tube. The samples were mixed well and left to stand at room temperature for 5 minutes. 200ul of phenol-chloroform was added to each sample, shaken vigorously for 20 seconds and then incubated at room temperature for 3 minutes. The samples were centrifuged at 13,000rpm for 15 minutes at 4°C in order to create 3 distinct layers: a lower red phenol-chloroform phase, an interphase and a colorless upper phase, with the RNA found in the upper phase. The aqueous upper phase was transferred to a fresh 1.5ml Eppendorf tube and the RNA was precipitated out by adding 500ul of 100% isopropanol. The sample was mixed and incubated at room temperature for 10 minutes. The sample was centrifuged at 13,000rpm for 15 minutes at 4°C and the supernatant was removed. The pellet was gently vortexed in 1ml of 75% ethanol, spun down at 7,500rpm for 5 minutes and the supernatant was removed. The pellet was left to air dry for 10 minutes in the cabinet and then resuspended in 87ul of DEPC dH₂O (pH 7) and redissolved on ice for 10 minutes. The RNA samples were then purified using the RNeasy spin column kit (Qiagen Crawley, UK) following the manufacturers protocol with the additional DNAse step added to ensure removal of genomic DNA. The quality of the RNA extracted from the 102 individuals used to make
the cDNA was checked using the Agilent 2100 Bioanalyzer and only RNA with RNA integrity number (RIN) > 6 were used. Due to differing RIN between the different regions from the 102 brain samples, this led to 89 DLPFC brain samples with RIN > 6 and 65 hippocampal brain samples with RIN > 6 which were used to make cDNA. cDNA was synthesized using the ImProm-II Reverse Transcriptase protocol (Promega, Madison, WI, USA).

For exon-skipping events, primers were designed flanking the predicted spliced-out exon such that a fragment could be amplified both with the exon spliced in and out (Figure 3). PCR amplification was performed on all samples using these primers. The PCR amplification was then run on a 2% agarose gel and the presence of splicing isoforms was identified based on size separation of amplified fragments.

**Figure 3. Schematic of the cDNA assay used to identify alternative splicing transcripts with exon skipping.** Primers were designed in the two exons flanking the predicted skipped out exon in order to amplify isoforms with the central exon spliced in or out.

For intron retention events, primers were designed in the exons flanking the retained intron such that a fragment could be amplified both with the intron spliced in and out
(Figure 4). PCR amplification was performed on all samples using these primers. The PCR amplification was then run on a 2% agarose gel and the presence of splicing isoforms was identified based on size separation of amplified fragments.

**Figure 4. Schematic of the cDNA assay used to identify alternative splicing transcripts with intron retention.** Primers were designed in the two exons flanking the predicted spliced in intron in order to amplify isoforms with the intron spliced in or out.

For alternative first exon events, separate primer assays were designed in either of the first exons such that two different PCR reactions could be performed to amplify the transcript with either first exon (Figure 5). PCR amplification was performed on all samples using these primers. The PCR amplification was then run on a 2% agarose gel to determine if there was amplification in the different PCRs that corresponded to the alternative first exons.
Figure 5. Schematic of the cDNA assay used to identify alternative splicing transcripts with alternative first exons. Different forward primers were designed in each of the two predicted first exons with the same reverse primer in the second exon in order to amplify isoforms with alternative first exons.

2.4.2 Quantification of Identified Splicing Variants Using ddPCR

Upon identification of multiple splicing isoforms, a quantitative digital droplet PCR (ddPCR) assay was designed in order to quantify the proportion of each splicing isoform in each sample. Digital droplet PCR (ddPCR) partitions a normal PCR reaction into 10,000-20,000 droplets and then counts the number of droplets that amplified at the end of the PCR cycling in order to determine the number of copies of target template that was in the original reaction. ddPCR works on the basis that only one copy of the target template will be partitioned into each of the droplets, with some droplets also containing no target template. Using ddPCR, the number of copies of each splicing isoform in each brain sample was determined.

Probe-based assays were designed with one assay amplifying the isoform with the exon spliced in and using a HEX fluorescent dye for the probe, and a second assay
amplifying the exon spliced out and using a FAM fluorescent dye for the probe (Figure 6).

Figure 6. Schematic of the probe-based assay used to identify the proportion of each splicing isoform.

These assays were run as duplex ddPCRs on each brain sample. Each droplet with either the full length or exon-skipped transcript, determined by either FAM or HEX fluorescence, was then counted (Figure 7).

Figure 7. Digital Droplet PCR output of probe-based exon skipping assay. Black dots represent negative droplets with no amplification, blue dots represent droplets with FAM fluorescence containing the exon skipped transcript, green dots represent droplets with HEX fluorescence containing the full length transcript and orange dots represent double droplets with both FAM and HEX fluorescence containing the exon skipped and full length transcripts.
The fractional abundance (FA) of the exon skipped transcript compared to the total number of transcripts amplified was determined by the following equation:

$$FA = \frac{A}{A+B}$$

A refers to the copies/uL of the alternatively spliced transcript, in this case the exon skipped transcript, and B refers to the copies/uL of the full length transcript, in this case the transcript with the exon spliced in.

The fractional abundance was determined for each individual brain sample and was then compared to their risk genotype for the candidate SNVs as an ANOVA. The fractional abundance of each sample compared to their genotype was also plotted as a boxplot using plotly in R. For each plot, the upper and lower hinges are the third and first quartiles, respectively, and the upper and lower whiskers are the largest and smallest values contained within 1.5 times the inter-quartile range. Outliers are values that extend above or below 1.5 times the inter-quartile range.

2.5 Predicting the Mechanism of the Splice Variants
Candidate SNVs whose splicing prediction was experimentally validated in the brain tissue samples through the identification of multiple splicing isoforms were input into splicing prediction servers to determine the mechanism by which the SNV change would likely cause a splicing change. If the candidate SNV was found within an exonic or intronic region in close proximity to the location of the predicted splicing change, two splicing prediction servers were used to determine the mechanism by which these SNVs could be causing the splicing change: i) RBPMap, which identifies RNA binding proteins that are predicted to bind to an input sequence (Paz, Kosti, Ares, Cline, & Mandel-Gutfreund, 2014) and ii) SpliceAid, which identifies differentially binding RNA motifs based on the two alleles of a SNV (Piva, Giulietti, Burini, & Principato, 2012). If the candidate SNV was not found within an exonic or intronic region in close proximity to
the location of the predicted splicing change, a literature search into the mechanism behind the splicing change was performed.

2.6 Predicting the Effect on Protein Function

Protein sequences of the splicing isoforms experimentally validated in the brain tissue samples through the identification of multiple splicing isoforms were input into protein prediction servers in order to determine the potential effect on protein function of the alternatively spliced isoforms.

Three sequence-motif based protein prediction methods were employed to analyze functional differences between the full length and alternatively spliced protein sequences: InterPro, PANTHER and Pfam (Finn et al., 2016).

Two automated function prediction methods were also used to determine the potential structural and functional differences between the full-length and alternatively spliced protein sequences: PredictProtein and I-TASSER (Yachdav et al., 2014; Yang et al., 2015). These servers use protein threading and ab initio methods to predict potential structural and binding regions as well as identifying protein domains similar to the input sequence which can aid in inferring function of the target protein.
Chapter 3

Results

3.1 Splicing SNVs associated with Schizophrenia and Bipolar Disorder

63 SNVs from the splicing datasets were found to be a GWAS or near GWAS significant SNV in the schizophrenia GWAS dataset ($p > 10^{-5}$) (Table 2). 142 SNVs from the splicing datasets were found to be a GWAS or near GWAS significant SNV in the bipolar disorder GWAS dataset ($p > 10^{-3}$) (Table 3). Near GWAS significant SNVs were analyzed in this study due to the fact that soon to be published larger datasets of cases and controls may lead these SNVs to become GWAS significant. As well, the bipolar disorder GWAS included a much smaller sample size than the schizophrenia GWAS, which would lead to fewer SNVs associated at more significant $p$ values, therefore the $p$ value cutoff was much lower for bipolar disorder than schizophrenia in order to capture potential disease-associated SNVs predicted to cause a splicing change. Further, SNVs that are not in LD with the index SNVs may still be contributing to disease and may be identified within the SNVs with less significant GWAS $p$ values. Of the 63 schizophrenia-associated SNVs and 142 bipolar disorder-associated SNVs, 37 of these SNVs overlapped. This led to a total of 26 schizophrenia-associated SNVs, 105 bipolar disorder-associated SNVs and 37 schizophrenia and bipolar disorder-associated SNVs found to also be associated with an alternative splicing event that were then further filtered based on which gene they were predicted to alter the splicing of. After filtering, 72 SNVs in 13 genes were chosen to be genotyped and have the associated alternative splicing change identified in the brain tissue sample cDNA. For 8 of these genes ($PPPIR16B$, $STAB1$, $NGEF$, $LRRC48$, $CACNA1D$, $KIF21B$, $SNAP91$ and $PRKAG1$), the predicted alternative splicing event was not identified in any of the brain tissue samples. However, in 6 genes ($APOPT1$, $AS3MT$, $BNIP3L$, $LRRC48$, $NEK4$, and $RPGRI1P1$), the predicted alternative splicing change was identified in the brain tissue samples and these genes were further studied.
Table 2. Schizophrenia and alternative splicing associated SNVs.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Gene</th>
<th>SCZ GWAS p value</th>
<th>Splicing Database</th>
<th>Splicing p value</th>
<th>Splicing Z-score</th>
<th>Splicing rho value</th>
<th>cDNA Tested?</th>
<th>Alternative Splicing Isoforms Identified?</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3740400</td>
<td>AS3MT</td>
<td>6.00E-17</td>
<td>SPIDEX</td>
<td></td>
<td>2.043</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs10431750</td>
<td>APOPT1</td>
<td>5.69E-13</td>
<td>PTV</td>
<td></td>
<td>3.1488E-06</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs4906337</td>
<td>APOPT1</td>
<td>5.72E-13</td>
<td>PTV</td>
<td></td>
<td>2.0799E-05</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs7140568</td>
<td>APOPT1</td>
<td>5.74E-13</td>
<td>PTV</td>
<td></td>
<td>2.0799E-05</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs2535627</td>
<td>NEK4</td>
<td>3.96E-11</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>3.26E-06</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs1108842</td>
<td>NEK4</td>
<td>2.27E-09</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>4.66E-07</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs6769789</td>
<td>NEK4</td>
<td>8.92E-09</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>4.69E-04</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs2071042</td>
<td>NEK4</td>
<td>1.10E-08</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>6.43E-04</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs2535629</td>
<td>NEK4</td>
<td>1.33E-08</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>1.19E-06</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs6445538</td>
<td>NEK4</td>
<td>4.10E-08</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>1.64E-03</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs2276817</td>
<td>NEK4</td>
<td>4.49E-08</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>1.73E-03</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs6445539</td>
<td>NEK4</td>
<td>4.58E-08</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>1.64E-03</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs736408</td>
<td>NEK4</td>
<td>4.58E-08</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>4.37E-06</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs2230535</td>
<td>NEK4</td>
<td>7.78E-08</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>9.15E-10</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs1042779</td>
<td>NEK4</td>
<td>9.26E-08</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>4.59E-10</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs3774354</td>
<td>NEK4</td>
<td>1.91E-07</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>1.15E-09</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs11130317</td>
<td>NEK4</td>
<td>1.94E-07</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>9.15E-10</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs1076425</td>
<td>NEK4</td>
<td>1.99E-07</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>2.48E-08</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs11870660</td>
<td>LRRC48</td>
<td>2.16E-07</td>
<td>SPIDEX</td>
<td></td>
<td>1.994</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs2300149</td>
<td>NEK4</td>
<td>2.80E-07</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>8.46E-10</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs1075653</td>
<td>NEK4</td>
<td>2.99E-07</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>2.27E-08</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs9324</td>
<td>NEK4</td>
<td>3.64E-07</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>2.27E-08</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs2071508</td>
<td>NEK4</td>
<td>3.98E-07</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>2.27E-08</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs11177</td>
<td>NEK4</td>
<td>4.25E-07</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>2.52E-09</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs6976</td>
<td>NEK4</td>
<td>4.40E-07</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>2.52E-09</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs10780035</td>
<td>NEK4</td>
<td>4.74E-07</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>2.52E-09</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs13071584</td>
<td>NEK4</td>
<td>6.32E-07</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>4.77E-09</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs3755799</td>
<td>NEK4</td>
<td>6.85E-07</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>9.83E-09</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs11716747</td>
<td>NEK4</td>
<td>1.56E-06</td>
<td>DLPFC sQTLs</td>
<td></td>
<td>4.77E-09</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SNP</td>
<td>Gene</td>
<td>Effect</td>
<td>p-value</td>
<td>Software</td>
<td>Significant</td>
<td>Interaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------</td>
<td>---------</td>
<td>----------</td>
<td>-------------------</td>
<td>-------------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4687657</td>
<td>NEK4</td>
<td>1.57E-06</td>
<td>1.21E-03</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2289247</td>
<td>NEK4</td>
<td>1.64E-06</td>
<td>4.77E-09</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1866268</td>
<td>NEK4</td>
<td>1.82E-06</td>
<td>4.77E-09</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3755798</td>
<td>NEK4</td>
<td>2.16E-06</td>
<td>4.77E-09</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12635140</td>
<td>NEK4</td>
<td>2.20E-06</td>
<td>4.77E-09</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1362572</td>
<td>RPGRIPI1</td>
<td>7.23E-06</td>
<td>1.54E-26</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2675959</td>
<td>NGEF</td>
<td>1.13E-09</td>
<td>3.415</td>
<td>SPIDEX</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs217323</td>
<td>SNAP91</td>
<td>2.71E-09</td>
<td>5.45E-23</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4368395</td>
<td>PPP1R16B</td>
<td>3.32E-09</td>
<td>7.2000E-05</td>
<td>sQTLseeker</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs217291</td>
<td>SNAP91</td>
<td>3.73E-09</td>
<td>5.45E-23</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs217328</td>
<td>SNAP91</td>
<td>4.16E-09</td>
<td>2.49E-22</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6093098</td>
<td>PPP1R16B</td>
<td>4.85E-09</td>
<td>7.2000E-05</td>
<td>sQTLseeker</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1546977</td>
<td>SNAP91</td>
<td>5.51E-09</td>
<td>2.49E-22</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2022265</td>
<td>SNAP91</td>
<td>6.22E-09</td>
<td>3.83E-19</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs217308</td>
<td>SNAP91</td>
<td>7.88E-09</td>
<td>5.45E-23</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10854216</td>
<td>PPP1R16B</td>
<td>3.66E-08</td>
<td>1.2400E-04</td>
<td>sQTLseeker</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4924832</td>
<td>LRRC48</td>
<td>4.89E-08</td>
<td>2.432</td>
<td>SPIDEX</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4584886</td>
<td>LRRC48</td>
<td>1.60E-07</td>
<td>-2.204</td>
<td>SPIDEX</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7741310</td>
<td>SNAP91</td>
<td>4.38E-07</td>
<td>4.50E-10</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1158059</td>
<td>SNAP91</td>
<td>4.46E-07</td>
<td>5.90E-14</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1171114</td>
<td>SNAP91</td>
<td>1.26E-06</td>
<td>9.77E-10</td>
<td>DLPFC sQTLs</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6682427</td>
<td>STAB1</td>
<td>3.58E-06</td>
<td>-2.492</td>
<td>SPIDEX</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs55742290</td>
<td>ARL6IP4</td>
<td>1.74E-10</td>
<td>-2.412</td>
<td>SPIDEX</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs61884274</td>
<td>ATG13</td>
<td>9.52E-09</td>
<td>2.352</td>
<td>SPIDEX</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1545762</td>
<td>PLEKHO1</td>
<td>1.01E-08</td>
<td>1.0992E-03</td>
<td>PTV</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3208509</td>
<td>PLEKHO1</td>
<td>1.07E-08</td>
<td>1.0992E-03</td>
<td>PTV</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs883563</td>
<td>OGFOD2</td>
<td>3.15E-08</td>
<td>2.014</td>
<td>SPIDEX</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs133335</td>
<td>WBP2NL</td>
<td>1.89E-07</td>
<td>2.6400E-04</td>
<td>sQTLseeker</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7537216</td>
<td>C1orf132</td>
<td>5.00E-07</td>
<td>0.521</td>
<td>Altrans</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2851435</td>
<td>MPHOSPH9</td>
<td>6.65E-07</td>
<td>-1.998</td>
<td>SPIDEX</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1142469</td>
<td>CD46</td>
<td>1.17E-06</td>
<td>3.3300E-07</td>
<td>sQTLseeker</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7541230</td>
<td>CD46</td>
<td>1.31E-06</td>
<td>3.3300E-07</td>
<td>sQTLseeker</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1891423</td>
<td>CD46</td>
<td>2.08E-06</td>
<td>3.3300E-07</td>
<td>sQTLseeker</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2796275</td>
<td>CD46</td>
<td>2.85E-06</td>
<td>3.3300E-07</td>
<td>sQTLseeker</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Bipolar disorder and alternative splicing associated SNVs.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Gene</th>
<th>BD GWAS p value</th>
<th>Splicing Database</th>
<th>Splicing p value</th>
<th>Splicing Z-score</th>
<th>Splicing rho value</th>
<th>cDNA Tested?</th>
<th>Alternative Splicing Isoforms Identified?</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11779570</td>
<td>BNIP3L</td>
<td>9.23E-06</td>
<td>sQTLSeeker</td>
<td>1.01E-04</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs11786753</td>
<td>BNIP3L</td>
<td>1.01E-05</td>
<td>sQTLSeeker</td>
<td>1.01E-04</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs60377146</td>
<td>BNIP3L</td>
<td>1.05E-05</td>
<td>sQTLSeeker</td>
<td>1.94E-04</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs4687657</td>
<td>NEK4</td>
<td>1.48E-05</td>
<td>DLPFC sQTLs</td>
<td>1.21E-03</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs11135938</td>
<td>BNIP3L</td>
<td>1.83E-05</td>
<td>sQTLSeeker</td>
<td>2.62E-04</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs8050354</td>
<td>RPGRIP1L</td>
<td>4.73E-05</td>
<td>DLPFC sQTLs</td>
<td>9.12E-42</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs1014969</td>
<td>NEK4</td>
<td>4.97E-05</td>
<td>DLPFC sQTLs</td>
<td>8.00E-09</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs3755799</td>
<td>NEK4</td>
<td>5.03E-05</td>
<td>DLPFC sQTLs</td>
<td>9.83E-09</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs1421091</td>
<td>RPGRIP1L</td>
<td>6.63E-05</td>
<td>DLPFC sQTLs</td>
<td>7.31E-41</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs736408</td>
<td>NEK4</td>
<td>9.18E-05</td>
<td>DLPFC sQTLs</td>
<td>4.37E-06</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs12635140</td>
<td>NEK4</td>
<td>9.64E-05</td>
<td>DLPFC sQTLs</td>
<td>4.77E-09</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs13071584</td>
<td>NEK4</td>
<td>9.67E-05</td>
<td>DLPFC sQTLs</td>
<td>4.77E-09</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs1866268</td>
<td>NEK4</td>
<td>9.68E-05</td>
<td>DLPFC sQTLs</td>
<td>4.77E-09</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs11716747</td>
<td>NEK4</td>
<td>9.70E-05</td>
<td>DLPFC sQTLs</td>
<td>4.77E-09</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs2289247</td>
<td>NEK4</td>
<td>9.87E-05</td>
<td>DLPFC sQTLs</td>
<td>4.77E-09</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs3755798</td>
<td>NEK4</td>
<td>1.04E-04</td>
<td>DLPFC sQTLs</td>
<td>4.77E-09</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs1042779</td>
<td>NEK4</td>
<td>1.23E-04</td>
<td>DLPFC sQTLs</td>
<td>4.59E-10</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs2535629</td>
<td>NEK4</td>
<td>1.36E-04</td>
<td>DLPFC sQTLs</td>
<td>1.19E-06</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs2071508</td>
<td>NEK4</td>
<td>2.14E-04</td>
<td>DLPFC sQTLs</td>
<td>2.27E-08</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs11130317</td>
<td>NEK4</td>
<td>2.15E-04</td>
<td>DLPFC sQTLs</td>
<td>9.15E-10</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs1108842</td>
<td>NEK4</td>
<td>2.34E-04</td>
<td>DLPFC sQTLs</td>
<td>4.66E-07</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs2230535</td>
<td>NEK4</td>
<td>2.40E-04</td>
<td>DLPFC sQTLs</td>
<td>9.15E-10</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs1075653</td>
<td>NEK4</td>
<td>2.62E-04</td>
<td>DLPFC sQTLs</td>
<td>2.27E-08</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs9324</td>
<td>NEK4</td>
<td>2.72E-04</td>
<td>DLPFC sQTLs</td>
<td>2.27E-08</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs2300149</td>
<td>NEK4</td>
<td>2.94E-04</td>
<td>DLPFC sQTLs</td>
<td>8.46E-10</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs6499632</td>
<td>RPGRIP1L</td>
<td>3.15E-04</td>
<td>DLPFC sQTLs</td>
<td>2.45E-15</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs1076425</td>
<td>NEK4</td>
<td>3.35E-04</td>
<td>DLPFC sQTLs</td>
<td>2.48E-08</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs1362572</td>
<td>RPGRIP1L</td>
<td>3.58E-04</td>
<td>DLPFC sQTLs</td>
<td>1.54E-26</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rs7193898</td>
<td>RPGRIP1L</td>
<td>4.02E-04</td>
<td>DLPFC sQTLs</td>
<td>4.46E-12</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>rsID</td>
<td>Gene</td>
<td>Minor Allele MAF</td>
<td>Phenotype</td>
<td>p-value</td>
<td>Haplotype MAF</td>
<td>Association</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>------------------</td>
<td>-------------</td>
<td>---------</td>
<td>---------------</td>
<td>-------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>rs9788828</td>
<td>RPGRIP1L</td>
<td>4.88E-04</td>
<td>DLPFC sQTLs</td>
<td>6.51E-05</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs8047089</td>
<td>RPGRIP1L</td>
<td>5.57E-04</td>
<td>DLPFC sQTLs</td>
<td>1.14E-39</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10780035</td>
<td>NEK4</td>
<td>6.03E-04</td>
<td>DLPFC sQTLs</td>
<td>2.52E-09</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11177</td>
<td>NEK4</td>
<td>6.32E-04</td>
<td>DLPFC sQTLs</td>
<td>2.52E-09</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3760008</td>
<td>RPGRIP1L</td>
<td>6.80E-04</td>
<td>DLPFC sQTLs</td>
<td>3.04E-39</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6976</td>
<td>NEK4</td>
<td>7.09E-04</td>
<td>DLPFC sQTLs</td>
<td>2.52E-09</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1946155</td>
<td>RPGRIP1L</td>
<td>7.57E-04</td>
<td>DLPFC sQTLs</td>
<td>1.14E-39</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3774354</td>
<td>NEK4</td>
<td>7.81E-04</td>
<td>DLPFC sQTLs</td>
<td>1.15E-09</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2071042</td>
<td>NEK4</td>
<td>8.96E-04</td>
<td>DLPFC sQTLs</td>
<td>6.43E-04</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10783299</td>
<td>PRKAG1</td>
<td>1.56E-06</td>
<td>DLPFC sQTLs</td>
<td>2.71E-06</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2276817</td>
<td>NEK4</td>
<td>3.21E-06</td>
<td>DLPFC sQTLs</td>
<td>1.73E-03</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2293445</td>
<td>PRKAG1</td>
<td>3.96E-06</td>
<td>DLPFC sQTLs</td>
<td>1.79E-06</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6445539</td>
<td>NEK4</td>
<td>4.22E-06</td>
<td>DLPFC sQTLs</td>
<td>1.64E-03</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11168839</td>
<td>PRKAG1</td>
<td>4.22E-06</td>
<td>DLPFC sQTLs</td>
<td>1.45E-06</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6445538</td>
<td>NEK4</td>
<td>4.72E-06</td>
<td>DLPFC sQTLs</td>
<td>1.64E-03</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2241726</td>
<td>PRKAG1</td>
<td>4.83E-06</td>
<td>DLPFC sQTLs</td>
<td>1.79E-06</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2022265</td>
<td>SNAP91</td>
<td>4.35E-05</td>
<td>DLPFC sQTLs</td>
<td>3.83E-19</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7741310</td>
<td>SNAP91</td>
<td>1.07E-04</td>
<td>DLPFC sQTLs</td>
<td>4.50E-10</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs217291</td>
<td>SNAP91</td>
<td>1.08E-04</td>
<td>DLPFC sQTLs</td>
<td>5.45E-23</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1171114</td>
<td>SNAP91</td>
<td>1.24E-04</td>
<td>DLPFC sQTLs</td>
<td>9.77E-10</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs217308</td>
<td>SNAP91</td>
<td>1.33E-04</td>
<td>DLPFC sQTLs</td>
<td>5.45E-23</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10875912</td>
<td>PRKAG1</td>
<td>1.37E-04</td>
<td>DLPFC sQTLs</td>
<td>5.98E-07</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs217323</td>
<td>SNAP91</td>
<td>1.53E-04</td>
<td>DLPFC sQTLs</td>
<td>5.45E-23</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12118735</td>
<td>KIF21B</td>
<td>1.82E-04</td>
<td>PTV</td>
<td>8.00E-03</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1546977</td>
<td>SNAP91</td>
<td>2.05E-04</td>
<td>DLPFC sQTLs</td>
<td>2.49E-22</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs217328</td>
<td>SNAP91</td>
<td>2.15E-04</td>
<td>DLPFC sQTLs</td>
<td>2.49E-22</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1158059</td>
<td>SNAP91</td>
<td>4.07E-04</td>
<td>DLPFC sQTLs</td>
<td>5.90E-14</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs17053453</td>
<td>CACNA1D</td>
<td>8.33E-04</td>
<td>SPIDEX</td>
<td>-2.41</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7195043</td>
<td>DEF8</td>
<td>2.41E-06</td>
<td>SPIDEX</td>
<td>-2.086</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs115749012</td>
<td>FLOT1</td>
<td>1.05E-05</td>
<td>DLPFC sQTLs</td>
<td>2.37E-22</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs115184190</td>
<td>FLOT1</td>
<td>1.06E-05</td>
<td>DLPFC sQTLs</td>
<td>2.37E-22</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs116387714</td>
<td>FLOT1</td>
<td>1.15E-05</td>
<td>DLPFC sQTLs</td>
<td>2.37E-22</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs115246198</td>
<td>FLOT1</td>
<td>1.35E-05</td>
<td>DLPFC sQTLs</td>
<td>5.90E-20</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs115007114</td>
<td>FLOT1</td>
<td>1.43E-05</td>
<td>DLPFC sQTLs</td>
<td>2.37E-22</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNP</td>
<td>Gene</td>
<td>p-value</td>
<td>Phenotype</td>
<td>Log10 p-value</td>
<td>Association</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>----------</td>
<td>-----------</td>
<td>---------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs115799332</td>
<td>FLOT1</td>
<td>1.50E-05</td>
<td>DLPFC sQTLs</td>
<td>5.90E-20</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs114480306</td>
<td>FLOT1</td>
<td>1.50E-05</td>
<td>DLPFC sQTLs</td>
<td>1.04E-14</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs116733291</td>
<td>FLOT1</td>
<td>1.58E-05</td>
<td>DLPFC sQTLs</td>
<td>5.90E-20</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs116265619</td>
<td>FLOT1</td>
<td>2.34E-05</td>
<td>DLPFC sQTLs</td>
<td>2.37E-22</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs8059821</td>
<td>CDK10</td>
<td>3.17E-05</td>
<td>DLPFC sQTLs</td>
<td>1.55E-04</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2282888</td>
<td>SP4</td>
<td>7.67E-05</td>
<td>DLPFC sQTLs</td>
<td>6.00E-03</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs463938</td>
<td>DPEP1</td>
<td>8.00E-05</td>
<td>SPIDEX</td>
<td>-2.053</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs115973930</td>
<td>FLOT1</td>
<td>8.02E-05</td>
<td>DLPFC sQTLs</td>
<td>1.35E-06</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs164749</td>
<td>CDK10</td>
<td>8.09E-05</td>
<td>DLPFC sQTLs</td>
<td>5.89E-06</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs154657</td>
<td>CDK10</td>
<td>8.26E-05</td>
<td>DLPFC sQTLs</td>
<td>2.71E-06</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs115559348</td>
<td>FLOT1</td>
<td>8.32E-05</td>
<td>DLPFC sQTLs</td>
<td>1.31E-30</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs114374585</td>
<td>FLOT1</td>
<td>8.62E-05</td>
<td>DLPFC sQTLs</td>
<td>1.35E-06</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs114259349</td>
<td>FLOT1</td>
<td>8.71E-05</td>
<td>DLPFC sQTLs</td>
<td>6.14E-07</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs8047581</td>
<td>ZNF276</td>
<td>8.96E-05</td>
<td>DLPFC sQTLs</td>
<td>3.40E-04</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs115614113</td>
<td>FLOT1</td>
<td>8.98E-05</td>
<td>DLPFC sQTLs</td>
<td>1.35E-06</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs164746</td>
<td>CDK10</td>
<td>1.09E-04</td>
<td>DLPFC sQTLs</td>
<td>2.71E-06</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11078935</td>
<td>MED24</td>
<td>1.22E-04</td>
<td>PTV</td>
<td>6.00E-03</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs460879</td>
<td>CDK10</td>
<td>1.24E-04</td>
<td>DLPFC sQTLs</td>
<td>7.40E-06</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6860</td>
<td>CDK10</td>
<td>1.31E-04</td>
<td>DLPFC sQTLs</td>
<td>1.99E-06</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2301249</td>
<td>MPI</td>
<td>1.33E-04</td>
<td>DLPFC sQTLs</td>
<td>7.32E-04</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7085</td>
<td>MPI</td>
<td>1.44E-04</td>
<td>DLPFC sQTLs</td>
<td>5.37E-04</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs114202113</td>
<td>FLOT1</td>
<td>1.47E-04</td>
<td>DLPFC sQTLs</td>
<td>1.07E-14</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1378938</td>
<td>MPI</td>
<td>1.52E-04</td>
<td>DLPFC sQTLs</td>
<td>8.50E-04</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7162232</td>
<td>MPI</td>
<td>1.65E-04</td>
<td>DLPFC sQTLs</td>
<td>7.66E-04</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs116575337</td>
<td>FLOT1</td>
<td>1.68E-04</td>
<td>DLPFC sQTLs</td>
<td>1.91E-12</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4393623</td>
<td>PRPSAP2</td>
<td>1.70E-04</td>
<td>SPIDEX</td>
<td>2.779</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs116523291</td>
<td>FLOT1</td>
<td>1.93E-04</td>
<td>DLPFC sQTLs</td>
<td>1.63E-06</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3828783</td>
<td>LEMD2</td>
<td>1.99E-04</td>
<td>DLPFC sQTLs</td>
<td>1.63E-04</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1205340</td>
<td>ITCH</td>
<td>2.02E-04</td>
<td>DLPFC sQTLs</td>
<td>2.39E-06</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs115211595</td>
<td>FLOT1</td>
<td>2.11E-04</td>
<td>DLPFC sQTLs</td>
<td>1.27E-36</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs115420460</td>
<td>FLOT1</td>
<td>2.40E-04</td>
<td>DLPFC sQTLs</td>
<td>9.40E-38</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs154665</td>
<td>CDK10</td>
<td>2.47E-04</td>
<td>DLPFC sQTLs</td>
<td>4.73E-05</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs116295105</td>
<td>FLOT1</td>
<td>2.61E-04</td>
<td>DLPFC sQTLs</td>
<td>9.40E-38</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs115339255</td>
<td>FLOT1</td>
<td>2.64E-04</td>
<td>DLPFC sQTLs</td>
<td>9.40E-38</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs8177375</td>
<td>FAM118B</td>
<td>2.81E-04</td>
<td>DLPFC sQTLs</td>
<td>6.95E-07</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs17127118</td>
<td>ADD3</td>
<td>2.89E-04</td>
<td>DLPFC sQTLs</td>
<td>1.63E-05</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs116582597</td>
<td>FLOT1</td>
<td>2.92E-04</td>
<td>DLPFC sQTLs</td>
<td>3.30E-22</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6088512</td>
<td>ITCH</td>
<td>3.01E-04</td>
<td>DLPFC sQTLs</td>
<td>1.88E-06</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2282640</td>
<td>MRPL11</td>
<td>3.09E-04</td>
<td>DLPFC sQTLs</td>
<td>1.61E-07</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12942088</td>
<td>SAT2</td>
<td>3.24E-04</td>
<td>DLPFC sQTLs</td>
<td>5.83E-30</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6059893</td>
<td>ITCH</td>
<td>3.33E-04</td>
<td>DLPFC sQTLs</td>
<td>1.80E-06</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1264308</td>
<td>GTF2H4</td>
<td>3.35E-04</td>
<td>SPIDEX</td>
<td>-2.535</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs886420</td>
<td>GTF2H4</td>
<td>3.42E-04</td>
<td>SPIDEX</td>
<td>2.931</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12924572</td>
<td>CDK10</td>
<td>5.13E-04</td>
<td>DLPFC sQTLs</td>
<td>2.50E-05</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11220432</td>
<td>FAM118B</td>
<td>5.42E-04</td>
<td>DLPFC sQTLs</td>
<td>6.83E-07</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1049666</td>
<td>FAM118B</td>
<td>5.61E-04</td>
<td>DLPFC sQTLs</td>
<td>6.83E-07</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1049665</td>
<td>FAM118B</td>
<td>5.81E-04</td>
<td>DLPFC sQTLs</td>
<td>6.83E-07</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.2 Genes studied for alternative splicing

