2009

Schizophrenia Candidate Genes Study

Grace Lee
Virginia Commonwealth University

School of Medicine
Virginia Commonwealth University

This is to certify that the thesis prepared by Grace Hyeiwon Lee entitled SCHIZOPHRENIA CANDIDATE GENES STUDY has been approved by his or her committee as satisfactory completion of the thesis or dissertation requirement of the degree of Masters of Science.

Xiangning Chen, Ph.D., School of Medicine

Kenneth S. Kendler, M.D., Ph.D., School of Medicine

Margaret Biber, D.Phil.., School of Medicine

Mohammad Kalimi, Ph.D., School of Medicine

Diomedes E. Logothetis, Ph.D., Department Chairman

Jerome F. Strauss III, M.D., Ph.D., Dean, School of Medicine

F. Douglas Boudinot, Dean of the Graduate School

SCHIZOPHRENIA CANDIDATE GENES STUDY

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science at Virginia Commonwealth University.

By

GRACE HYEIWON LEE
B.A., University of Virginia, 2006

Director: Xiangning Chen
Associate Professor, School of Medicine

Virginia Commonwealth University
Richmond, Virginia
August 2009

Acknowledgement

First, I would like to thank my research advisor, Dr. Xiangning (Sam) Chen. The completion of this research project owes much to his guidance and support. His office door has always been open to me regardless of his numerous responsibilities. I learned much from his dedication and enthusiasm to research, patience, and insight. I have been very fortunate in coming to know him as a mentor who puts his priority in cultivating the potential of his mentees.

I gratefully acknowledge the assistance and guidance provided by my distinguished committee members, Drs Kenneth Kendler, Margaret Biber, and Mohammad Kalimi. It has been a great honor to know and work with them, and I have learned much from watching and interacting with them. I would also like to thank the members of Dr. Chen's lab for training me and working with me throughout the duration of this project. I have enjoyed our camaraderie.

Finally, and most importantly, I am deeply grateful for my family. I have an amazing family who brought me where I am today with their unconditional love. Esther and John, my life would be dull without you and the moments we cherish. I am thankful to have you as best friends, partners in crime, and counselors. Let's continue to paint this world beautifully with our dreams. I am especially thankful for my parents. They have taught me love, faith, ethical principles, and the life of contributing to the society by example. I am truly blessed to have parents that have always believed in me, prayed for me, and made every effort to ensure my well-being. This thesis embodies the times they encouraged my inquisitive pursuits and cultivated my social consciousness. To them I dedicate this thesis.

Table of Contents

List of Tables

Page

List of Figures

Page

List of Abbreviations

CATIE: Clinical Antipsychotic Trials of Intervention Effectiveness

CMYA5: Cardiomyopathy Associated 5

CNV: Copy number variation

DSM: Diagnostic and Statistical Manual of Mental Disorders

EML5: Echinoderm Microtubule-Associated Protein-Like 5

FH: Family history

GAIN: Genetic Association Information Network

GWAS: Genome-wide association studies

HWE: Hardy-Weinberg Equilibrium

ICCSS: Irish Case-Control Study of Schizophrenia

ISHDSF: The Irish Study of High-Density Schizophrenia Families

ITRIO: The Irish Trio Study of Schizophrenia

LD: Linkage disequilibrium

NIMH: National Institute of Mental Health

NOS1AP: Nitric Oxide Synthase 1 Adaptor Protein

OR: Odds ratio

PDT: Pedigree Disequilibrium Test

PTP: Protein Tyrosine Phosphatase

PTPN21: Protein Tyrosine Phosphate Non-receptor Type 21

SNP: Single nucleotide polymorphism

SZ: Schizophrenia

TDT: Transmission Disequilibrium Test

UTR: Untranslated region

Abstract

SCHIZOPHRENIA CANDIDATE GENES STUDY

By Grace Hyeiwon Lee, M.S.

A thesis submitted in partial fulfillment of the requirements for the degree of Masters at Virginia Commonwealth University.

Virginia Commonwealth University, 2009

Major Director: Xiangning Chen
Associate Professor, School of Medicine

Schizophrenia is a debilitating disorder caused by the interaction of genetic and environmental factors. In this study, we identified candidate genes and single nucleotide polymorphisms from two genome-wide association studies, GAIN and CATIE. Nine SNPs representing four candidate genes were selected for replication studies with our Irish samples: Irish Case-Control Study of Schizophrenia (ICCSS), the Irish Study of High-Density Schizophrenia Families (ISHDSF), and the Irish Trio Study of Schizophrenia (ITRIO). In the ITRIO sample, rs4704591 (CMYA5 gene) showed nominal significance ($p = 0.0447947$). Combining ICCSS, ISHDSF, and ITRIO samples for rs4704591 increased sample size and power and yielded a p-value of 0.00388. This marker remained significant after Bonferroni correction for 9 markers genotyped in this study. CMYA5 gene binds to dysbindin protein in muscle. The dysbindin gene may influence glutamatergic neurotransmission, which has been suspected of being a mechanism by which the pathophysiology of schizophrenia is manifest. Our data suggest CMYA5 gene may be associated with schizophrenia in Caucasian subjects.