3.2.1 APOPT1

Apoptogenic 1, mitochondrial (APOPT1) was identified by Yasuda et al. (2006) as an apoptogenic protein that localizes to the mitochondria when it was cloned from Apo-E deficient mice smooth muscle cells in culture. By analyzing the amino acid sequence of APOPT1, it was determined to be a soluble protein with no transmembrane domains and a mitochondrial targeting sequence at the N terminus (Yasuda et al., 2006). When APOPT1 expression vectors were transfected into wild-type mouse smooth muscle cells, the cells died within 24 hours and displayed features of apoptosis. Caspase-dependent apoptosis occurs through mitochondrial outer membrane permeabilization and release of cytochrome c to the cytoplasm (Green & Kroemer, 2004). Mitochondrial outer membrane permeabilization can occur through two main pathways: the Bax/Bak-related pathway or the Cyp-D pathway. Yasuda et al. (2006) performed experiments blocking either the Bax/Bak-related apoptosis pathway with Bcl-2 or the Cyp-D apoptosis pathway with CyA and identified that the apoptosis induced by APOPT1 was occurring through only the Cyp-D pathway. Further studies on the inhibition of the APOPT1 pathway determined that the apoptosis-preventing PI3K signaling pathway, which involves the activation of Akt and further inactivation of pro-apoptotic proteins and activation of anti-apoptotic proteins, can prevent apoptosis induced by APOPT1 (Sun et al., 2008). APOPT1 was also identified to interact with Bcl2-associated athanogene 3 (BAG3),
which is also involved in regulating apoptosis and the cellular response to stress (Y. Chen et al., 2013). Thus, many studies have implicated APOPT1 as an apoptogenic protein.

However, Melchionda et al. (2014) identified a different functional role for *APOPT1*. APOPT1 mutations were identified in patients with cavitating leukoencephalopathy and cytochrome c oxidase deficiency, and functional studies were performed in control fibroblasts and fibroblasts from one of these patients with homozygous *APOPT1* mutations leading to a truncated protein. Under normal conditions, APOPT1 was found to be degraded in cell culture as a protein by the proteasome system, therefore these cells were placed under two other conditions: an apoptosis inducing condition and an oxidative stress condition. When the apoptosis inducer was used on these fibroblasts, still no APOPT1 protein was detected in both the control and patient cells. However, under oxidative stress conditions, levels of the APOPT1 protein were detected. Examining the effect of oxidative stress on fibroblasts with mutant *APOPT1* compared to controls demonstrated increased levels of reactive oxygen species (ROS) after oxidative stress only in the mutant *APOPT1* fibroblasts, suggesting that full-length functional *APOPT1* is necessary to maintain low levels of ROS (Melchionda et al., 2014).

### 3.2.1.1 *APOPT1* SNVs Associated with Alternative Splicing

Three SNVs in *APOPT1* have been identified by the GTEx PTV dataset to be associated with alternative splicing and lead to an exon-skipping event and a truncated APOPT1 protein (Table 4).
Table 4. SNVs associated with alternative splicing of *APOPT1*.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Location</th>
<th>EUR MAF</th>
<th>SCZ p value</th>
<th>Splicing p value</th>
<th>Alternative Splicing Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4906337</td>
<td>NM_001302652.1:c.361-30C&gt;A</td>
<td>A = 0.2744</td>
<td>5.27E-13</td>
<td>3.1488E-06</td>
<td>Exon 3 skipping of <em>APOPT1</em> resulting in a truncated protein (p.Glu121Valfs*6)</td>
</tr>
<tr>
<td>rs10431750</td>
<td>NM_001302652.1:c.361-379T&gt;C</td>
<td>C = 0.2744</td>
<td>5.69E-13</td>
<td>2.0799E-05</td>
<td></td>
</tr>
<tr>
<td>rs7140568</td>
<td>NM_001302652.1:c.424+867C&gt;T</td>
<td>T = 0.2744</td>
<td>5.74E-13</td>
<td>2.0799E-05</td>
<td></td>
</tr>
</tbody>
</table>

SNV = single nucleotide polymorphism, EUR MAF = minor allele frequency for European Caucasian population, SCZ = schizophrenia

Rs7140568, rs10431750 and rs4906337 were inputted into RBPMAP and SpliceAid to determine the mechanism by which these SNVs are likely causing exon 3 skipping.

RBPMAP and SpliceAid both did not yield significant results for rs7140568.

For rs10431750, SpliceAid identified hnRNP C 1/2, ELAV1, Sam68 and SLM-2 to only bind to the T allele and TIA-1 to only bind to the C allele. RBPMAP also predicted hnRNP C and hnRNP CL1 to be more likely to bind to the T allele compared to the C allele. Heterogeneous nuclear ribonucleoproteins (hnRNPs) bind to intronic splicing silencers and promote exon skipping, therefore the T allele of rs10431750 likely contributes to splicing out of exon 3 through binding to hnRNPs.

For rs4906337, SpliceAid identified hnRNP H 1/2 and SFRS7 to only bind to the A allele. RBPMAP did not yield significant results for rs4906337. Splicing factor arginine/serine-rich 7 is a member of the serine/arginine rich family of proteins that promote exon inclusion. Interestingly, Crawford and Patton (2006) showed that the SFRS7 and hnRNP H proteins competed for the same splicing regulatory elements to promote or inhibit splicing of exon 2 of a-Tropomyosin (a-TM). Overexpression of the hnRNP H proteins blocked SFRS7 binding and led to decreased exon 2 splicing, while knockdown of the hnRNP H proteins allowed SFRS7 binding and increased exon 2 splicing (Crawford & Patton, 2006). hnRNP H2 and SFRS7 are both expressed in the...
brain however their expression in the brain may differ throughout development and contribute to \textit{APOPT1} exon 3 splicing in or out depending on when each of these proteins are more highly expressed.

3.2.1.2 \textbf{Identification of Two APOPT1 Transcript Isoforms}

92 brain tissue samples were genotyped for the three \textit{APOPT1} SNVs predicted to alter splicing. For rs4906337 and rs7140568 respectively, 50 individuals were homozygous for the C and C alleles, 34 individuals were heterozygous and 8 individuals were homozygous for the A and T alleles. For rs10431750, 48 individuals were homozygous for the T allele, 36 individuals were heterozygous and 8 individuals were homozygous for the C allele.

cDNA from these brain tissue samples was amplified using primers in exons 2 and 4 of \textit{APOPT1} (forward primer: 5’ GCCATGATTGGATAGGACCC 3’, reverse primer: 5’ GAAGTCCGCCATTTCTTCTG 3’). Two \textit{APOPT1} transcripts were identified at varying levels between individuals: a 272bp fragment corresponding to the full \textit{APOPT1} transcript including exon 3 and a 208bp fragment corresponding to a smaller \textit{APOPT1} transcript with exon 3 spliced out (Figure 8). Both fragments were gel extracted from a 2.5% agarose Ethidium Bromide (EtBr) post-stained gel, gel purified and sequenced, and both the full length and exon 3 skipped \textit{APOPT1} sequences were identified.
3.2.1.3 Quantification of APOPT1 Isoforms and Association to rs4906337, rs10431750 and rs7140568

The APOPT1 isoforms with exon 3 spliced in or out were quantified using probe-based assays run as ddPCRs (forward primer: 5’ ACACAAGAATGGAATCAACAG 3’, reverse primer: 5’ TGTAGAAGTCCGCCATTTTC 3’, probe for full length: 5’ TGATTCACTTCTCAGGCCAGGC 3’, probe for exon 3 skipping: 5’ TAGTAAGGTCAGAAAGCAACATTGAATGC 3’). The fractional abundance of the number of copies of the exon 3 skipped isoform compared to the total number of copies of both isoforms was identified in the DLPFC and hippocampus of each sample and was run as an ANOVA with each sample’s genotype. The APOPT1 transcript with exon 3 spliced out is significantly decreased in the DLPFC of individuals with the schizophrenia risk alleles A, C and T of rs4906337, rs10431750 and rs7140568, respectively, with a dose allele effect (rs4906337 and rs7140568: $p = 1.26e-15$, $F (2, 84) = 53.35$, rs10431750: $p = 1.71e-15$, $F = (2, 84) 52.66$) and is most significant for rs4906337 and rs7140568 (Figure 9).
Figure 9. Quantification of APOPT1 transcripts in the DLPFC. Splicing ratio of the APOPT1 exon 3 spliced out transcript compared to all APOPT1 transcripts amplified shows a significant decrease with the schizophrenia risk alleles in an allele dose dependent manner in the DLPFC. (Hom. = Homozygous, Maj. = Major Allele, Min. = Minor Allele, Het. = Heterozygous)

The APOPT1 transcript with exon 3 spliced out is also significantly decreased in the hippocampus of individuals with the schizophrenia risk alleles A, C and T of rs4906337, rs10431750 and rs7140568, respectively, with a dose allele effect (rs4906337 and rs7140568: $p = 1.1e-13$, $F(2, 60) = 51.11$, rs10431750: $p = 5.79e-13$, $F(2, 60) = 46.74$) and is most significant for rs4906337 and rs7140568 (Figure 10).
Figure 10. Quantification of \textit{APOPT1} transcripts in the hippocampus. Splicing ratio of the \textit{APOPT1} exon 3 spliced out transcript compared to all \textit{APOPT1} transcripts amplified shows a significant decrease with the schizophrenia risk alleles in an allele dose dependent manner in the hippocampus. (Hom. = Homozygous, Maj. = Major Allele, Min. = Minor Allele, Het. = Heterozygous)

3.2.1.4 Protein Function Prediction of APOPT1 Isoforms
The shorter isoform of \textit{APOPT1} encodes a transcript that skips exon 3 which leads to a frameshift and a premature stop codon. The premature stop codon is predicted to occur 6 amino acids into exon 4 (p.Glu121Valfs*6) and predicts a truncated protein that is 125 amino acids in length, as opposed to 206 amino acids in the full length transcript. The full-length and truncated \textit{APOPT1} protein sequences were input into InterPro, PANTHER, Pfam, PredictProtein and I-TASSER.

The protein families and domains corresponding to the \textit{APOPT1} protein sequences were identified as uncharacterized protein family UPF0671 (InterPro), unnamed family PTHR31107 (PANTHER) and domain of unknown function DUF2315 (Pfam). Limited information exists to describe the function of these uncharacterized families and domains.
PredictProtein results compared between the full-length and truncated APOPT1 proteins identified the loss of 5 potential protein-binding, 3 potential DNA-binding and 1 potential RNA-binding regions, 1 strand and 5 helix structural regions and 15 exposed and 16 buried solvent accessibility regions (Figure 11).

**Figure 11.** PredictProtein structural and binding site results for full-length and truncated APOPT1 protein sequences. The first row of stick figures represent predicted protein-binding (red diamonds), DNA-binding (yellow circles) and RNA-binding (purple circles) sites. The second row of block figures represents predicted strand (blue boxes) and helix (red boxes) structural regions. The third row of block figures represents predicted exposed (blue boxes) and buried (yellow boxes) solvent accessibility regions. A) The full-length APOPT1 protein predicted structural and binding site results identified 17 protein-binding, 6 DNA-binding and 1 RNA-binding sites, 4 strand and 6 helix structural regions and 31 exposed and 33 buried solvent accessibility regions. B) The truncated APOPT1 protein predicted structural and binding site results identified 12 protein-binding, 3 DNA-binding and 0 RNA-binding sites, 3 strand and 1 helix structural regions and 16 exposed and 17 buried solvent accessibility regions (Yachdav et al., 2014).

I-TASSER protein prediction results identified top protein domains and GO terms that were different for the full-length and truncated APOPT1 protein sequences. The top protein domains were categorized as hydrolases and transferases for the full-length protein but were categorized as oxidoreductases and some hydrolases for the truncated protein. GO molecular terms predicted for the full-length protein were ion binding and DNA binding, but for the truncated protein were monoxygenase activity and nucleotide binding. The GO biological term identified for the full-length protein was DNA integration whereas for the truncated protein it was oxidation-reduction process.

Due to the limited number of experimental studies on APOPT1 protein function, the
exact function of the truncated protein resulting from the exon 3 skipped \textit{APOPT1} transcript and the resulting effect at the cellular level would have to be tested experimentally. The APOPT1 protein is induced upon oxidative stress, therefore the effect of both \textit{APOPT1} transcripts and their translation into proteins upon oxidative stress should be tested. This exon 3 skipped transcript may also be targeted for non-sense mediated decay (NMD) or may escape NMD but create a non-functional protein as a mechanism of decreasing the amount of functional protein that is present in the cell. Lewis, Green & Brenner (2002) found that one third of alternatively spliced transcripts create premature stop codons in order to target these transcripts for non-sense mediated decay (NMD), and propose that this is a mechanism of controlling protein expression (Lewis et al., 2003). This process of linking alternative splicing to NMD in order to regulate gene and protein expression is called regulated unproductive splicing and translation (RUST) (Lareau, Brooks, Soergel, Meng, & Brenner, 2007). It is possible that RUST may occur through exon skipping of exon 3 in \textit{APOPT1} in order to regulate the amount of functional APOPT1 protein. As the schizophrenia risk allele is associated with a decrease in the amount of exon 3 skipped transcript and an increase in the full length transcript, this could result in too much functional protein and an imbalance in the amount of apoptosis required for the stressor to which the cell is responding to, or perhaps the formation of functional protein at incorrect times.

3.2.2 \textit{AS3MT}

Arsenic (+3 Oxidation State) Methyltransferase (\textit{AS3MT}) is a key component in the metabolism of arsenic and catalyzes inorganic arsenic into methylated products, which are then further metabolized through reduction and methylation reactions to monomethylarsonic acid (MMA) and dimethylarsinic acid (DMA) that are then excreted in the urine (Figure 12).
Figure 12. Arsenic metabolism pathway. Metabolism of inorganic arsenic (iAs\(^{V}\)) to dimethylarsinic acid (DMA\(^{III}\)) through a series of reduction and methylation reactions carried out by glutathione reductase (GSH) and arsenic (+3 oxidation state) methyltransferase (AS3MT), respectively.

DMA, the most methylated form of arsenic in the human body, is associated with decreased reactivity and increased urinary arsenic excretion, which suggests efficient methylation of arsenic is a key component of arsenic detoxification. The efficiency of arsenic methylation and metabolism can be measured based on the amount and ratio of excreted arsenic products, with a higher amount of iAs and MMA being associated with less efficient arsenic methylation and metabolism while a higher amount of DMA is associated with more efficient arsenic methylation and metabolism. Differences in arsenic metabolism and susceptibility to increased iAs and MMA buildup have been associated with SNVs in AS3MT (K. Engstrom et al., 2011; K. S. Engstrom et al., 2013; K. S. Engstrom et al., 2015).

As mentioned previously, a novel exon 2 and 3 skipped AS3MT transcript was recently reported that is associated with schizophrenia and bipolar disorder, most highly expressed in the brain and specific to humans (Li et al., 2016). However the mechanism of exon skipping was not identified.

3.2.2.1 AS3MT SNV Predicted to Alter Splicing
The major allele, T, of one SNV in AS3MT was found to be significantly associated with schizophrenia and was predicted by the SPIDEX dataset to cause exon 2 skipping (Table 5).
Table 5. SNV predicted to alter splicing of AS3MT.

| SNV       | Location                | EUR MAF | SCZ  
p value | Splicing  
Z score | Alternative Splicing Event |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3740400</td>
<td>NM_020682.3: c.2-97T&gt;G</td>
<td>G = 0.3698</td>
<td>6.00E-17</td>
<td>2.043</td>
<td>Exon 2 skipping of AS3MT</td>
</tr>
</tbody>
</table>

SNV = single nucleotide polymorphism, EUR MAF = minor allele frequency for European Caucasian population, SCZ = schizophrenia.

Rs3740400 was inputted into RBPMaP and SpliceAid to determine the mechanism by which this SNV is likely causing exon 2 skipping.

RBPMaP did not yield any predictions for rs3740400. SpliceAid identified one RNA binding protein that is predicted to only bind to the minor allele, G, of rs3740400: SF2/ASF, also known as serine/arginine-rich splicing factor 1 (SRSF1) (Figure 13).

![Figure 13. rs3740400 SpliceAid results.](image)

A) SF2/ASF does not bind to major allele, T. B) SF2/ASF binds to minor allele, G (Piva et al., 2012).

SRSF1 is a member of the serine/arginine-rich family of splicing factors. SRSF1 promotes recruitment of the U1 and U2 snRNPs, which are both key components of the spliceosome and are necessary for proper exon inclusion. Based on these predictions, the likely mechanism by which the major allele, T, of rs3740400 may be contributing to the splicing out of exon 2 in AS3MT is through decreased binding of SRSF1 which in turn
would lead to a decrease in U1 and U2 snRNP binding and therefore lack of correct splicing of this exon.

Ming Li et al. (2016) identified a novel alternatively spliced AS3MT transcript associated with schizophrenia that has both exon 2 and 3 skipped. Although rs3740400, which is in intron 1 of AS3MT, is only predicted to cause exon 2 skipping by the SPIDEX dataset, it is possible that the decreased binding of SRSF1 may cause exon 2 and 3 skipping. A similar situation has been identified by Guo et al. (2013) in the IRF-3 gene, whereby a SNV in a SRSF1 binding site in intron 1 of this gene caused decreased binding of SRSF1 which led to skipping of exons 2 and 3 in IRF-3. A minigene assay determined that the single base pair change to the SRSF1 binding sequence was enough to prevent SRSF1 binding and therefore decreased exon 2 and 3 skipping (Guo et al., 2013). Therefore it is possible that disrupting an SRSF1 binding site in intron 1 of AS3MT could not only cause skipping of exon 2 but also skipping of exons 2 and 3, thus making rs3740400 the likely causal variant leading to the newly identified, schizophrenia and bipolar disorder-associated AS3MT splice variant.

3.2.2.2 Identification of Two AS3MT Transcript Isoforms
92 brain tissue samples were genotyped for the AS3MT SNV predicted to alter splicing. For rs3740400, 31 individuals were homozygous for the T allele, 48 individuals were heterozygous and 13 individuals were homozygous for the G allele.

cDNA from these brain tissue samples was amplified using primers in exons 1 and 4 of AS3MT (forward primer: 5’ AGCCCGCCGTCTGAGTC 3’, reverse primer: 5’ CTACCCAGATCCAAAATCCAGCAGT 3’). The full length AS3MT transcript corresponding to a 341 or 305bp fragment and the exon 2 and 3 skipped AS3MT transcript corresponding to a 172 or 136bp fragment were identified at varying levels between individuals (Figure 14). The altered size of the full length and exon 2 and 3 skipped transcripts is due to a 36bp VNTR that exists in exon 1 of AS3MT. All fragments were gel extracted from a 2.5% agarose EtBr post-stained gel, gel purified and
sequenced, and the full length and exon 2 and 3 skipped AS3MT sequences were identified.

Figure 14. Identification of two AS3MT transcript isoforms through cDNA PCR amplification. cDNA PCR amplification using forward and reverse primers in exons 1 and 4 of AS3MT, respectively, amplified a 341bp fragment corresponding to the full length AS3MT transcript and a 172bp fragment corresponding to a smaller AS3MT transcript isoform excluding exons 2 and 3.

3.2.2.3 Quantification of AS3MT Isoforms and Association to rs3740400
The AS3MT isoforms with exons 2 and 3 spliced in or out were quantified using probe-based assays run as ddPCRs (forward primer: 5’ GCCGAGGAGACAGTGAGT 3’, reverse primer: 5’ TAGATGCTCAGGGATCACCA 3’, probe for full length: 5’ TGACGCTGAGATACAAGAGGACGTGC 3’, probe for exon 2 and 3 skipping: 5’ TCGCAGGCGAGGACGACAATATTATGGC 3’). The fractional abundance of the number of copies of the exon 2 and 3 skipped isoform compared to the total number of copies of both isoforms was identified in the DLPFC and hippocampus of each sample and was run as an ANOVA with each sample’s genotype. The AS3MT transcript with exons 2 and 3 spliced out is significantly increased in the DLPFC of individuals with the
schizophrenia risk allele, T, of rs3740400 with a dose allele effect ($p = 2.85e-06$, $F(2, 84) = 14.92$) (Figure 15).

**Figure 15. Quantification of AS3MT transcripts in the DLPFC.** Splicing ratio of the AS3MT exons 2 and 3 spliced out transcript compared to all AS3MT transcripts amplified shows a significantly increased with the schizophrenia risk allele in an allele dose dependent manner in the DLPFC.

The AS3MT transcript with exons 2 and 3 spliced out is also significantly increased in the hippocampus of individuals with the schizophrenia risk allele, T, of rs3740400 with a dose allele effect ($p = 8.81e-06$, $F(2, 60) = 14.22$) (Figure 16).
Figure 16. Quantification of AS3MT transcripts in the hippocampus. Splicing ratio of the AS3MT exons 2 and 3 spliced out transcript compared to all AS3MT transcripts amplified shows a significantly increase with the schizophrenia risk allele in an allele dose dependent manner in the hippocampus.

3.2.2.4 Protein Function Prediction of AS3MT Isoforms
The full length AS3MT start codon exists at the exon 1-2 junction. The smaller isoform of AS3MT encodes a transcript that skips exons 2 and 3, therefore disrupting the start codon at the exon 1-2 junction, leading to the predicted use of an alternative start codon 103 amino acids downstream of the constitutive start site. The full length and downstream start codon protein sequences of AS3MT were input into InterPro, PANTHER and Pfam.

The methyltransferase domain was identified for both AS3MT protein sequences: IPR025714 (InterPro), PTHR43675:SF2 (PANTHER) and PF13847 (Pfam). However, the methyltransferase domain of the full length AS3MT protein spans amino acids 71-215 and is truncated in the smaller AS3MT isoform, spanning amino acids 5-113 (Figure 17).
Figure 17. **AS3MT Pfam protein predictions.** A) Pfam predicted methyltransferase domain (in green) on full length AS3MT protein sequence. B) Pfam predicted truncated methyltransferase domain (in green) on shortened AS3MT protein sequence.

The shortened AS3MT protein sequence coding for a predicted truncated methyltransferase domain likely would result in inefficient methylation and metabolism of arsenic. Interestingly, rs3740400 has been associated with decreased arsenic metabolism efficiency (K. S. Engstrom et al., 2015). This shortened AS3MT isoform may also be associated with decreased arsenic metabolism efficiency as they are both associated with rs3740400, and the mechanism by which the decreased arsenic metabolism efficiency occurs may be due to the truncation of the AS3MT methyltransferase domain.

### 3.2.2.5 Hypothesis of inefficient arsenic metabolism contributing to risk of schizophrenia

Arsenic exists naturally in the world in the Earth’s crust, and is found in water, soil and the atmosphere. The major source of arsenic ingestion to humans is through drinking water, but humans can also come into contact with arsenic through vegetation whose irrigation systems contain arsenic, especially rice, pollution and many manufactured products that contain arsenic, such as computer chips, fertilizer, glass manufacturing and mining waste (Ayres, 1992). The central nervous system is a major target of arsenic, with an unknown entrance mechanism through the blood-brain-barrier but an identified accumulation in the brain (Rodriguez, Limon-Pacheco, Carrizales, Mendoza-Trejo, & Giordano, 2010). Chronic exposure to arsenic has been found in rat brains to cause increased reactive oxygen species, mitochondrial dysfunction and changes to the dopamine system, including elevated striatal dopamine, increased striatal dopamine
receptor 1 (DRD1) mRNA, decreased dopamine receptor 2 (DRD2) mRNA in the nucleus accumbens, decreased dopamine and tyrosine hydroxylase in the striatum and cerebral cortex as well as locomotor changes (Chandravanshi, Shukla, Sultana, Pant, & Khanna, 2014; Kim, Seo, Sung, & Kim, 2014; Moreno Avila, Limon-Pacheco, Giordano, & Rodriguez, 2016; Rodriguez et al., 2010).

An association between arsenic and developmental problems in humans has also been identified in many studies. Decreased IQ, perceptual reasoning, working memory, verbal comprehension and increased incidence of intellectual disabilities have been found in children exposed to powdered milk contaminated with arsenic and arsenic-contaminated water all over the world (Dakeishi, Murata, & Grandjean, 2006; Rosado et al., 2007; Wasserman et al., 2014). Hsieh et al. (2017) examined the relationship between arsenic methylation capacity, AS3MT polymorphisms and developmental delay in children from Taiwan and found that certain AS3MT SNVs were associated with decreased arsenic methylation, and in turn both of these factors were associated with increased developmental delay. SNVs in enzymes involved in the reduction of arsenic were not found to be associated with arsenic methylation or developmental delay. These children from Taiwan were not in an area of increased exposure to arsenic and drank water regulated at the World Health Organization standards of <10ug/L, therefore the toxic effects of arsenic do not require extreme concentrations (Hsieh et al., 2017). Other studies have looked at the association of AS3MT SNVs and arsenic methylation capacity, without looking at developmental delay, and have found that multiple AS3MT SNVs are associated with arsenic methylation and metabolism, including rs3740400, where the schizophrenia and alternative splicing-associated major allele, T, was associated with less efficient or decreased arsenic methylation and metabolism (K. S. Engstrom et al., 2015).

Rs3740400, among other AS3MT SNVs, has been associated with schizophrenia, decreased arsenic methylation and metabolism and an alternatively spliced AS3MT isoform that is most highly expressed in the brain. Therefore, it is hypothesized that the AS3MT variants associated with exon 2 and 3 skipping and the truncated AS3MT protein have poorer arsenic methylation which may lead to increased inorganic or toxic arsenic
exposure in these individuals, even at normal arsenic exposure levels. This arsenic exposure would lead to increased ROS, mitochondrial dysfunction and likely even an altered dopamine system. In combination with additional environmental and genetic risk factors, these detrimental mitochondrial and dopaminergic changes, especially if occurring during development, may contribute to the manifestation of schizophrenia or bipolar disorder in at-risk individuals who have the risk allele for rs3740400, or the other AS3MT SNVs associated with schizophrenia and bipolar disorder, as well as other risk variants.

3.2.3 **BNIP3L**
BCL2/Adenovirus E18 19kDa Interacting Protein 3-Like (*BNIP3L*) is a member of the BCL2 family of proteins that contain a single BCL2 homology 3 domain and are involved in the mitochondrial apoptosis and autophagy pathways (J. Zhang & Ney, 2009). The BNIP proteins have been found to be pro-apoptotic by suppressing the anti-apoptotic function of E1B 19kDa or BCL2, and the overexpression of BNIP3 and BNIP3L has been found to induce autophagy (Boyd et al., 1994; Daido et al., 2004; Hamacher-Brady et al., 2007; Sandoval et al., 2008). Of the BNIP proteins, BNIP3L was found to be most highly expressed in the hippocampus and cerebellum in the developing and adult rat brain (Cho, Choi, Park, Sun, & Geum, 2012). Oxygen-glucose deprivation to cultured cerebral cortex rat neurons dramatically increased BNIP3L mRNA expression (Cho et al., 2012). BNIP3L was also found to positively upregulate NF-kB under hypoxic conditions in the U251 human glioblastoma cell line, and this upregulation was lost in BNIP3L-knockout cell line (Lu et al., 2012).

3.2.3.1 **SNVs Associated with Alternative BNIP3L Transcripts**
Four SNVs were found to have a trend for association with bipolar disorder and were identified by the sQTLSeeker dataset to be associated with the use of an alternative first exon of *BNIP3L* (Table 6).
Table 6. SNVs associated with alternative *BNIP3L* transcripts.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Location</th>
<th>EUR MAF</th>
<th>BD p value</th>
<th>Splicing p value</th>
<th>Alternative transcript event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11779570</td>
<td>NC_000008.10: g.26317889G&gt;A</td>
<td>A = 0.1968</td>
<td>9.23E-06</td>
<td>1.01E-04</td>
<td></td>
</tr>
<tr>
<td>rs11786753</td>
<td>NC_000008.10: g.26318240C&gt;A</td>
<td>A = 0.1968</td>
<td>1.01E-05</td>
<td>1.01E-04</td>
<td></td>
</tr>
<tr>
<td>rs60377146</td>
<td>NC_000008.10: g.26316766G&gt;A</td>
<td>A = 0.1899</td>
<td>1.05E-05</td>
<td>1.94E-04</td>
<td></td>
</tr>
<tr>
<td>rs11135938</td>
<td>NC_000008.10: g.26313736A&gt;C</td>
<td>C = 0.1909</td>
<td>1.83E-05</td>
<td>2.62E-04</td>
<td></td>
</tr>
</tbody>
</table>

SNV = single nucleotide polymorphism, EUR MAF = minor allele frequency for European Caucasian population, BD = bipolar disorder.

These variants are all in an intragenic region downstream of *BNIP3L* and are not in close proximity to the alternative first exons, therefore SpliceAid and RBPMap were not used to identify the mechanism by which these variants could be contributing to this transcript change. However, these SNVs were found to be in intragenic DNA methylation regions, which have been found to regulate alternative promoter usage (Maunakea et al., 2010). This suggests that these variants may affect alternative promoter usage and therefore alternative first exon usage.

### 3.2.3.2 Identification of Two *BNIP3L* Transcript Isoforms

92 brain tissue samples were genotyped for the four *BNIP3L* SNVs predicted to alter first exon usage. For rs11779570, rs11786753, rs60377146 and rs11135938 respectively, 50 individuals were homozygous for the G, C, G and A alleles, 35 individuals were heterozygous and 7 individuals were homozygous for the A, A, A and C alleles.

cDNA from these brain tissue samples was amplified using two different forward primers in each of the potential first exons of *BNIP3L* and the same reverse primer in exon 2 of *BNIP3L* (forward primer 1A: 5’ GGACTCGGCTTTGTTGTTG 3’, forward primer 1B:
5’ CTGTTGCCTCCTGGGGTCTT 3’ reverse primer: 5’ GGAGGATGAGGATGGTACGTG 3’). Two BNIP3L transcripts were identified at varying levels between individuals: a 272bp fragment corresponding to the BNIP3L transcript with one first exon named here as exon 1A and a 157bp fragment corresponding to the BNIP3L transcript with another first exon named here as exon 1B (Figure 18). Both fragments were gel extracted from a 2.5% agarose EtBr post-stained gel, gel purified and sequenced, and both BNIP3L transcripts with the alternative first exon sequences were identified.

![Figure 18. Identification of two BNIP3L transcript isoforms through cDNA PCR amplification. cDNA PCR amplification using two forward primers in each of the alternative first exons and one reverse primer in exons 2 of BNIP3L amplified a 272bp fragment corresponding to the BNIP3L transcript using exon 1A and a 157bp fragment corresponding to the BNIP3L transcript using exon 1B.](image)

3.2.3.3 Quantification of BNIP3L Isoforms and Association to rs11779570, rs11786753, rs60377146 and rs11135938

The BNIP3L isoforms with alternative first exons were quantified using probe-based assays run as ddPCRs (forward 1A primer: 5’ ACAATGTCGTCCCACCTAGTC 3’, forward 1B primer: 5’ TTGCCTCCTGGGGTCTT 3’, reverse primer: 5’
GGTAGCTCCACCCAGGAA 3’, probe for exon 1A isoform: 5’
CCCTGCACAACAACAACAAACTGCG 3’, probe for exon 1B isoform: 5’
AGGGGGATCCGCGCTACC 3’). The fractional abundance of the number of copies of
the isoform with exon 1B compared to the total number of copies of both isoforms was
identified in the DLPFC and hippocampus of each sample and was run as an ANOVA
with each sample’s genotype. The BNIP3L transcript using exon 1B as the first exon was
not significantly increased or decreased in the DLPFC of individuals with the bipolar risk
alleles (p = 0.64, F (2, 84) = 0.449) (Figure 19).

![Figure 19. Quantification of BNIP3L transcripts in the DLPFC.](image)

**Figure 19. Quantification of BNIP3L transcripts in the DLPFC.** Splicing ratio of the
BNIP3L transcript using exon 1B compared to all BNIP3L transcripts amplified shows no
difference with the bipolar disorder risk alleles in the DLPFC.

The BNIP3L transcript using exon 1B as the first exon was not significantly increased or
decreased in the hippocampus of individuals with the bipolar risk alleles (p = 0.889, F (2,
60) = 0.118) (Figure 20).
Figure 20. Quantification of \textit{BNIP3L} transcripts in the hippocampus. Splicing ratio of the \textit{BNIP3L} transcript using exon 1B compared to all \textit{BNIP3L} transcripts amplified shows no difference with the bipolar disorder risk alleles in the hippocampus.

3.2.3.4 Protein Function Prediction of \textit{BNIP3L} Isoforms

The \textit{BNIP3L} transcript utilizing exon 1A encodes for the full length \textit{BNIP3L} protein with 219 amino acids. The \textit{BNIP3L} transcript utilizing exon 1B uses a start codon 40 amino acids downstream of the start codon from the full length \textit{BNIP3L} isoform, and therefore encodes a protein that is 40 amino acids shorter than the full length isoform. The full length and shortened \textit{BNIP3L} protein sequences were input into InterPro, PANTHER, Pfam and PredictProtein.

The BNIP3 domain was identified for both BNIP3L protein sequences: IPR010548 (InterPro), PTHR15186:SF9 (PANTHER) and PF06553 (Pfam). However, the BNIP3 domain of the full length \textit{BNIP3L} protein spans amino acids 29-218 and is truncated in the smaller \textit{BNIP3L} isoform, spanning amino acids 7-178.

PredictProtein results compared between the full-length and truncated \textit{BNIP3L} proteins identified the loss of 3 potential protein-binding regions and the creation of a potential
polynucleotide binding region (Figure 21).

**Figure 21. PredictProtein binding site results for full-length and truncated BNIP3L protein sequences.** The row of stick figures represents predicted protein binding (red diamonds) and polynucleotide binding (yellow circles) sites. A) The full-length BNIP3L protein predicted binding site results identified 11 protein-binding sites. B) The truncated BNIP3L protein predicted binding site results identified 8 protein binding and 1 DNA-binding sites (Yachdav et al., 2014).

GO terms also predicted binding as a molecular function of both isoforms but catalytic activity as a molecular function of only the truncated protein.

Many of the BCL2 family of proteins express two alternatively spliced proteins, and the ratio of these different protein isoforms has been found to be very important in determining the effect on apoptosis. For example, Bcl-x has two isoforms: a short, pro-apoptotic Bcl-xS isoform and a long, anti-apoptotic Bcl-xL isoform. Changes in the ratio of these isoforms can lead to increased or decreased apoptosis depending on whether Bcl-xS or Bcl-xL is increased, respectively (Kunisada et al., 2002; Revil et al., 2007). A similar situation has been identified with BNIP3L in a mouse model of cardiac hypertrophy, where a smaller BNIP3L protein was found to form a heterodimer with the full-length BNIP3L protein and prevent its pro-apoptotic function, however this isoform is not listed in the UCSC human genome browser (Yussman et al., 2002).

Although the specific function of these BNIP3L proteins and how their functions may differ is unclear and still in need of further study, based on extensive evidence from other BCL2 proteins that the ratio of isoforms is highly regulated it is likely that a change in
the ratio of the expression of these BNIP3L proteins would impact their effect on apoptosis and cell function.

### 3.2.4 CACNA1D

Calcium Voltage-Gated Channel Subunit Alpha1 D (CACNA1D) encodes a subunit of the voltage-gated long-lasting L-type calcium channels, which are important for many brain functions, including synaptic transmission and dopamine signaling (Berger & Bartsch, 2014). Variants in this gene have been associated with bipolar disorder, schizophrenia and autism spectrum disorder (Guan et al., 2015; Pinggera et al., 2015; Ross et al., 2016). One SNV in CACNA1D was found to have a trend for association with bipolar disorder and was predicted by the SPIDEX dataset to cause exon 25 skipping (Table 7).

#### Table 7. SNV predicted to alter splicing of CACNA1D.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Location</th>
<th>EUR MAF</th>
<th>BD p value</th>
<th>Splicing Z score</th>
<th>Alternative Splicing Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs17053453</td>
<td>NM_001128839.2: c.3114+18A&gt;G</td>
<td>G = 0.0755</td>
<td>8.33E-04</td>
<td>-4.4027</td>
<td>Exon 25 skipping of CACNA1D</td>
</tr>
</tbody>
</table>

SNV = single nucleotide polymorphism, EUR MAF = minor allele frequency for European Caucasian population, BD = bipolar disorder.

92 brain tissue samples were genotyped for the CACNA1D SNV predicted to alter splicing. For rs17053453, 79 individuals were homozygous for the A allele, 13 individuals were heterozygous and no individuals were homozygous for the G allele.

cDNA from these brain tissue samples was amplified using primers in exons 24 and 26 of CACNA1D (forward primer: 5’ TCCAGTGCCATCTCCGTTG 3’, reverse primer: 5’ TCAGGGTTACTTTTGGCTTCA 3’). Only a 239bp fragment corresponding to the full CACNA1D transcript including exon 25 was identified for each genotype but no 131bp fragment corresponding to the smaller CACNA1D transcript with exon 25 spliced out was identified in the DLPFC or hippocampus (Figure 22).
No identification of exon 25 skipped *CACNA1D* transcript. A) Amplification of *CACNA1D* DLPFC cDNA from 12 individuals heterozygous for rs17053453 and 11 individuals homozygous for the A allele only identified a 239bp fragment corresponding to the full *CACNA1D* transcript but not a 131bp fragment corresponding to the *CACNA1D* transcript with exon 25 spliced out. B) Amplification of *CACNA1D* hippocampal cDNA from 11 individuals heterozygous for rs17053453 and 11 individuals homozygous for the A allele only identified a 239bp fragment corresponding to the full *CACNA1D* transcript but not a 131bp fragment corresponding to the *CACNA1D* transcript with exon 25 spliced out.

Therefore, rs17053453 was not identified in this study to be associated with splicing out of exon 25 of *CACNA1D* in the DLPFC or hippocampus.

### 3.2.5  **KIF21B**

Kinesin Family Member 21B (KIF21B) encodes a member of the kinesin superfamily. This gene has been found to be necessary for proper neuronal morphology and synapse transmission, as Kif21b knockout mice show decreased spine density and synaptic
transmission deficits (Muhia et al., 2016). One SNV in KIF21B was found to have a trend for association with bipolar disorder and was found by the PTV dataset to be associated with exon 28 skipping of KIF21B (Table 8).

**Table 8. SNV associated with alternative splicing of KIF21B.**

<table>
<thead>
<tr>
<th>SNV</th>
<th>Location</th>
<th>EUR MAF</th>
<th>BD</th>
<th>Splicing p value</th>
<th>Alternative Splicing Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12118735</td>
<td>NM_001252100.1: c.3843-736G&gt;A</td>
<td>A = 0.2833</td>
<td>1.82E-04</td>
<td>8.00E-3</td>
<td>Exon 28 skipping of KIF21B</td>
</tr>
</tbody>
</table>

SNV = single nucleotide polymorphism, EUR MAF = minor allele frequency for European Caucasian population, BD = bipolar disorder.

92 brain tissue samples were genotyped for the KIF21B SNV associated with alternative splicing. For rs12118735, 55 individuals were homozygous for the G allele, 30 individuals were heterozygous and 7 individuals were homozygous for the A allele.

cDNA from these brain tissue samples was amplified using primers in exons 27 and 29 of KIF21B (forward primer: 5’ ACCGCAATGTCTTCTCTCTC 3’, reverse primer: 5’ AACTCATCTGTGGCATCCAGG 3’). Only a 214bp fragment corresponding to the full KIF21B transcript including exon 28 was identified for each genotype but no 175bp fragment corresponding to the smaller KIF21B transcript with exon 28 spliced out was identified in the DLPFC or hippocampus (Figure 23).
A) Amplification of KIF21B DLPFC cDNA from 7 individuals homozygous for the G allele, 7 heterozygous individuals and 7 individuals homozygous for the A allele of rs12118735 only identified a 214bp fragment corresponding to the full KIF21B transcript but not a 175bp fragment corresponding to the KIF21B transcript with exon 28 spliced out.

B) Amplification of KIF21B hippocampal cDNA from 4 individuals homozygous for the G allele, 4 heterozygous individuals and 4 individuals homozygous for the A allele of rs12118735 only identified a 214bp fragment corresponding to the full KIF21B transcript but not a 175bp fragment corresponding to the KIF21B transcript with exon 28 spliced out.

Therefore, rs12118735 was not identified in this study to be associated with splicing out of exon 28 of KIF21B in the DLPFC or hippocampus.

3.2.6 LRRC48

LRRC48 (Leucine Rich Repeat Containing 48), also known as DRC3 (Dynein Regulatory Complex subunit 3), encodes subunit 3 of the nexin-dynein regulatory complex (N-DRC). The N-DRC plays an important role in controlling flagellar and ciliary motility (Awata et al., 2015). A mutant mouse model of LRRC48 which introduces a missense
mutation and decrease in \textit{LRRC48} transcription was reported to have hydrocephalus and laterality defects (Ha, Lindsay, Timms, & Beier, 2016). Laterality defects have been identified in schizophrenia, including handedness and abnormal hemispheric asymmetry (Brandler & Paracchini, 2014; Ribolsi, Daskalakis, Siracusano, & Koch, 2014). Hydrocephalus involves enlargement of the ventricles, which is also one of the hallmarks of schizophrenia (Olabi et al., 2011).

3.2.6.1 \textit{LRRC48} SNVs Predicted to Alter Exon 7 and 8 Splicing
Two SNVs in \textit{LRRC48} were found to be associated with schizophrenia and were predicted by the SPIDEX dataset to each cause different exon skipping events (Table 9).

<table>
<thead>
<tr>
<th>SNV</th>
<th>Location</th>
<th>EUR MAF</th>
<th>SCZ ( p ) value</th>
<th>Splicing Z Score</th>
<th>Alternative Splicing Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4924832</td>
<td>NM_001130090.1: c.711+7T&gt;C</td>
<td>C = 0.3767</td>
<td>4.89E-08</td>
<td>2.432</td>
<td>Exon 8 skipping of \textit{LRRC48}</td>
</tr>
<tr>
<td>rs4584886</td>
<td>NM_001130090.1: c.571C&gt;T</td>
<td>T = 0.3748</td>
<td>1.60E-07</td>
<td>-2.204</td>
<td>Exon 7 skipping of \textit{LRRC48}</td>
</tr>
</tbody>
</table>

\(\text{SNV} = \text{single nucleotide polymorphism, EUR MAF} = \text{minor allele frequency for European Caucasian population, SCZ} = \text{schizophrenia.}\)

92 brain tissue samples were genotyped for the two \textit{LRRC48} SNVs predicted to alter splicing. For rs4924832, 40 individuals were homozygous for the T allele, 38 individuals were heterozygous and 14 individuals were homozygous for the C allele. For rs4584886, 43 individuals were homozygous for the C allele, 36 individuals were heterozygous and 13 individuals were homozygous for the T allele.

cDNA from these brain tissue samples was amplified using primers in exons 6 and 9 of \textit{LRRC48} (forward primer: 5’ GGACCTGAGCTTGTTCAACA 3’, reverse primer: 5’ GAGTCCTCAGCGTACATGCT 3’). Only a 495bp fragment corresponding to the full \textit{LRRC48} transcript including both exons 7 and 8 was identified for each genotype but no
348bp fragment or 432bp fragment corresponding to the smaller \textit{LRRC48} transcripts with exon 7 or 8 spliced out, respectively, were identified in the DLPFC or hippocampus (Figure 24).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure24.png}
\caption{No identification of exon 7 or 8 skipped \textit{LRRC48} transcripts. A) Amplification of \textit{LRRC48} DLPFC cDNA from 7 individuals homozygous for the T and C alleles, 7 heterozygous individuals and 7 individuals homozygous for the C and T alleles of rs4924832 and rs4584886, respectively, only identified a 495bp fragment corresponding to the full \textit{LRRC48} transcript but not a 348bp or 432bp fragment corresponding to the \textit{LRRC48} transcripts with exon 7 or 8 spliced out, respectively. B) Amplification of \textit{LRRC48} hippocampal cDNA from 4 individuals homozygous for the T and C alleles, 4 heterozygous individuals and 4 individuals homozygous for the C and T alleles of rs4924832 and rs4584886, respectively, only identified a 495bp fragment corresponding to the full \textit{LRRC48} transcript but not a 348bp or 432bp fragment corresponding to the \textit{LRRC48} transcripts with exon 7 or 8 spliced out, respectively.}
\end{figure}

Therefore, rs4924832 and rs4584886 were not identified in this study to be associated with splicing out of exons 7 or 8 of \textit{LRRC48} in the DLPFC or hippocampus.
3.2.6.2  **LRRC48 SNV Predicted to Alter Exon 2 Splicing**

One SNV in *LRRC48* was found to be associated with schizophrenia and was predicted by the SPIDEX dataset to cause exon 2 skipping of *LRRC48* (Table 10).

**Table 10. SNV predicted to alter splicing of LRRC48.**

<table>
<thead>
<tr>
<th>SNV</th>
<th>Location</th>
<th>EUR MAF</th>
<th>SCZ</th>
<th>Splicing Z score</th>
<th>Alternative Splicing Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11870660</td>
<td>NM_001033551.2:c.-1469G&gt;A</td>
<td>A = 0.3956</td>
<td>2.16E-07</td>
<td>1.994</td>
<td>Exon 2 skipping of LRRC48</td>
</tr>
</tbody>
</table>

SNV = single nucleotide polymorphism, EUR MAF = minor allele frequency for European Caucasian population, SCZ = schizophrenia.

Rs11870660 was inputted into RBPMap and SpliceAid to determine the mechanism by which this SNV is likely causing exon 2 skipping.

SpliceAid identified one RNA binding protein that is predicted to only bind to the minor allele, A, of rs11870660: hnRNP I, also known as heterogeneous nuclear ribonucleoprotein I or polypyrimidine tract binding protein 1. Heterogeneous nuclear ribonucleoproteins bind to intronic splicing silencers and promote exon skipping. RBPMap identified SRSF3, SRSF5 and U2AF2 as RNA binding proteins that are predicted to only bind to the major allele, G of rs11870660. Serine and arginine rich splicing factor 3 (SRSF3) and serine and arginine rich splicing factor 5 (SRSF5) are both serine/arginine rich proteins that promote exon inclusion. The U2 small nuclear RNA auxiliary factor 2 (U2AF2) is a non-snRNP that is necessary for the binding of the U2 snRNP to the branch site, thus a variant in the binding site of this protein may cause decreased binding of U2AF2 and thus decreased binding of the U2 snRNP to the branch site. Therefore the A allele of rs11870660 likely to promote exon 2 skipping of *LRRC48* and the G allele of rs11870660 likely to promote exon 2 inclusion of *LRRC48*. 
3.2.6.3 Identification of Three LRRC48 Transcript Isoforms

92 brain tissue samples were genotyped for the LRRC48 SNV associated with alternative splicing. For rs11870660, 34 individuals were homozygous for the G allele, 42 individuals were heterozygous and 16 individuals were homozygous for the A allele.

cDNA from these brain tissue samples was amplified using primers in exons 1 and 4 of LRRC48 (forward primer: 5’ TTCACCCAGGATGTCTCCCT 3’, reverse primer: 5’ ACATCCTTGAAGGAGGATGCC 3’). Three LRRC48 transcripts were identified at varying levels between individuals: a 471bp fragment corresponding to the full LRRC48 transcript including exon 2, a 382bp fragment corresponding to a smaller LRRC48 transcript with exon 2 spliced out, and a 320bp fragment corresponding to a third novel LRRC48 transcript that has not been reported before with exon 2 and part of exon 3 spliced out (Figure 25). All fragments were gel extracted from 2.5% agarose EtBr post-stained gel, gel purified and sequenced, and the full length, exon 2 and exon 2 and partial exon 3 skipped LRRC48 sequences were identified.

Figure 25. Identification of three LRRC48 transcript isoforms through cDNA PCR amplification. cDNA PCR amplification using forward and reverse primers in exons 1 and 4 of LRRC48 amplified a 471bp fragment corresponding to the full length LRRC48 transcript, a 382bp fragment corresponding to a smaller LRRC48 transcript isoform excluding exon 2 and a 320bp fragment corresponding to a novel LRRC48 transcript isoform excluding exon 2 and part of exon 3.
3.2.6.4  Quantification of LRRC48 Isoforms and Association to rs11870660

The LRRC48 isoforms with exon 2 skipping, exon 2 and partial exon 3 skipping and the full length were quantified using probe-based assays run as ddPCRs (forward primer: 5’ GTCTGCTGCGTGGAACC 3’, reverse primer: 5’ TCCCTGATGATAGTGTTGTTGAAG 3’, probe for full length: 5’ AGTTCAAGGATGTTCAAGACCAGCCC 3’, probe for exon 2 skipping: 5’ TCCCAGCGCTTGAGAAGGAAAATTCTG 3’, probe for exon 2 and partial exon 3 skipping: 5’ CGCTTGAGAAGCTCCAACCTCTCT 3’). The fractional abundance of the number of copies of the exon 2 or exon 2 and partial exon 3 skipped isoforms compared to the total number of copies of all isoforms was identified in the DLPFC and hippocampus of each sample and was run as an ANOVA with each sample’s genotype. The LRRC48 transcript with exon 2 or exon 2 and partial exon 3 spliced out was not significantly increased or decreased in the DLPFC of individuals with the schizophrenia risk allele (exon 2 skipping: \( p = 0.0532, F(2, 84) = 3.04 \), exon 2 and partial exon 3 skipping: \( p = 0.823, F(2, 84) = 0.196 \) (Figure 26).
Figure 26. Quantification of *LRRC48* transcripts in the DLPFC. A) Splicing ratio of the *LRRC38* exon 2 spliced out transcript compared to all *LRRC48* transcripts amplified shows no difference with the schizophrenia risk allele in the DLPFC. B) Splicing ratio of the *LRRC38* exon 2 and partial exon 3 spliced out transcript compared to all *LRRC48* transcripts amplified shows no difference with the schizophrenia risk allele in the DLPFC.

The *LRRC48* transcript with exon 2 or exon 2 and partial exon 3 spliced out was not significantly increased or decreased in the hippocampus of individuals with the
schizophrenia risk allele (exon 2 skipping: \( p = 0.461, F(2, 84) = 0.785 \), exon 2 and partial exon 3 skipping: \( p = 0.385, F(2, 84) = 0.971 \)) (Figure 27).

**Figure 27. Quantification of LRRC48 transcripts in the hippocampus.** A) Splicing ratio of the LRRC38 exon 2 spliced out transcript compared to all LRRC48 transcripts amplified shows no difference with the schizophrenia risk allele in the hippocampus. B) Splicing ratio of the LRRC38 exon 2 and partial exon 3 spliced out transcript compared to all LRRC48 transcripts amplified shows no difference with the schizophrenia risk allele in the hippocampus.
3.2.6.5 Protein Function Prediction of LRRC48 Isoforms
The LRRC48 start codon exists in exon 4, meaning the protein sequences of the full length, exon 2 skipped and exon 2 and partial exon 3 skipped LRRC48 isoforms all have identical protein sequences. Therefore InterPro, PANTHER, Pfam, PredictProtein and I-TASSER were not used to predict the protein function of the alternatively spliced isoforms. Instead, the 5'UTR sequences were input into IRESite to predict the presence of internal ribosome entry sites (IRES), which are translation initiation sites. IRESite predicted the presence of an IRES in the 5' sequence of the full length LRRC48 isoform but not the exon 2 or exon 2 and partial 3 skipped isoforms (Mokrejs et al., 2006). Therefore altering the ratio of these splicing isoforms may affect the rate at which the LRRC48 protein is expressed.

3.2.7 NEK4
NEK4 (never in mitosis gene A related kinase 4) is a member of the NEK family of proteins that are involved in multiple aspects of cell cycle control as well as ciliary function (Fry, O'Regan, Sabir, & Bayliss, 2012). NEK4 has been found to play a role in regulating the autophagy process and autophagy induction in MCF-7 human breast carcinoma cells, but acts independently or downstream of mTORC1 (Szylnarowski et al., 2011). Knockdown of NEK4 in cells from a mouse model of lymphoma promoted resistance to taxol, a microtubule stabilizing drug, but promoted sensitization to vincristine, a microtubule destabilizing drug, and also prevented polarization of microtubules upon depletion, implicating a role in microtubule regulation (Doles & Hemann, 2010). In the brain, NEK4 has been found to associate with ciliary rootlets and is involved in ciliary integrity. Coene et al. (2011) found a decrease in ciliated cells after knocking down NEK4, but no difference in mitotic profiles compared to non-NEK4 knock down. Multiple possible functions of NEK4 have been identified in many cell types, however in the brain NEK4 has been found to be involved in ciliary function.
3.2.7.1  **NEK4 SNVs Predicted to Alter Exon 5 Splicing**

Three SNVs were found to be associated with schizophrenia and bipolar disorder and were identified by the DLPFC sQTLs dataset to be associated with exon 5 skipping of *NEK4* (Table 11).

**Table 11. SNVs associated with alternative splicing of NEK4.**

<table>
<thead>
<tr>
<th>SNV</th>
<th>Location</th>
<th>EUR MAF</th>
<th>BD p value</th>
<th>SCZ p value</th>
<th>Splicing p value</th>
<th>Alternative Splicing Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2276817</td>
<td>NM_001166449.1:c.390G&gt;A</td>
<td>A = 0.2584</td>
<td>3.21E-06</td>
<td>4.49E-08</td>
<td>1.73E-03</td>
<td>Exon 5 skipping of NEK4</td>
</tr>
<tr>
<td>rs6445539</td>
<td>NM_001198974.2:c.879C&gt;T</td>
<td>T = 0.2654</td>
<td>4.22E-06</td>
<td>4.58E-08</td>
<td>1.64E-03</td>
<td></td>
</tr>
<tr>
<td>rs6445538</td>
<td>NM_001198974.2:c.879+228A&gt;G</td>
<td>G = 0.2664</td>
<td>4.72E-06</td>
<td>4.10E-08</td>
<td>1.64E-03</td>
<td></td>
</tr>
</tbody>
</table>

SNV = single nucleotide polymorphism, EUR MAF = minor allele frequency for European Caucasian population, SCZ = schizophrenia.

92 brain tissue samples were genotyped for the *NEK4* SNVs predicted to alter splicing. For rs2276817, rs6445539 and rs6445538 respectively, 50 individuals were homozygous for the G, C and A alleles, 34 individuals were heterozygous, and 8 individuals were homozygous for the A, T and G alleles.

cDNA from these brain tissue samples was amplified using primers in exons 4 and 6 of *NEK4* (forward primer: 5’ GAAATGGCCACCTTGAAGCA 3’, reverse primer: 5’ AGCCCTCAGAAGAGAGTTGGA 3’). Only a 355bp fragment corresponding to the full *NEK4* transcript including exon 5 was identified for each genotype but no 200bp fragment corresponding to the smaller *NEK4* transcript with exon 5 spliced out was identified in the DLPFC or hippocampus (Figure 28).
Figure 28. No identification of exon 5 skipped NEK4 transcript. A) Amplification of NEK4 DLPFC cDNA from 7 individuals homozygous for the G, C and A alleles, 7 heterozygous individuals and 7 individuals homozygous for the A, T and G alleles of rs2276817, rs6445539 and rs6445538, respectively, only identified a 355bp fragment corresponding to the full NEK4 transcript but not a 200bp fragment corresponding to the NEK4 transcript with exon 5 spliced out. B) Amplification of NEK4 hippocampal cDNA from 4 individuals homozygous for the G, C and A alleles, 4 heterozygous individuals and 4 individuals homozygous for the A, T and G alleles of rs2276817, rs6445539 and rs6445538, respectively, only identified a 355bp fragment corresponding to the full NEK4 transcript but not a 200bp fragment corresponding to the NEK4 transcript with exon 5 spliced out.

Therefore, rs2276817, rs6445539 and rs6445538 were not identified in this study to be associated with splicing out of exon 5 of NEK4 in the DLPFC or hippocampus.

3.2.7.2 NEK4 SNVs Predicted to Alter Exon 4 Splicing
Twenty-seven SNVs were found to be associated with bipolar disorder and schizophrenia and were identified by the DLPFC sQTLs dataset to be associated with exon 4 skipping of NEK4 (Table 12).
Table 12. SNVs associated with alternative splicing of *NEK4*.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Location</th>
<th>EUR MAF</th>
<th>BD p value</th>
<th>SCZ p value</th>
<th>Splicing p value</th>
<th>Alternative Splicing Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4687657</td>
<td>NM_001166449.1:c.2002C&gt;A</td>
<td>A = 0.3022</td>
<td>1.48E-05</td>
<td>1.57E-06</td>
<td>1.21E-03</td>
<td></td>
</tr>
<tr>
<td>rs1014969</td>
<td>NC_000003.11:g.52808341G&gt;A</td>
<td>A = 0.4235</td>
<td>4.97E-05</td>
<td>6.83E-01</td>
<td>8.00E-09</td>
<td></td>
</tr>
<tr>
<td>rs3755799</td>
<td>NC_000003.11:g.52809193G&gt;A</td>
<td>A = 0.3439</td>
<td>5.03E-05</td>
<td>6.85E-07</td>
<td>9.83E-09</td>
<td></td>
</tr>
<tr>
<td>rs736408</td>
<td>NM_002217.3:c.1383+192C&gt;T</td>
<td>T = 0.3847</td>
<td>9.18E-05</td>
<td>4.58E-08</td>
<td>4.37E-06</td>
<td></td>
</tr>
<tr>
<td>rs12635140</td>
<td>NM_001010983.2:c.-37+575A&gt;G</td>
<td>G = 0.4235</td>
<td>9.64E-05</td>
<td>6.85E-01</td>
<td>4.77E-09</td>
<td></td>
</tr>
<tr>
<td>rs13071584</td>
<td>NM_001193533.1:c.93+183C&gt;T</td>
<td>G = 0.4294</td>
<td>9.67E-05</td>
<td>6.32E-07</td>
<td>4.77E-09</td>
<td></td>
</tr>
<tr>
<td>rs1866268</td>
<td>NM_014366.4:c.-711C&gt;T</td>
<td>A = 0.4304</td>
<td>9.68E-05</td>
<td>1.82E-06</td>
<td>4.77E-09</td>
<td></td>
</tr>
<tr>
<td>rs11716747</td>
<td>NM_001193533.1:c.2167-2972G&gt;A</td>
<td>A = 0.4304</td>
<td>9.70E-05</td>
<td>1.56E-06</td>
<td>4.77E-09</td>
<td></td>
</tr>
<tr>
<td>rs2289247</td>
<td>NM_014366.4:c.1099G&gt;A</td>
<td>A = 0.4304</td>
<td>9.87E-05</td>
<td>1.64E-06</td>
<td>4.77E-09</td>
<td></td>
</tr>
<tr>
<td>rs3755798</td>
<td>NM_001010983.2:c.-1963C&gt;T</td>
<td>T = 0.4294</td>
<td>1.04E-04</td>
<td>2.16E-06</td>
<td>4.77E-09</td>
<td></td>
</tr>
<tr>
<td>rs1042779</td>
<td>NM_001166434.2:c.1358A&gt;G</td>
<td>G = 0.3986</td>
<td>1.23E-04</td>
<td>9.26E-08</td>
<td>4.59E-10</td>
<td></td>
</tr>
<tr>
<td>rs2535629</td>
<td>NM_002217.3:c.789+112G&gt;A</td>
<td>A = 0.3728</td>
<td>1.36E-04</td>
<td>1.43E-08</td>
<td>4.77E-09</td>
<td></td>
</tr>
<tr>
<td>rs2071508</td>
<td>NM_002217.3:c.-1974G&gt;A</td>
<td>A = 0.3907</td>
<td>2.14E-04</td>
<td>3.98E-07</td>
<td>6.37E-09</td>
<td></td>
</tr>
<tr>
<td>rs11130317</td>
<td>NM_001010983.2:c.329+254G&gt;A</td>
<td>A = 0.4205</td>
<td>2.15E-04</td>
<td>1.94E-07</td>
<td>9.15E-10</td>
<td></td>
</tr>
<tr>
<td>rs1108842</td>
<td>NM_014366.4:c.129A&gt;C</td>
<td>A = 0.4791</td>
<td>2.34E-04</td>
<td>2.27E-09</td>
<td>4.66E-07</td>
<td></td>
</tr>
<tr>
<td>rs2230535</td>
<td>NM_001193533.1:c.201A&gt;G</td>
<td>G = 0.4195</td>
<td>2.40E-04</td>
<td>7.78E-08</td>
<td>9.15E-10</td>
<td></td>
</tr>
<tr>
<td>rs1075653</td>
<td>NM_001166434.2:c.2069-5T&gt;C</td>
<td>C = 0.3926</td>
<td>2.62E-04</td>
<td>2.99E-07</td>
<td>2.27E-08</td>
<td></td>
</tr>
<tr>
<td>rs9324</td>
<td>NM_001166434.2:c.2121T&gt;C</td>
<td>C = 0.3926</td>
<td>2.72E-04</td>
<td>3.64E-07</td>
<td>2.27E-08</td>
<td></td>
</tr>
<tr>
<td>rs2300149</td>
<td>NM_001166434.2:c.1693+560C&gt;T</td>
<td>T = 0.4076</td>
<td>2.94E-04</td>
<td>2.80E-07</td>
<td>8.46E-10</td>
<td></td>
</tr>
<tr>
<td>rs1076425</td>
<td>NM_001166434.2:c.2069-71A&gt;G</td>
<td>G = 0.3917</td>
<td>3.35E-04</td>
<td>1.99E-07</td>
<td>2.48E-08</td>
<td></td>
</tr>
</tbody>
</table>
SNV = single nucleotide polymorphism, EUR MAF = minor allele frequency for European Caucasian population, BD = bipolar disorder, SCZ = schizophrenia.

Of the 27 SNVs associated with exon 4 skipping of NEK4, only two were in close proximity to exon 4 of NEK4 and are the most plausible candidates to cause this splicing change. The two SNVs were rs13071584, which is in intron 1 of NEK4, and rs2230535, which is in exon 3 of NEK4. These two SNVs were inputted into RBPMap and SpliceAid to determine the mechanism by which these SNVs are likely to cause exon 4 skipping.

RBPMap did not yield any significant results for rs13071584 or rs2230535.

For rs13071584, SpliceAid identified SRSF2 to bind to both the A and G alleles but SRSF9 only to bind to the G allele. Serine and arginine rich splicing factor 2 (SRSF2) and serine and arginine rich splicing factor 9 (SRSF9) are both likely to contribute to exon inclusion.

For rs2230535, SpliceAid identified ETR-3 to only bind to the G allele but not the A allele. ELAV-Type RNA-binding protein 3 (ETR-3) has previously been identified to cause exon skipping in the Tau protein (Leroy et al., 2006). Therefore the most likely variant to contribute to exon 4 skipping of NEK4 would be the G allele of rs2230535, which binds to ETR-3.
3.2.7.3 Identification of Two *NEK4* Transcript Isoforms

92 brain tissue samples were genotyped for the twenty-seven *NEK4* SNVs associated with alternative splicing (Table 13).

**Table 13. Brain tissue sample genotypes of SNVs associated with alternative splicing of *NEK4*.**

<table>
<thead>
<tr>
<th>SNV</th>
<th>Major Allele</th>
<th>Minor Allele</th>
<th># Individuals homozygous for Major Allele</th>
<th># Heterozygous Individuals</th>
<th># Individuals homozygous for Minor Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4687657</td>
<td>C</td>
<td>A</td>
<td>46</td>
<td>39</td>
<td>7</td>
</tr>
<tr>
<td>rs1014969</td>
<td>G</td>
<td>A</td>
<td>26</td>
<td>46</td>
<td>20</td>
</tr>
<tr>
<td>rs3755799</td>
<td>G</td>
<td>A</td>
<td>33</td>
<td>46</td>
<td>13</td>
</tr>
<tr>
<td>rs736408</td>
<td>C</td>
<td>T</td>
<td>31</td>
<td>43</td>
<td>18</td>
</tr>
<tr>
<td>rs12635140</td>
<td>A</td>
<td>G</td>
<td>25</td>
<td>46</td>
<td>21</td>
</tr>
<tr>
<td>rs13071584</td>
<td>A</td>
<td>G</td>
<td>26</td>
<td>45</td>
<td>21</td>
</tr>
<tr>
<td>rs1866268</td>
<td>C</td>
<td>A</td>
<td>25</td>
<td>46</td>
<td>21</td>
</tr>
<tr>
<td>rs11716747</td>
<td>G</td>
<td>A</td>
<td>25</td>
<td>46</td>
<td>21</td>
</tr>
<tr>
<td>rs2289247</td>
<td>G</td>
<td>A</td>
<td>25</td>
<td>46</td>
<td>21</td>
</tr>
<tr>
<td>rs3755798</td>
<td>C</td>
<td>T</td>
<td>25</td>
<td>46</td>
<td>21</td>
</tr>
<tr>
<td>rs1042779</td>
<td>A</td>
<td>G</td>
<td>29</td>
<td>47</td>
<td>16</td>
</tr>
<tr>
<td>rs2535629</td>
<td>G</td>
<td>A</td>
<td>31</td>
<td>44</td>
<td>17</td>
</tr>
<tr>
<td>rs2071508</td>
<td>G</td>
<td>A</td>
<td>38</td>
<td>40</td>
<td>14</td>
</tr>
<tr>
<td>rs11130317</td>
<td>G</td>
<td>A</td>
<td>31</td>
<td>47</td>
<td>14</td>
</tr>
<tr>
<td>rs1108842</td>
<td>C</td>
<td>A</td>
<td>29</td>
<td>46</td>
<td>17</td>
</tr>
<tr>
<td>rs2230535</td>
<td>A</td>
<td>G</td>
<td>31</td>
<td>47</td>
<td>14</td>
</tr>
<tr>
<td>rs1075653</td>
<td>T</td>
<td>C</td>
<td>34</td>
<td>42</td>
<td>16</td>
</tr>
<tr>
<td>rs9324</td>
<td>T</td>
<td>C</td>
<td>33</td>
<td>42</td>
<td>17</td>
</tr>
<tr>
<td>rs2300149</td>
<td>C</td>
<td>T</td>
<td>36</td>
<td>41</td>
<td>15</td>
</tr>
<tr>
<td>rs1076425</td>
<td>A</td>
<td>G</td>
<td>39</td>
<td>39</td>
<td>14</td>
</tr>
<tr>
<td>rs10780035</td>
<td>G</td>
<td>A</td>
<td>35</td>
<td>43</td>
<td>14</td>
</tr>
<tr>
<td>rs11177</td>
<td>G</td>
<td>A</td>
<td>35</td>
<td>43</td>
<td>14</td>
</tr>
<tr>
<td>rs6976</td>
<td>G</td>
<td>A</td>
<td>35</td>
<td>44</td>
<td>13</td>
</tr>
<tr>
<td>SNP</td>
<td>Reference</td>
<td>Alternative</td>
<td>Count 1</td>
<td>Count 2</td>
<td>Count 3</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td>-------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>rs3774354</td>
<td>G</td>
<td>A</td>
<td>36</td>
<td>44</td>
<td>12</td>
</tr>
<tr>
<td>rs2071042</td>
<td>T</td>
<td>C</td>
<td>25</td>
<td>45</td>
<td>22</td>
</tr>
<tr>
<td>rs6769789</td>
<td>A</td>
<td>G</td>
<td>25</td>
<td>44</td>
<td>23</td>
</tr>
<tr>
<td>rs2535627</td>
<td>A</td>
<td>G</td>
<td>21</td>
<td>42</td>
<td>29</td>
</tr>
</tbody>
</table>

cDNA from the DLPFC brain tissue samples was amplified using primers in exons 3 and 6 of NEK4 (forward primer: 5’ GGACCTAGGAATTGCCCGAG 3’, reverse primer: 5’ AGCCCTCAGAAGAGAGTGGT 3’). Two NEK4 transcripts were identified at varying levels between individuals: a 503bp fragment corresponding to the full NEK4 transcript including exon 4 and a 395bp fragment corresponding to a smaller NEK4 transcript with exon 4 spliced out (Figure 29). Both fragments were gel extracted from 2.5% agarose EtBr post-stained gel, gel purified and sequenced, and both the full length and exon 4 skipped NEK4 sequences were identified.

Figure 29. Identification of two NEK4 transcript isoforms through DLPFC cDNA PCR amplification. cDNA PCR amplification using forward and reverse primers in exons 3 and 6 of NEK4 amplified a 503bp fragment corresponding to the full length NEK4 transcript and a 395bp fragment corresponding to a smaller NEK4 transcript isoform excluding exon 4.
3.2.7.4 Quantification of NEK4 Isoforms and Association to SNVs associated with NEK4 exon 4 skipping

The NEK4 isoforms with exon 4 spliced in or out were quantified using probe-based assays run as ddPCRs (forward primer: 5’ GCTCTGGGCTGTAATCTCTTG 3’, reverse primer: 5’ CATTGGCACACCCTACTACAT 3’, probe for full length: 5’ ACACAGCATCCTAGAGCCCCAAACA 3’, probe for exon 4 skipping: 5’ TGGTGGCAGCTTATAGTTGTAGG 3’). The fractional abundance of the number of copies of the exon 4 skipped isoform compared to the total number of copies of both isoforms was identified in the DLPFC and hippocampus of each sample and was run as an ANOVA with each sample’s genotype. The NEK4 transcript with exon 4 spliced out is significantly increased in the DLPFC of individuals with the schizophrenia and bipolar disorder risk allele, G, of both rs13071584 and rs2230535 with a dose allele effect (rs13071584: \( p = 2.39\times10^{-11}, \ F(2, 84) = 33.19 \); rs2230535: \( p = 2.18\times10^{-13}, \ F(2, 84) = 42.08 \)) and is most significant for rs2230535 (Figure 30).

**Figure 30. Quantification of NEK4 transcripts in the DLPFC.** Splicing ratio of the NEK4 exon 4 spliced out transcript compared to all NEK4 transcripts amplified shows a significant increase with the schizophrenia and bipolar disorder risk alleles in an allele dose dependent manner in the DLPFC.

The NEK4 transcript with exon 4 spliced out is also significantly increased in the hippocampus of individuals with the schizophrenia and bipolar disorder risk allele, G, of
rs2230535 with a dose allele effect (rs2230535: \( p = 0.0274, F (2, 60) = 3.823 \)) (Figure 31).

**Figure 31. Quantification of NEK4 transcripts in the hippocampus.** Splicing ratio of the NEK4 exon 4 spliced out transcript compared to all NEK4 transcripts amplified shows a significant increase with the schizophrenia and bipolar disorder risk allele in an allele dose dependent manner in the hippocampus.

### 3.2.7.5 Protein Function Prediction of NEK4 Isoforms

The full length NEK4 transcript encodes for a protein with 841 amino acids, while the NEK4 transcript skipping exon 4 causes an in-frame deletion of 36 amino acids and results in a protein with 805 amino acids. The full length and exon 4 skipped NEK4 protein sequences were input into InterPro, PANTHER, Pfam and PredictProtein.

The protein kinase domain was identified for both NEK4 protein sequences: IPR000719 (InterPro), PTHR24361:SF452 (PANTHER) and PF00069 (Pfam). However, in the full length NEK4 protein, the protein kinase domain spans amino acids 9-261 and includes all of exon 4. Therefore in the NEK4 protein with exon 4 skipped, 36 amino acids of the protein kinase domain are deleted (Figure 32).
Figure 32. NEK4 Pfam protein predictions. A) Pfam predicted protein kinase domain (in green) on full length NEK4 protein sequence. B) Pfam predicted truncated protein kinase domain (in green) on shortened NEK4 protein sequence.

PredictProtein also identified different potential protein-binding regions for the full length and exon 4 skipped NEK4 protein sequences.

Therefore, the NEK4 protein with exon 4 skipped may have decreased protein kinase activity, however this would have to be experimentally tested in order to determine this protein’s exact function.

3.2.8 NGEF

NGEF (neuronal guanine nucleotide exchange factor) is a guanine exchange factor protein that has been identified to be involved in axon guidance by activating EphA receptors (Shamah et al., 2001). One SNV in NGEF was found to be associated with schizophrenia and was predicted by the SPIDEX dataset to cause exon 2 skipping (Table 14).

Table 14. SNV predicted to alter splicing of NGEF.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Location</th>
<th>EUR MAF</th>
<th>SCZ p value</th>
<th>Splicing Z Score</th>
<th>Alternative Splicing Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2675959</td>
<td>NM_001114090.1:c.108-92T&gt;A</td>
<td>A = 0.2913</td>
<td>1.13E-09</td>
<td>3.415</td>
<td>Exon 2 skipping of NGEF</td>
</tr>
</tbody>
</table>

SNV = single nucleotide polymorphism, EUR MAF = minor allele frequency for European Caucasian population, SCZ = schizophrenia.
92 brain tissue samples were genotyped for the NGEF SNV predicted to alter splicing. For rs2675959, 53 individuals were homozygous for the T allele, 29 individuals were heterozygous, 10 individuals were homozygous for the A allele.

cDNA from these brain tissue samples was amplified using primers in exons 1 and 3 of NGEF (forward primer: 5’ GTGGACCACGACAGTTCCA 3’, reverse primer: 5’ GCTGTAAGGATCTCAAGCACC 3’). Only a 490bp fragment corresponding to the full NGEF transcript including exon 2 was identified for each genotype but no 347bp fragment corresponding to the smaller NGEF transcript with exon 2 spliced out was identified in the DLPFC or hippocampus (Figure 33).
Figure 33. No identification of exon 2 skipped NGEF transcript. A) Amplification of NGEF DLPFC cDNA from 7 individuals homozygous for the T allele, 7 heterozygous individuals and 7 individuals homozygous for the A allele of rs2675959 only identified a 490bp fragment corresponding to the full NGEF transcript but not a 347bp fragment corresponding to the NGEF transcript with exon 2 spliced out. B) Amplification of NGEF hippocampal cDNA from 4 individuals homozygous for the T allele, 4 heterozygous individuals and 4 individuals homozygous for the A allele of rs2675959 only identified a 490bp fragment corresponding to the full NGEF transcript but not a 347bp fragment corresponding to the NGEF transcript with exon 2 spliced out.

Therefore, rs2675959 was not identified in this study to be associated with splicing out of exon 2 of NGEF in the DLPFC or hippocampus.

3.2.9 PRKAG1

*PRKAG1* (protein kinase AMP-activated non-catalytic subunit gamma 1) is a protein that encodes a regulatory subunit of the AMP-activated protein kinase. PRKAG1 has to been
found to interact with the netrin-DSCAM signaling pathway to induce neurite outgrowth (Zhu et al., 2013). Five SNVs were found to be associated with bipolar disorder and were identified by the DLPFC sQTLs dataset to be associated with partial intron 2 retention of PRKAG1 (Table 15).

Table 15. SNVs associated with alternative splicing of PRKAG1.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Location</th>
<th>EUR MAF</th>
<th>BD p value</th>
<th>Splicing p value</th>
<th>Alternative Splicing Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10875912</td>
<td>NM_003482.3: c.16053-286A&gt;G</td>
<td>G = 0.3042</td>
<td>1.37E-04</td>
<td>5.98E-07</td>
<td></td>
</tr>
<tr>
<td>rs2241726</td>
<td>NM_003482.3: c.2826C&gt;T</td>
<td>T = 0.3479</td>
<td>4.83E-06</td>
<td>1.79E-06</td>
<td></td>
</tr>
<tr>
<td>rs11168839</td>
<td>NM_000012.11: g.49457635G&gt;A</td>
<td>A = 0.3489</td>
<td>4.22E-06</td>
<td>1.45E-06</td>
<td>Partial intron 2 retention of PRKAG1</td>
</tr>
<tr>
<td>rs2293445</td>
<td>NM_001206709.1: c.382+41C&gt;T</td>
<td>T = 0.3489</td>
<td>3.96E-06</td>
<td>1.79E-06</td>
<td></td>
</tr>
<tr>
<td>rs10783299</td>
<td>NM_015086.1: c.1982A&gt;G</td>
<td>G = 0.3539</td>
<td>1.56E-06</td>
<td>2.71E-06</td>
<td></td>
</tr>
</tbody>
</table>

SNV = single nucleotide polymorphism, EUR MAF = minor allele frequency for European Caucasian population, BD = bipolar disorder.

92 brain tissue samples were genotyped for the five PRKAG1 SNVs associated with alternative splicing (Table 16).

Table 16. Brain tissue sample genotypes of SNVs associated with alternative splicing of PRKAG1.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Major Allele</th>
<th>Minor Allele</th>
<th># Individuals homozygous for Major Allele</th>
<th># Heterozygous Individuals</th>
<th># Individuals homozygous for Minor Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10875912</td>
<td>A</td>
<td>G</td>
<td>37</td>
<td>41</td>
<td>14</td>
</tr>
<tr>
<td>rs2241726</td>
<td>C</td>
<td>T</td>
<td>36</td>
<td>39</td>
<td>17</td>
</tr>
<tr>
<td>rs11168839</td>
<td>G</td>
<td>A</td>
<td>35</td>
<td>40</td>
<td>17</td>
</tr>
<tr>
<td>rs2293445</td>
<td>C</td>
<td>T</td>
<td>38</td>
<td>38</td>
<td>16</td>
</tr>
<tr>
<td>rs10783299</td>
<td>A</td>
<td>G</td>
<td>34</td>
<td>41</td>
<td>17</td>
</tr>
</tbody>
</table>
cDNA from the brain tissue samples was amplified using primers in exons 2 and 3 of *PRKAG1* (forward primer: 5’ TCTTCAGATAGCTCCCAGC 3’, reverse primer: 5’ AGCTTTGTGGAATCAGGTCA 3’). Only a 121bp fragment corresponding to the *PRKAG1* transcript without partial intron 2 retention was identified for each genotype but no 295bp fragment corresponding to the *PRKAG1* transcript with partial intron 2 retention was identified in the DLPFC or hippocampus (Figure 34).

**Figure 34. No identification of *PRKAG1* transcript with partial intron 2 retention.**  
A) Amplification of *PRKAG1* DLPFC cDNA from 7 individuals homozygous for the A, C, G, C and A alleles, 7 heterozygous individuals and 7 individuals homozygous for the G, T, A, T and G alleles of rs10875912, rs2241726, rs11168839, rs2293445 and rs10783299, respectively, only identified a 121bp fragment corresponding to the *PRKAG1* transcript without partial intron 2 retention but not a 295bp fragment corresponding to the *PRKAG1* transcript with partial intron 2 retention. B) Amplification of *PRKAG1* hippocampal cDNA from 4 individuals homozygous for the A, C, G, C and A alleles, 4 heterozygous individuals and 4 individuals homozygous for the G, T, A, T and G alleles of rs10875912, rs2241726, rs11168839, rs2293445 and rs10783299, respectively, only identified a 121bp fragment corresponding to the *PRKAG1* transcript without partial intron 2 retention but not a 295bp fragment corresponding to the *PRKAG1* transcript with partial intron 2 retention.
Therefore, these five SNVs were not identified in this study to be associated with partial intron 2 retention of PRKAG1 in the DLPFC or hippocampus.

3.2.10 PPP1R16B

PPP1R16B (protein phosphatase 1, regulatory subunit 16B) encodes a regulatory protein for protein phosphatase 1, which has many regulatory and signaling roles involved in neuronal activities, brain signaling cascades, apoptosis and other processes (Shopik et al., 2013). PPP1R16B was implicated in schizophrenia by the most recent schizophrenia GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), and although it has not been implicated in schizophrenia in any other study, it regulates a gene that plays a role in many neuronal processes and is therefore a potential schizophrenia candidate gene. Three SNVs in PPP1R16B were found to be associated with schizophrenia and were identified by the sQTLSeekeR dataset to be associated with the PPP1R16B isoform with exon 7 spliced out (Table 17).

Table 17. SNVs associated with alternative splicing of PPP1R16B.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Location</th>
<th>EUR MAF</th>
<th>SCZ p value</th>
<th>Splicing p value</th>
<th>Alternative Splicing Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4368395</td>
<td>NM_001172735.2: c.250+5635G&gt;A</td>
<td>G = 0.329</td>
<td>3.32E-09</td>
<td>7.2E-05</td>
<td>Exon 7 skipping of PPP1R16B</td>
</tr>
<tr>
<td>rs6093098</td>
<td>NM_001172735.2: c.250+11459T&gt;C</td>
<td>T = 0.327</td>
<td>4.85E-09</td>
<td>7.2E-05</td>
<td></td>
</tr>
<tr>
<td>rs10854216</td>
<td>NM_001172735.2: c.-102+13484A&gt;G</td>
<td>A = 0.3231</td>
<td>3.66E-08</td>
<td>1.24E-04</td>
<td></td>
</tr>
</tbody>
</table>

SNV = single nucleotide polymorphism, EUR MAF = minor allele frequency for European Caucasian population, SCZ = schizophrenia.

92 brain tissue samples were genotyped for the three PPP1R16B SNVs associated with alternative splicing. For rs4368395, 54 individuals were homozygous for the A allele, 30 individuals were heterozygous and 8 individuals were homozygous for the G allele. For rs6093098, 54 individuals were homozygous for the C allele, 31 individuals were heterozygous and 7 individuals were homozygous for the A allele. For rs10854216, 55
individuals were homozygous for the G allele, 28 individuals were heterozygous and 9 individuals were homozygous for the A allele.

cDNA from the brain tissue samples was amplified using primers in exons 6 and 8 of \textit{PPP1R16B} (forward primer: 5’ AAAATCAACGAGATCGGGT 3’, reverse primer: 5’ CATGGATGAGATGCCAATAG 3’). Only a 313bp fragment corresponding to the full \textit{PPP1R16B} transcript including exon 7 was identified for each genotype but no 187bp fragment corresponding to the smaller \textit{PPP1R16B} transcript with exon 7 spliced out was identified in the DLPFC or hippocampus (Figure 35).
Figure 35. No identification of exon 7 skipped PPP1R16B transcript. A) Amplification of PPP1R16B DLPFC cDNA from 7 individuals homozygous for the A, C and G alleles, 7 heterozygous individuals and 7 individuals homozygous for the G, A, and A alleles of rs4368395, rs6093098 and rs10854216, respectively, only identified a 313bp fragment corresponding to the PPP1R16B transcript corresponding to the full length PPP1R16B isoform with exon 7 spliced in but not a 187bp fragment corresponding to the PPP1R16B isoform with exon 7 spliced out. B) Amplification of PPP1R16B hippocampal cDNA from 4 individuals homozygous for the A, C and G alleles, 4 heterozygous individuals and 4 individuals homozygous for the G, A, and A alleles of rs4368395, rs6093098 and rs10854216, respectively, only identified a 313bp fragment corresponding to the PPP1R16B transcript corresponding to the full length PPP1R16B isoform with exon 7 spliced in but not a 187bp fragment corresponding to the PPP1R16B isoform with exon 7 spliced out.

Therefore, rs4368395, rs6093098 and rs10854216 were not identified in this study to be associated with splicing out of exon 7 of PPP1R16B in the DLPFC or hippocampus.
3.2.11 **RPGRIP1L**
Retinitis pigmentosa GTPase regulator interacting protein 1-like (*RPGRIP1L*) encodes for a protein that localizes to the basal bodies and transition zone of primary cilia. Mutations in this gene have been found to cause Joubert Syndrome type 7 and Meckel Syndrome type 5, both ciliopathies, and *RPGRIP1L* knockout mice have a ciliopathy phenotype and die at birth (Delous et al., 2007; Vierkotten, Dildrop, Peters, Wang, & Ruther, 2007). Vierkotten et al. (2007) showed that RPGRIP1L is necessary for multiple processes during embryonic development, including left-right asymmetry development, neural tube formation and cilium-related sonic hedgehog signaling. RPGRIP1L has been found to regulate proteasomal activity in the primary cilia of mouse embryonic fibroblasts, and interacts with Psmd2, a 19S proteasomal subunit component (Gerhardt et al., 2015). RPGRIP1L has also been found to interact with NEK4 at the basal bodies of the cilia and is thought to act as a scaffolding protein to recruit proteins to the primary cilium (Coene et al., 2011). RPGRIP1L has also been found to be in a region associated with obesity (Thorleifsson et al., 2009). Stratigopoulos et al. (2008) found that the expression of RPGRIP1L in the mouse hypothalamus decreased after fasting as well as exposure to cold (Stratigopoulos et al., 2008). These researchers later identified a GWAS obesity risk variant that effects a proximal enhancer leading to decreased RPGRIP1L levels and in turn decreased leptin receptor localization to the cilium in the hypothalamus, and therefore affects leptin signaling and food intake (Stratigopoulos, LeDuc, Cremona, Chung, & Leibel, 2011). RPGRIP1L is a protein with many functions that has been found to recruit multiple proteins to the primary cilia and whose expression has been shown to be regulated by environmental factors and as well genetic variants.

3.2.11.1 **RPGRIP1L SNVs Associated with Alternative Splicing**
Ten SNVs were found to have a trend for association with bipolar disorder and were identified by the DLPFC sQTLs dataset to be associated with exon 20 skipping of *RPGRIP1L* (Table 18). One of these variants was also found to have a trend for association with schizophrenia.
Table 18. SNVs associated with alternative splicing of *RPGRIP1L*.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Location</th>
<th>EUR MAF</th>
<th>BD p value</th>
<th>SCZ p value</th>
<th>Splicing p value</th>
<th>Alternative Splicing Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs8050354</td>
<td>NM_001127897.3: c.3376+3769G&gt;A</td>
<td>A = 0.3231</td>
<td>4.73E-05</td>
<td></td>
<td>9.12E-42</td>
<td></td>
</tr>
<tr>
<td>rs1421091</td>
<td>NM_001127897.3: c.-2066T&gt;G</td>
<td>G = 0.3230</td>
<td>6.63E-05</td>
<td></td>
<td>7.31E-41</td>
<td></td>
</tr>
<tr>
<td>rs6499632</td>
<td>NM_001127897.3: c.*2907A&gt;G</td>
<td>G = 0.4851</td>
<td>3.15E-04</td>
<td></td>
<td>2.45E-15</td>
<td></td>
</tr>
<tr>
<td>rs1362572</td>
<td>NC_000016.10:g.5354008G&gt;A</td>
<td>A = 0.3330</td>
<td>3.58E-04</td>
<td>7.23E-06</td>
<td>1.54E-26</td>
<td>Exon 20 skipping of <em>RPGRIP1L</em></td>
</tr>
<tr>
<td>rs7193898</td>
<td>NC_000016.10:g.5351009G&gt;A</td>
<td>A = 0.2744</td>
<td>4.02E-04</td>
<td></td>
<td>4.46E-12</td>
<td></td>
</tr>
<tr>
<td>rs9788828</td>
<td>NC_000016.10:g.5354307G&gt;T</td>
<td>T = 0.2495</td>
<td>4.88E-04</td>
<td></td>
<td>6.51E-05</td>
<td></td>
</tr>
<tr>
<td>rs8047089</td>
<td>NM_001127897.3: c.3461+1220G&gt;A</td>
<td>A = 0.2992</td>
<td>5.57E-04</td>
<td></td>
<td>1.14E-39</td>
<td></td>
</tr>
<tr>
<td>rs3760008</td>
<td>NM_001127897.3: c.*737T&gt;G</td>
<td>G = 0.2962</td>
<td>6.80E-04</td>
<td></td>
<td>3.04E-39</td>
<td></td>
</tr>
<tr>
<td>rs1946155</td>
<td>NM_001127897.3: c.*1693G&gt;A</td>
<td>A = 0.2962</td>
<td>7.57E-04</td>
<td></td>
<td>1.14E-39</td>
<td></td>
</tr>
<tr>
<td>rs4784321</td>
<td>NM_001127897.3: c.3596-309G&gt;A</td>
<td>A = 0.2326</td>
<td>1.85E-03</td>
<td></td>
<td>5.61E-19</td>
<td></td>
</tr>
</tbody>
</table>

SNV = single nucleotide polymorphism, EUR MAF = minor allele frequency for European Caucasian population, BD = bipolar disorder, SCZ = schizophrenia.

Of the 10 SNVs associated with exon 20 skipping of *RPGRIP1L*, only five were in close proximity to exon 20 of *RPGRIP1L* and are the most plausible candidates to cause this splicing change. The five SNVs were rs8050354, which is in intron 24 of *RPGRIP1L*, rs8047089, which is in intron 25 of *RPGRIP1L*, rs4784321, which is in intron 26 of *RPGRIP1L*, and rs3760008 and rs1946155, which are in exon 27 of *RPGRIP1L*. These five SNVs were inputted into RBPMap and SpliceAid to determine the mechanism by which these SNVs are likely to cause exon 20 skipping.

RBPMap did not yield any significant results for any of the five SNVs and SpliceAid did not yield any significant results for rs8050354 or rs4784321.
For rs8047089, SpliceAid identified MBNL1 to only bind to the G allele but not the A allele. Muscleblind like splicing regulator 1 (MBNL1) has been found to bind to and compete for the same sequence as the U2AF65 splicing factor, which recruits the U2 snRNP necessary for proper splicing, and therefore MBNL1 promotes exon skipping (Warf, Diegel, von Hippel, & Berglund, 2009).

For rs3760008, SpliceAid identified Sam68 to only bind to the T allele but not the G allele. Sam68 has been found to cause exon 7 skipping in the SMN2 gene (Pedrotti et al., 2010).

For rs1946155, SpliceAid identified hnRNP I to only bind to the A allele but not the G allele. Heterogeneous nuclear ribonucleoprotein I (hnRNP I) or polypyrimidine tract binding protein 1 (PTBP1) is a member of the heterogeneous nuclear ribonucleoproteins that bind to intronic splicing silencers and promote exon skipping. However, PTBP1 is only expressed in neural stem and progenitor cells but not mature neurons, and therefore would not contribute to splicing in the adult brain tissues used in this study but may contribute to RPGRIP1L splicing during development in neural stem cells.

3.2.11.2 Identification of Two RPGRIP1L Transcript Isoforms
92 brain tissue samples were genotyped for the ten RPGRIP1L SNVs associated with alternative splicing (Table 19).
Table 19. Brain tissue sample genotypes of SNVs associated with alternative splicing of *RPGRIP1L*.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Major Allele</th>
<th>Minor Allele</th>
<th># Individuals homozygous for Major Allele</th>
<th># Heterozygous Individuals</th>
<th># Individuals homozygous for Minor Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs8050354</td>
<td>G</td>
<td>A</td>
<td>42</td>
<td>37</td>
<td>13</td>
</tr>
<tr>
<td>rs1421091</td>
<td>T</td>
<td>G</td>
<td>40</td>
<td>37</td>
<td>15</td>
</tr>
<tr>
<td>rs6499632</td>
<td>A</td>
<td>G</td>
<td>21</td>
<td>44</td>
<td>27</td>
</tr>
<tr>
<td>rs1362572</td>
<td>G</td>
<td>A</td>
<td>42</td>
<td>42</td>
<td>8</td>
</tr>
<tr>
<td>rs7193898</td>
<td>G</td>
<td>A</td>
<td>44</td>
<td>44</td>
<td>4</td>
</tr>
<tr>
<td>rs9788828</td>
<td>G</td>
<td>T</td>
<td>46</td>
<td>43</td>
<td>3</td>
</tr>
<tr>
<td>rs8047089</td>
<td>G</td>
<td>A</td>
<td>46</td>
<td>37</td>
<td>9</td>
</tr>
<tr>
<td>rs3760008</td>
<td>T</td>
<td>G</td>
<td>47</td>
<td>36</td>
<td>9</td>
</tr>
<tr>
<td>rs1946155</td>
<td>G</td>
<td>A</td>
<td>47</td>
<td>36</td>
<td>9</td>
</tr>
<tr>
<td>rs4784321</td>
<td>G</td>
<td>A</td>
<td>46</td>
<td>43</td>
<td>3</td>
</tr>
</tbody>
</table>

cDNA from these brain tissue samples was amplified using primers in exons 19 and 21 of *RPGRIP1L* (forward primer: 5’ TCGTGGATATCATGCCACATCA 3’, reverse primer: 5’ GCTTTGTTCTGCAAGCTGAC 3’). Two *RPGRIP1L* transcripts were identified at varying levels between individuals: a 239bp fragment corresponding to the full *RPGRIP1L* transcript including exon 20 and a 137bp fragment corresponding to a smaller *RPGRIP1L* transcript with exon 20 spliced out (Figure 36). Both fragments were gel extracted from 2.5% agarose EtBr post-stained gel, gel purified and sequenced, and both the full length and exon 20 skipped *RPGRIP1L* sequences were identified.
3.2.11.3 Quantification of RPGRIP1L Isoforms and Association to SNVs associated with RPGRIP1L exon 20 skipping

The RPGRIP1L isoforms with exon 20 spliced in or out were quantified using probe-based assays run as ddPCRs (forward primer: 5’
AGCTGACCTTCAGATAGTAAAGAC 3’, reverse primer:
5’ AGAAGGTATCTTTCGTGGATATCAT 3’, probe for full length: 5’
TGGGAACATGTGGAACAGTCAGCA 3’, probe for exon 20 skipping: 5’
ACTGCTTCTTTGTGAAACACTCTGATGT 3’). The fractional abundance of the number of copies of the exon 20 skipped isoform compared to the total number of copies of both isoforms was identified in the DLPFC and hippocampus of each sample and was run as an ANOVA with each sample’s genotype. The RPGRIP1L transcript with exon 20 spliced out is significantly decreased in the DLPFC of individuals with the schizophrenia and bipolar disorder risk alleles A, G and A of rs8047089, rs3760008 and rs1946155, respectively, with a dose allele effect (rs8047089: $p < 2e^{-16}$, $F (2, 84) = 338.9$; rs1946155: $p < 2e^{-16}$, $F (2, 84) = 271.9$; rs3760008: $p < 2e^{-16}$, $F (2, 84) = 135.9$) (Figure 37).
Figure 37. Quantification of *RPGRIP1L* transcripts in the DLPFC. Splicing ratio of the *RPGRIP1L* exon 20 spliced out transcript compared to all *RPGRIP1L* transcripts amplified shows a significant decrease with the schizophrenia and bipolar disorder risk allele in an allele dose dependent manner in the DLPFC. (Hom. = Homozygous, Maj. = Major Allele, Min. = Minor Allele, Het. = Heterozygous)

The *RPGRIP1L* transcript with exon 20 spliced out is significantly decreased in the hippocampus of individuals with the schizophrenia and bipolar disorder risk alleles A, G and A of rs8047089, rs3760008 and rs1946155, respectively, with a dose allele effect (rs8047089: $p < 2e^{-16}$, $F(2, 60) = 229.9$; rs1946155: $p < 2e^{-16}$, $F(2, 60) = 182.6$; rs3760008: $p < 2e^{-16}$, $F(2, 60) = 182.6$) (Figure 38).
Figure 38. Quantification of *RPGRIP1L* transcripts in the hippocampus. Splicing ratio of the *RPGRIP1L* exon 20 spliced out transcript compared to all *RPGRIP1L* transcripts amplified shows a significant decrease with the schizophrenia and bipolar disorder risk alleles in an allele dose dependent manner in the hippocampus. (Hom. = Homozygous, Maj. = Major Allele, Min. = Minor Allele, Het. = Heterozygous)

3.2.11.4 Protein Function Prediction of *RPGRIP1L* Isoforms

The full length *RPGRIP1L* transcript encodes for a protein with 1,315 amino acids, while the *RPGRIP1L* transcript skipping exon 20 causes an in-frame deletion of 34 amino acids and results in a protein with 1,281 amino acids. The full length and exon 20 skipped *RPGRIP1L* protein sequences were input into InterPro, PANTHER, Pfam and PredictProtein.

The deletion of 34 amino acids resulting from exon 20 skipping of *RPGRIP1L* does not affect any known domains identified by InterPro, PANTHER and Pfam. PredictProtein identified two protein binding regions in the full length *RPGRIP1L* protein sequence and four protein binding regions in the exon 20 skipped *RPGRIP1L* protein sequence.
The effect on RPGRIP1L protein function from exon 20 skipping of this transcript are unclear from the protein prediction results, and require experimental testing to determine their effects.

### 3.2.12 SNAP91

*SNAP91* (synaptosome associated protein 91) encodes a protein that localizes to the synapse and has been found to play a role in regulating synaptic vesicles in the rat hippocampus (Petralia et al., 2013). Nine SNVs were found to be associated with both schizophrenia and bipolar disorder and were identified by the DLPFC sQTLs dataset to be associated with intron 11 retention of *SNAP91* (Table 20).

#### Table 20. SNVs associated with alternative splicing of SNAP91.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Location</th>
<th>EUR MAF</th>
<th>BD $p$ value</th>
<th>SCZ $p$ value</th>
<th>Splicing $p$ value</th>
<th>Alternative Splicing Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2022265</td>
<td>NM_001242792.1:c.2015-1196T&gt;C</td>
<td>C = 0.4592</td>
<td>4.35E-05</td>
<td>6.22E-09</td>
<td>1.38E-17</td>
<td>Intron 11 retention of SNAP91</td>
</tr>
<tr>
<td>rs7741310</td>
<td>NC_000006.11:g.84250330C&gt;T</td>
<td>C = 0.4254</td>
<td>1.07E-04</td>
<td>4.38E-07</td>
<td>1.35E-08</td>
<td></td>
</tr>
<tr>
<td>rs217291</td>
<td>NM_001242792.1:c.131-18642A&gt;G</td>
<td>G = 0.4443</td>
<td>1.08E-04</td>
<td>3.73E-09</td>
<td>1.69E-21</td>
<td></td>
</tr>
<tr>
<td>rs1171114</td>
<td>NM_001170423.1:c.-125-3608T&gt;C</td>
<td>T = 0.4254</td>
<td>1.24E-04</td>
<td>1.26E-06</td>
<td>2.74E-08</td>
<td></td>
</tr>
<tr>
<td>rs217308</td>
<td>NM_001242792.1:c.453-1058A&gt;G</td>
<td>G = 0.4443</td>
<td>1.33E-04</td>
<td>7.88E-09</td>
<td>1.96E-21</td>
<td></td>
</tr>
<tr>
<td>rs217323</td>
<td>NM_001242792.1:c.659-3180C&gt;T</td>
<td>T = 0.4433</td>
<td>1.53E-04</td>
<td>2.71E-09</td>
<td>2.12E-21</td>
<td></td>
</tr>
<tr>
<td>rs1546977</td>
<td>NM_001242792.1:c.1141-981T&gt;C</td>
<td>C = 0.4433</td>
<td>2.05E-04</td>
<td>5.51E-09</td>
<td>9.45E-21</td>
<td></td>
</tr>
<tr>
<td>rs217328</td>
<td>NM_001242792.1:c.765-3101C&gt;T</td>
<td>T = 0.4433</td>
<td>2.15E-04</td>
<td>4.16E-09</td>
<td>9.70E-21</td>
<td></td>
</tr>
<tr>
<td>rs1158059</td>
<td>NM_001242792.1:c.*11-329G&gt;A</td>
<td>G = 0.4831</td>
<td>4.07E-04</td>
<td>4.46E-07</td>
<td>1.83E-12</td>
<td></td>
</tr>
</tbody>
</table>

SNV = single nucleotide polymorphism, EUR MAF = minor allele frequency for European Caucasian population, BD = bipolar disorder, SCZ = schizophrenia.

92 brain tissue samples were genotyped for the ten *SNAP91* SNVs associated with alternative splicing (Table 21).
Table 21. Brain tissue sample genotypes of SNVs associated with alternative splicing of \textit{SNAP91}.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Major Allele</th>
<th>Minor Allele</th>
<th># Individuals homozygous for Major Allele</th>
<th># Heterozygous Individuals</th>
<th># Individuals homozygous for Minor Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2022265</td>
<td>T</td>
<td>C</td>
<td>21</td>
<td>49</td>
<td>22</td>
</tr>
<tr>
<td>rs7741310</td>
<td>T</td>
<td>C</td>
<td>28</td>
<td>52</td>
<td>12</td>
</tr>
<tr>
<td>rs217291</td>
<td>A</td>
<td>G</td>
<td>30</td>
<td>47</td>
<td>15</td>
</tr>
<tr>
<td>rs1171114</td>
<td>C</td>
<td>T</td>
<td>28</td>
<td>52</td>
<td>12</td>
</tr>
<tr>
<td>rs217308</td>
<td>A</td>
<td>G</td>
<td>30</td>
<td>47</td>
<td>15</td>
</tr>
<tr>
<td>rs217323</td>
<td>C</td>
<td>T</td>
<td>30</td>
<td>47</td>
<td>15</td>
</tr>
<tr>
<td>rs1546977</td>
<td>T</td>
<td>C</td>
<td>27</td>
<td>48</td>
<td>17</td>
</tr>
<tr>
<td>rs217328</td>
<td>C</td>
<td>T</td>
<td>28</td>
<td>47</td>
<td>17</td>
</tr>
<tr>
<td>rs1158059</td>
<td>A</td>
<td>G</td>
<td>21</td>
<td>55</td>
<td>16</td>
</tr>
</tbody>
</table>

cDNA from the brain tissue samples was amplified using primers in exon 11 and exon 12 of \textit{SNAP91} (forward primer: 5’ ACTATTGACACATCCCCACCG 3’, reverse primer: 5’ CCCTCCAGAGGAAAAGTCTGG 3’). Only a 120bp fragment corresponding to the \textit{SNAP91} transcript without intron 11 retention was identified for each genotype but no 1,895bp fragment corresponding to the \textit{SNAP91} transcript with intron 11 retention was identified in the DLPFC or hippocampus (Figure 39).
Figure 39. No identification of SNAP91 transcript with intron 11 retention. A) Amplification of SNAP91 DLPFC cDNA from 7 individuals homozygous for the T, T, C, A, A, C, T, C and A alleles, 7 heterozygous individuals and 7 individuals homozygous for the C, C, T, G, G, T, C, T and G alleles of rs2022265, rs7741310, rs1171114, rs217291, rs217308, rs217323, rs1546977, rs217328 and rs1158059, respectively, only identified a 120bp fragment corresponding to the SNAP91 transcript without intron 11 retention but not a 1,895bp fragment corresponding to the SNAP91 transcript with intron 11 retention. B) Amplification of SNAP91 hippocampal cDNA from 4 individuals homozygous for the T, T, C, A, A, C, T, C and A alleles, 4 heterozygous individuals and 4 individuals homozygous for the C, C, T, G, G, T, C, T and G alleles of rs2022265, rs7741310, rs1171114, rs217291, rs217308, rs217323, rs1546977, rs217328 and rs1158059, respectively, only identified a 120bp fragment corresponding to the SNAP91 transcript without intron 11 retention but not a 1,895bp fragment corresponding to the SNAP91 transcript with intron 11 retention.

Therefore, these nine SNVs were not identified in this study to be associated with intron 11 retention of SNAP91 in the DLPFC or hippocampus.
3.2.13 **STAB1**

*STAB1* (stabilin 1) encodes a transmembrane protein that binds to gram negative and gram positive bacteria and is thought to be involved in the defense against bacterial infections (Adachi & Tsujimoto, 2002). It has been previously implicated in schizophrenia through pathway analysis of schizophrenia GWAS results (Y. H. Lee, Kim, & Song, 2013) and was identified again to be associated with schizophrenia in the most recent and comprehensive schizophrenia GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). One SNV in *STAB1* was found to be associated with schizophrenia and was found in the SPIDEX dataset to be predicted to cause exon 42 skipping (Table 22).

Table 22. SNV predicted to alter splicing of *STAB1*.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Location</th>
<th>EUR MAF</th>
<th>SCZ p value</th>
<th>Splicing Z Score</th>
<th>Alternative Splicing Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs66824127</td>
<td>NM_015136.2:c.4490-71T&gt;C</td>
<td>T = 0.4334</td>
<td>3.58E-06</td>
<td>-2.492</td>
<td>Exon 42 skipping of <em>STAB1</em></td>
</tr>
</tbody>
</table>

SNV = single nucleotide polymorphism, EUR MAF = minor allele frequency for European Caucasian population, SCZ = schizophrenia.

92 brain tissue samples were genotyped for the *STAB1* SNV predicted to alter splicing. For rs66824127, 23 individuals were found to be homozygous for the C allele, 47 individuals were found to be heterozygous and 22 individuals were found to be homozygous for the T allele.

cDNA from these brain tissue samples was amplified using primers in exons 41 and 44 of *STAB1* (forward primer: 5’ CTCCGGCAATGGCATCTTCT 3’, reverse primer: 5’ GGATGCCATCCCCGCTGTA 3’), however no or very little amplification was visible after gel electrophoresis due to the low expression of *STAB1* in the adult DLPFC and hippocampus. *STAB1* is expressed in the adolescent brain and would be an interesting candidate to test in adolescent brain tissue samples.
Chapter 4

Discussion and Conclusions

This study identified alternative transcripts that are associated with risk variants for schizophrenia, bipolar disorder or both disorders. Therefore these alternatively spliced transcripts may confer risk for these disorders. This study used multiple datasets, including two GWAS and five datasets identifying SNVs that are associated with a splicing change or predicted to cause a splicing change. The predicted alternative splicing event was then tested using cDNA from brain tissue samples to confirm the identification of these events. In six genes, multiple alternative transcripts were identified. In four of these genes, the alternative splicing events were quantified in each of the brain tissue samples and were found to be significantly associated with the schizophrenia and bipolar disorder risk alleles (Table 23). Thus, these alternative transcripts may be involved in risk for schizophrenia and bipolar disorder.
Table 23. Summary of major findings: alternatively spliced transcripts associated with schizophrenia and bipolar disorder risk alleles.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Gene</th>
<th>Splicing Dataset</th>
<th>BD p value</th>
<th>SCZ p value</th>
<th>Splicing p value from brain samples</th>
<th>Alternative Splicing Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4906337</td>
<td>APOPT1</td>
<td>PTV</td>
<td>5.27E-13</td>
<td></td>
<td>DLPFC: 1.26E-15, hippocampus: 1.1E-13</td>
<td>Decreased exon 3 skipping associated with risk allele</td>
</tr>
<tr>
<td>rs3740400</td>
<td>AS3MT</td>
<td>SPIDEX</td>
<td>6.00E-17</td>
<td></td>
<td>DLPFC: 2.85E-06, hippocampus: 8.81E-06</td>
<td>Increased exons 2 and 3 skipping associated with risk allele</td>
</tr>
<tr>
<td>rs2230535</td>
<td>NEK4</td>
<td>DLPFC sQTLs</td>
<td>2.40E-04</td>
<td>7.78E-08</td>
<td>DLPFC: 2.18E-13, hippocampus: 0.0274</td>
<td>Increased exon 4 skipping associated with risk allele</td>
</tr>
<tr>
<td>rs8047089</td>
<td>RPGRIP1L</td>
<td>DLPFC sQTLs</td>
<td>5.57E-04</td>
<td></td>
<td>DLPFC: &lt; 2E-16, hippocampus: &lt; 2E-16</td>
<td>Decreased exon 20 skipping associated with risk allele</td>
</tr>
</tbody>
</table>

4.1 Implications
Previous focus of GWAS variants was to identify non-synonymous, protein-coding changing variants that cause more straightforward changes to protein structure. However, due to the fact that many of the GWAS variants identified in psychiatric disorders could not be explained by a coding region SNV the focus was shifted to identify variants affecting gene expression, including splicing. This study identified the presence of alternative splicing changes that are associated with psychiatric disorders and suggests that these transcripts may contribute risk for these disorders.

In addition to identifying individual genes with disease associated splicing changes, these splicing changes collectively indicate pathogenic gene networks and pathways implicated in these disorders. Identifying these pathogenic gene networks and pathways is critical to
understand the underlying biology of psychiatric disorders in order to develop therapeutic targets. For example, psychiatric disorder-associated splicing changes in *NRG1, NRG3* and *ERBB4* indicate pathogenic changes to the NRG-ErbB signalling pathway, thus emphasizing the importance of identifying gene networks contributing to disease and opening the possibility of therapeutic interventions of signalling pathways impacting the network of risk genes (Law et al., 2012). The findings from this study also provide avenues for further study of the pathways where these alternatively spliced genes are found, which can help to better understand the mechanisms behind the disorder and potentially allow for new animal models and treatment targets.

### 4.1.1 Alternative transcripts converge on the mitochondria and the cilia

The four genes in which alternative transcripts were identified in this study converged on two cellular components: the mitochondria and the cilia.

*APOPT1*, a pro-apoptotic protein that localizes to the mitochondria, was identified in this study to have an alternatively spliced transcript with exon 3 skipped that was found to be decreased in the DLPFC and hippocampus of individuals with the schizophrenia risk alleles compared to individuals without the schizophrenia risk alleles. The role of mitochondrial dysfunction has already been highly implicated in both schizophrenia and bipolar disorder. Robicsek et al (2013) showed that abnormal neuronal differentiation occurred due to mitochondrial dysfunction in induced pluripotent stem cells (iPSCs) derived from schizophrenia patient hair follicles. Evidence of mitochondrial dysfunction including altered proteins and transcripts involved in mitochondrial function, energy metabolism and the oxidative stress response has been found in postmortem DLPFC tissues from schizophrenia patients compared to controls (Prabakaran et al., 2004). A decreased number of mitochondria has also been found in the prefrontal cortex and caudate of schizophrenia patients compared to controls (Uranova et al., 2001). Further, decreased complex I subunit levels and increased oxidative stress have been identified in the prefrontal cortex of patients with schizophrenia and bipolar disorder (Andreazza, Wang, Salmasi, Shao, & Young, 2013).
AS3MT, a component of the arsenic metabolism pathway, was identified in this study to have an alternatively spliced transcript with exons 2 and 3 skipped that was found to be increased in the DLPFC and hippocampus of individuals with the schizophrenia risk allele compared to individuals without the schizophrenia risk allele. As discussed previously, arsenic exposure in rats has been found to cause increased ROS and affect the mitochondria, therefore anything that affects arsenic metabolism may have an indirect consequence on the mitochondria. The predicted effect on the AS3MT protein with exons 2 and 3 skipped is a decrease in the methyltransferase efficiency, and therefore decreased arsenic metabolism efficiency. Individuals with the schizophrenia risk allele associated with the AS3MT protein with exons 2 and 3 skipped will likely have increased ROS and mitochondrial dysfunction due to decreased arsenic metabolism.

Therefore two genes involved in mitochondrial function: APOPT1 and AS3MT, were identified in this study. This supports the already established role of mitochondrial dysfunction in psychiatric disorders and also provides specific mitochondrial targets for further study.

NEK4 was identified in this study to have an alternatively spliced transcript with exon 4 skipped that was found to be increased in the DLPFC and hippocampus of individuals with the schizophrenia and bipolar disorder risk alleles compared to individuals without the schizophrenia and bipolar disorder risk alleles. RPGRIP1L was found to have an alternatively spliced transcript with exon 20 skipped that was found to be decreased in the DLPFC and hippocampus of individuals with the schizophrenia and bipolar disorder risk alleles compared to individuals without the schizophrenia and bipolar disorder risk alleles. NEK4, a serine-threonine kinase, was found to co-localize to the basal body of the primary cilium with RPGRIP1L, a known ciliopathy-associated protein (Coene et al., 2011). Therefore these proteins both converge on the primary cilia. Limited research exists on the involvement or dysregulation of the primary cilia and its relationship to schizophrenia and bipolar disorder, however one of the most highly implicated risk genes for these disorders, disrupted in schizophrenia 1 (DISC1), has been found to localize to
and play a role in the function of primary cilia (Marley & von Zastrow, 2010). DISC1 was found to localize to the basal bodies of the primary cilia in rat striatal neurons and a fibroblast cell line, and knockdown of DISC1 in these cells resulted in decreased primary cilia which was restored after re-expression of DISC1 (Marley & von Zastrow, 2010). Interestingly, this study also showed that the dopamine receptors D1R and D2R, both highly implicated in psychiatric disorders, localized to the primary cilia on these cells (Marley & von Zastrow, 2010). A more recent study by Marley & von Zastrow (2012) showed that knockdown of 20 genes associated with psychiatric disorders either through interactions with DISC1, rare variants associated with psychiatric disorders or association by GWAS also resulted in decreased ciliated cells. Interestingly, transfecting these cells with one of the top schizophrenia GWAS hits, MIR137, increased ciliation and the length of the cilia (Marley & von Zastrow, 2012). Furthermore, lithium, the most common and effective treatment of bipolar disorder, has also been found to increase the length of primary cilia in striatal neuronal cell culture and in the mouse brain (Miyoshi, Kasahara, Miyazaki, & Asanuma, 2009). Thus, multiple lines of evidence implicate the primary cilium in psychiatric disorders.

RPGRIP1L has also been found to be important for left-right asymmetry development (Vierkotten et al., 2007). Interestingly, abnormal asymmetry has been identified in schizophrenia patients. Oertel et al. (2010) identified reduced left compared to right asymmetry in the temporal lobe of schizophrenia patients when compared to controls based on gray matter volume and functional activation. First degree unaffected relatives of schizophrenia patients were also studied and showed an intermediate reduced asymmetry phenotype, suggesting a genetic component (Oertel et al., 2010). Additionally, motile cilia in the brain have been found to play an important role in neuronal migration, left-right asymmetry and body plan development (Sawamoto et al., 2006). Therefore, psychiatric disorder risk variants, such as those identified in RPGRIP1L causing alternative splicing changes in this ciliary-associated gene, may contribute to the abnormalities in left-right asymmetry and laterality observed in schizophrenia.
Therefore two genes involved in ciliary function: *NEK4* and *RPGRIP1L*, were identified in this study. This further supports role of ciliary dysfunction in psychiatric disorders and also provides specific ciliary targets for further study.

The identification of psychiatric disorder-associated alternatively spliced isoforms that converge on the mitochondria and cilia further implicates these cellular components as dysfunctional in psychiatric disorders. It not only provides novel genes to target and continue to research in terms of their relationship and contribution to risk for psychiatric disorders, but also identifies specific alternatively spliced transcripts of these genes as risk transcripts for psychiatric disorders.

### 4.1.2 Combined Alternative Splicing Model of Psychiatric Risk

Psychiatric disorders are complex genetic disorders with multiple genes and variants involved, likely each contributing a small amount of risk in an individual that manifests in disease when the genetic load passes a certain threshold or environmental factors interact with genetic risk factors to pass the threshold of disease. The identification of multiple variants that each causes a splicing change fits with this additive model. Multiple small transcript changes converging on similar pathways could cause significantly increased risk to schizophrenia and bipolar disorder. For example, one could argue that an individual with the *AS3MT* risk allele and increased amount of AS3MT exon 2 and 3 skipped transcript and protein would have a decreased tolerance to the detrimental effects of arsenic, which would increase this individual’s risk for psychiatric disorders. Along with other risk variants that affect alternative splicing, such as variants affecting *APOPT1*, this would lead to an increase in the risk for psychiatric disorders. An elevated level of variants that cause alternative splicing changes would increase an individual’s risk for psychiatric disorders, especially if these variants affect genes that act in similar pathways.

*APOPT1* is a pro-apoptotic protein that targets the mitochondria, therefore different amounts of this protein’s isoforms may affect when and under what conditions apoptosis
takes place. In conjunction with potentially increased ROS due to decreased arsenic metabolism by inefficient AS3MT, this may cause dysfunctional mitochondria that may harm the cell’s energy levels and affect neuronal function, and may not be able to be properly degraded if the apoptosis mechanisms, including the function of APOPT1, are not properly regulated.

RPGRIP1L is a protein that localizes to the primary cilia and has been found to play an important role in sonic hedgehog signaling, a signaling pathway that occurs at the primary cilia and is crucial for embryonic development and neuronal proliferation and differentiation (Vierkotten, Dildrop, Peters, Wang, & Ruther, 2007). Alterations in the levels of RPGRIP1L transcripts, in combination with alterations in other primary cilia genes including NEK4, could affect primary ciliary function as well as sonic hedgehog signaling and embryonic development, contributing to the altered neural development and connections found in psychiatric disorders.

A combination of small changes affecting mitochondrial or primary cilia function may collectively lead to dysfunction of these organelles, and ultimately lead to the manifestation of schizophrenia or bipolar disorder. Mitochondrial and primary ciliary dysfunction would lead to altered neurodevelopment, neuronal function, immuno-inflammatory pathways and response, and altered neurotransmitter system function, including dopaminergic function, all of which are hallmarks of schizophrenia and bipolar disorder and would contribute to the improper neuronal connections underlying the symptoms of these disorders.

4.1.3 Negative Results
Alternatively spliced isoforms were identified in six genes that were predicted by five datasets of sQTLs. However, in eight genes the alternative splicing change that was predicted by the five sQTL datasets used in this study was not identified. The alternative transcripts of the eight genes may not be expressed in the DLPFC or hippocampus of the brain, may be expressed in other developmental time periods or may be a result of activity dependent splicing.
Not all of the splicing prediction datasets used data from brain tissue samples to identify associations between genetic variants and splicing changes. The PTV and DLPFC sQTLs datasets used RNAseq and genotype data from brain tissue samples, however the Altrans, sQTLseekeR, and SPIDEX datasets used RNAseq and genotype data from a combination of multiple tissues to identify sQTLs. Therefore these alternatively spliced isoforms may exist in other tissues but would not be detected in this study as only brain tissues from two brain regions were used: the DLPFC and the hippocampus. Furthermore, these alternative splicing changes may not exist specifically in the adult brain, but may be present in the brain during other developmental stages. For example, the neuregulin 3 gene (NRG3) has many transcript isoforms that have differential expression in the DLPFC during different developmental periods (Paterson et al., 2017). The genes tested in this study may also have transcript isoforms that are expressed only in certain developmental periods, and therefore may not be identified due to the use of only adult brain tissue samples in this study.

Alternatively spliced transcripts that are associated with environmental stressors have also been identified, such as an increase in intron 4 retention of the acetylcholinesterase gene (AChE) following inhibition of the AChE protein or forced swimming stress in mice (Kaufer et al., 1998). Therefore these alternatively spliced transcripts may result from activity-dependent splicing and would not be identified in this study.

4.1.4 Useful method to identify splicing changes in other complex disorders
This study provides a method to identify GWAS variants that are functionally contributing risk to psychiatric disorders by altering splicing, and would be a useful model for identifying functional GWAS SNVs for other complex genetic disorders.

This study used datasets of sQTLs from the Genotype Tissue Expression Project (GTEx) and the CommonMind Consortium as well as a splicing model to determine variants that are associated with and predicted to cause an alternative splicing change. These datasets
were then mined for variants that had been found to be associated with schizophrenia and bipolar disorder in order to identify candidate sQTLs. The candidate sQTLs were then tested in brain tissue samples to determine the presence of alternatively spliced isoforms and their association to the schizophrenia and bipolar disorder risk alleles.

Alternative splicing changes have been identified in many complex genetic disorders. Therefore this method could be used for any genetic disorder by comparing the GWAS variants of the disorder with the variants found in datasets of sQTLs to identify candidate SNVs and splicing isoforms. These candidate sQTLs and their associated splicing changes could then be tested in tissue samples related to the specific disorder of interest.

4.2 Future Directions

4.2.1 Causal SNV confirmation
Genetic risk variants for psychiatric disorders were found to be associated with an alternative splicing change in this study. The mechanism of how these risk variants are causing the splicing change was predicted in this study, however whether it is these specific risk variants, a combination of variants or another variant in LD with the risk variants that are causing the associated splicing change cannot be concretely known unless it is experimentally tested. Predictions were inferred based on the sequence and SNV change, however many proteins that bind to these sequences can induce both exon inclusion and exclusion depending on the larger surrounding sequence as well the expressed trans factors. For example, Nova proteins bound to exonic sequences block exon inclusion but promote exon inclusion when bound to intronic sequences (Ule et al., 2006). Therefore a future direction of this study would be to determine whether the SNVs identified in this study are the functional SNVs causing these associated splicing changes. A mini-gene assay with the risk and non-risk alleles transfected into cell culture could be used to test the effect of each allele on the splicing of the predicted gene. A CRISPR/Cas9 assay could also be used to edit the predicted splice-altering SNVs to determine the effect of the risk and non-risk alleles on splicing.
Different splicing trans factors that bind to these sequences are also present in different cell types in different amounts, therefore a variant may cause a splicing change only in a certain cell type or tissue. It would also be important to determine whether the risk allele causes its associated splicing change only in neuronal cell types or if the splicing change is also caused by this variant in other cell types. This could be done by quantifying the expression of these alternatively spliced isoforms in other human tissues and determining the association to genotype or by performing mini-gene assays in multiple cell lines that express different splicing factors. Furthermore, determining which trans splicing factor is binding preferentially to the risk or non-risk allele and causing the splicing change is also necessary. An electrophoretic mobility shift assay (EMSA) could be performed to determine whether the splicing trans factors identified by RBPMap and SpliceAid bind preferentially to the risk or non-risk allele.

4.2.2 Experimentally tested protein function
Function at the protein level can be predicted based on the protein sequences of these alternatively spliced transcripts, however the specific protein function of each alternatively spliced transcript, especially in different cell types and environmental conditions, cannot be known until it is tested experimentally. Therefore, the function of the proteins encoded by these alternatively spliced transcripts and how they contribute to cellular and system function would need to be experimentally tested.

For each of the four genes identified in this study to have an alternatively spliced transcript associated with psychiatric disease, a mini-gene assay in which either the full length or alternatively spliced sequence was transfected into neuronal cell culture could be performed in order to determine the function of the proteins encoded by the alternatively spliced transcripts and how this function may differ from the proteins encoded by the full length transcripts. The cellular effects from both ends of the spectrum, with only the protein encoded by the full length transcript or only the protein encoded by the alternatively spliced transcript present, should be tested as well as the effects of biologically relevant levels with both proteins present at different levels based on the risk and non-risk alleles.
For APOPT1, the function of the exon 3 skipped transcript, whether it is non-sense mediated decay or a functional protein, would be identified under normal and oxidative stress conditions in neural cell types. The effect of changing the ratio of the exon 3 skipped and un-skipped APOPT1 transcripts would also be identified under normal and oxidative stress conditions due to serum withdrawal. As well, no studies of APOPT1 in brain cells exist to date to our knowledge, therefore the role of APOPT1 in neural cell types would also be determined. For AS3MT, the function of the exon 2 and 3 skipped transcript would be determined under normal and arsenic-treated conditions in neural cell types. The effect of changing the ratio of the exon 2 and 3 skipped and un-skipped AS3MT transcripts would also be identified under normal and arsenic-treated conditions. In these studies, the effect on the mitochondria, reactive oxygen species and apoptosis would be determined by examining mitochondrial morphology, oxygen consumption, the enzymatic function of mitochondrial subunits, the amount of hydroxyl radicals, carbonyl groups and nitration and the amount of cell death. As well, for AS3MT, the effect on downstream neurotransmitter systems would be determined by measuring the levels of dopamine and dopamine receptors.

For NEK4, the function of the exon 4 skipped transcript would be determined in neural cell types, as well as the effect of changing the ratio of the exon 4 skipped and un-skipped NEK4 transcripts. For RPGRIP1L, the function of the exon 20 skipped transcript would be determined in neural cell types, as well as the effect of changing the ratio of the exon 20 skipped and un-skipped RPGRIP1L transcripts. In these studies, the effect on the primary cilia would be determined. As well, because DISC1 has been found to localize to the primary cilia, the potential interaction between RPGRIP1L, DISC1 and NEK4 would be explored.

Identifying not only the alternative transcripts themselves but also the specific variants that are causing these alternative splicing changes can be beneficial when testing patient samples where the genotype for these variants is known. Cell models of schizophrenia have been created by reprogramming cells from schizophrenia patients into iPSCs and
further differentiating them into neural cells types (Tran, Ladran, & Brennand, 2013). Creating cell models from patients with a known genotype for the sQTLs identified in this study would be a useful method to test the effects of these alternatively spliced transcripts and their protein functions in neuronal cell types known to express these genes.

4.2.3 Combined alternative splicing and environmental effects

Due to the fact that these four genes exist at similar cellular structures and likely are involved in similar pathways, the combined effects of these alternative risk transcripts on cellular and neuronal function would be important to test and would be more relevant to psychiatric risk than testing their individual effects.

A future direction of the work from this study would be to determine the effect on the mitochondria of expressing the alternatively spliced transcripts of APOPT1 and AS3MT. The effect on the primary cilia of expressing the alternatively spliced transcripts of NEK4 and RPGRIP1L would also be determined. As well, the effects of expressing the alternatively spliced transcripts of all four genes (APOPT1, AS3MT, NEK4 and RPGRIP1L) would be determined as an additive model of alternative splicing. In the mitochondria, the levels of complex I subunit, oxidation, nitration and lipid peroxidation would be identified as these have been identified to be altered in the prefrontal cortex of patients with schizophrenia and bipolar disorder (Andreazza et al., 2013). The number of ciliated cells as well as the length of the cilia would be identified as knockdown of genes implicated in psychiatric disorders have been found to decrease the primary cilia (Marley & von Zastrow, 2012).

This potential dysfunction to the mitochondria and cilia would likely negatively affect neuronal differentiation as mitochondrial dysfunction has been found to effect neuronal differentiation in iPSC from schizophrenia patient hair follicles (Robicsek et al., 2013). As well, the effects of these combined alternative splicing changes on dopamine receptors 1 and 2, dopamine and tyrosine hydroxylase expression would be determined as
arsenic exposure has been found to effect the dopaminergic system, dopamine receptors have been found to localize to the primary cilia and the dopaminergic system has been found to be dysregulated in schizophrenia (Bauer, Praschak-Rieder, Kasper, & Willeit, 2012; Chandravanshi et al., 2014; Cross-Disorder Group of the Psychiatric Genomics et al., 2013; Kim et al., 2014; Marley & von Zastrow, 2010; Rodriguez et al., 2010).

Many studies of long term arsenic exposure in rats have identified harmful changes to the mitochondria and downstream neurotransmitter systems (Rodriguez et al., 2010). The $AS3MT$ isoform discussed in this study with exon 2 and 3 skipping that would likely lead to a protein with decreased arsenic metabolism is human specific, therefore it would not be present in the rats used in these studies (Li et al., 2016). An interesting future study would be to express the human specific $AS3MT$ isoform at biologically relevant levels in rats with long term arsenic exposure and non-exposure, and compare the effects to rats with only the full-length $AS3MT$ isoform expressed. It is predicted that the already detrimental effects found in rats with chronic arsenic exposure would be exacerbated in rats expressing the human-specific $AS3MT$ alternatively spliced isoform that are also chronically exposed to arsenic, and that rats not exposed to arsenic would not have these detrimental effects.

4.2.4 Splicing Therapies
Identifying alternatively spliced transcripts that are associated with disease unlocks a potential therapeutic area. Multiple mechanisms of splicing therapies have been proposed, some of which are being tested in Mendelian diseases.

One approach is to directly target the alternatively spliced transcript associated with disease. For example, the survival motor neuron protein (SMN) has two alternatively spliced isoforms, SMN1 and SMN2, which only differ due to increased exon 7 skipping in SMN2 and therefore less functional SMN protein. Spinal muscular atrophy (SMA) is caused by mutations in SMN1, which leads to a reduction in the levels of the SMN protein. Small molecule drugs have been developed as a potential treatment of SMA that
decrease the amount of exon 7 skipping of SMN2 and therefore increase the amount of functional SMN protein (Andreassi et al., 2001).

Antisense oligonucleotides (AONs) are also being developed to correct the reading frame of alternatively spliced transcripts associated with disease. Duchenne Muscular Dystrophy (DMD) is caused by mutations that change the reading frame or result in premature stop codons of the DMD gene and do not produce functional dystrophin protein. AONs have been developed to bind to the exons that contain these mutations so that they are not spliced into the dystrophin transcript and a functional protein is produced (Goyenvalle et al., 2015; Jarmin, Kymalainen, Popplewell, & Dickson, 2014).

Another approach is to target the protein isoform encoded by the disease-associated alternatively spliced transcript. For example, cyclooxygenase (COX) is expressed as two protein isoforms: cyclooxygenase-1, which is constitutively expressed, and cyclooxygenase-2, which is expressed under inflammatory conditions. Small molecule drugs that only inhibit cyclooxygenase-2 have been developed for the treatment of arthritis to block its expression under inflammatory conditions (Patrignani, Capone, & Tacconelli, 2003).

The most novel method is through spliceosome-mediated RNA trans-splicing (SMaRT). This method forms a hybrid of the target mRNA with a corrected mRNA sequence at the location of the aberrant splicing region. Frontotemporal dementia with Parkinsonism linked to chromosome 17 (FTDP-17) is caused by cis-acting mutations that cause increased splicing of exon 10 in MAPT. The SMaRT method has been used as a potential therapy for FTDP-17 that creates a hybrid of a corrected mRNA sequences without exon 10 onto the target mRNA with exon 10 to create a final mRNA sequence that does not include exon 10 (Rodriguez-Martin et al., 2009).

Identifying pathways that contain multiple genes with splicing changes associated with psychiatric disorders also provides a potential therapeutic area. Schizophrenia-associated alternative splicing changes in NRG1 and ERBB4 implicated the NRG1-ErbB4 signaling
pathway in schizophrenia. Upon further study of this pathway, the ERBB4 schizophrenia-risk haplotype associated with alternative splicing changes in ERBB4 was also found to be associated with increased PIK3CD mRNA and protein, however this increased PIK3CD was not found in post-mortem brain tissue samples from schizophrenia patients (Law et al., 2012). This was hypothesized to be a result of anti-psychotic medication use potentially decreasing the elevated PIK3CD levels, and this hypothesis was supported by amelioration of amphetamine-induced schizophrenia symptoms in rats treated with a PIK3CD inhibitor (Law et al., 2012). Thus, identification of NRG1 and ERBB4 schizophrenia-associated alternatively spliced transcripts led Law et al. (2012) to study the NRG1-ErbBr signaling pathway which was critical in identifying a PIK3CD inhibitor as a new potential treatment for schizophrenia.

Splicing based therapies and therapies identified through alternative splicing targets are proving to be very successful, therefore identifying targets for these therapies is beneficial. The alternatively spliced transcripts identified in this study could be used as targets for splicing based therapies in the future.

4.3 Conclusions
In conclusion, this study identified schizophrenia and bipolar disorder risk variants that are associated with an alternative splicing event. This study also demonstrated a unique and useful method for identifying alternatively spliced transcripts associated with GWAS variants that could be used to study other complex genetic disorders. For psychiatric disorders, the alternatively spliced transcripts identified in this study likely confer risk for these disorders and are candidates that can be further studied for their specific contribution to schizophrenia and bipolar disorder at a cellular and systems level. As well, these genes exist in similar cellular components, therefore further implicating these cellular structures in schizophrenia and bipolar disorder. Finally, the transcripts identified in this study provide useful targets for splicing therapies. Identifying variants that cause alternatively spliced transcripts is important for identifying functional variants that contribute risk to psychiatric disorders in order to better understand the mechanisms of risk for these disorders as well as to discover new therapeutic options.


associated with DNA methylation and expression of multiple genes around AS3MT. *PLoS One, 8*(1), e53732. doi:10.1371/journal.pone.0053732


(RPGRIP1L) genes and coordinates leptin receptor signaling. *J Biol Chem*, 286(3), 2155-2170. doi:10.1074/jbc.M110.188482