CHAPTER 1: INTRODUCTION

1.1. Schizophrenia

1.1.1 Epidemiology & Significance of Schizophrenia

Schizophrenia (SZ) is a chronic, severely debilitating brain disorder with a lifetime risk of approximately 1%. The National Institute of Mental Health (NIMH) reports the number of Americans affected by SZ to be approximately 2.4 million (NIMH, 2009). Factors such as the gravity of physical and emotional suffering undergone by patients and caretakers are important in determining the significance of a disorder. In addition, the economic burden caused by a disorder is also noteworthy. A study (Wu et al., 2005) reported the overall U.S. 2002 cost of SZ was estimated to be $62.7 billion. Of the total, excess direct health care cost contributed $22.7 billion ($7.0 billion outpatient, $5.0 billion drugs, $2.8 billion inpatient, and $8.0 billion long-term care). The total indirect excess costs contributed $32.4 billion, and the total direct non-health care excess costs contributed $7.6 billion. Only 10% to 15% of people with SZ are estimated to be able to maintain full-time employment, which led to the direct excess cost due to unemployment being the largest component of overall SZ excess costs. There is no consensus on geographic and temporal variations in incidence rates. Some claim SZ occurs in diverse populations over the world at comparable rates (Jablensky et al., 1992), and that its incidence rate has remained similar over the past two centuries (Jablensky & Kalaydjieva, 2003). Others, concluded up to fivefold variation is seen in world-wide

incidence of SZ (McGrath et al., 2004). Regardless, the prevalence of SZ in the U.S. and the world establishes SZ as a vital area of research efforts.

SZ typically manifests in late adolescence or in early adult life. Females tend to have a later onset than males, and females also tend to have a milder form of the disorder with a better outcome than males. McGrath et al. (2004) reported the occurrence of SZ to be higher among males with a mean rate ratio of 1.4. SZ is also known to prevail to a greater extent in cities. Poverty, poor nutrition, inadequate healthcare and education are thought to attribute to the risk-increasing effects of urbanization. A certain migrant groups also show a higher incidence of SZ especially if they are relatively isolated with their own ethnic group in a small minority. Some of the well-replicated findings are with African-Caribbean migrants in the United Kingdom and Norwegians in the United States (Murray et al., 2008).

1.1.2. Symptoms of Schizophrenia

Schizophrenia is a multi-dimensional disorder with variable phenotype expressions (Figure 1). Such heterogeneity is in part determined by the age of onset, and patients' predominant symptoms may change as the disease progresses. Nothing the various materializations of the disorder, SZ is generally characterized by symptoms of psychosis such as hallucinations and delusions, and they are a major component of the diagnostic criteria for SZ in the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) (Figure 2). Hallucinations are generally auditory sensation in which patients believe they hear abusive or derogatory comments, although somatic or visual sensations manifest in some cases as well. Delusions, or false beliefs, range from

believing oneself to be under surveillance as a part of religious or political conspiracy to believing one's thoughts to be controlled by an evil will. Most delusions are of a paranoid nature, and some are religious, grandiose, or sexual. Hallucinations and delusions are termed positive symptoms, and they become more resistant to antipsychotic drug treatments with each succeeding schizophrenic episode. SZ patients also show derailment of thought and incoherent, illogical speech (Murray et al., 2008).

Negative symptoms of SZ include: social withdrawal, apathy, loss of motivation, slowness of thought and action, and poverty of thought and speech. As suggested by manifestations, negative symptoms accompany cognitive impairment especially in memory, attention and executive functions. Negative symptoms are prominent in chronic patients, and they accumulate gradually. These symptoms are harder to measure than positive symptoms, but they tend to be more persistent and are related to worse prognostics for patients (Murray et al., 2008). SZ patients also experience mood symptoms which make them feel hopeless and suicidal. In fact, research indicates that at least 5-13% of schizophrenia patients die by suicide, with the higher end of the range being the more accurate estimate (Pompili et al., 2007).

1.1.3. Etiology of Schizophrenia

Schizophrenia is caused by the interaction of genetic and environmental factors (Figure 3), and the genetic contribution is regarded as the most important of the known etiologic factors. The heritability of SZ liability is estimated to be up to 85%. The main sources of evidence for the genetic predisposition are studies of relatives, twin studies, and adoption studies (Murray et al., 2008). In a study with all patients with SZ on the

Roscommon County Case Register in Ireland, Kendler et al. (1993) found the lifetime risk of SZ for the first-degree relative of SZ patients to be 6.5% and that for the relatives of control subjects to be 0.5%. Other studies also report a higher incidence of SZ with relatives of SZ patients. According to NIMH, although SZ occurs in about 1 percent of the general population, it occurs at around 10 percent of the people with a first-degree relative with the disorder. NIMH also states people with a second-degree relative with SZ are also more likely than the general population to develop SZ (NIMH, 2009). Twin studies have further established genetic contributions to SZ. A monozygotic twin of a SZ patient has 41 to 65 percent chance of developing the disorder. The concordance rate is 0 to 28 percent for dizygotic twins (Cardino & Gottesman, 2000). Such studies firmly establish the role of genetics in SZ, but it is important to note that although SZ is mediated by genetics, it is not determined by it. In addition to the genetic factor, various environmental risk factors have been suggested. Although environmental risk factors operate throughout the life course, some believe the risk factors in action early in life – before or shortly after birth—cause abnormalities in the nervous system formation making an individual vulnerable to psychosis. These early inflictions include prenatal exposure to viruses, malnutrition in the womb, pregnancy complications such as bleeding, diabetes, and pre-eclampsia, complications of delivery such as uterine atony, emergency C section and asphyxia, and abnormal fetal growth and development such as low birth weight, and small head circumference. Neurodevelopmental abnormalities caused by early environmental insults may be aggravated and may lead to brain dysfunction by additional environmental factors in childhood or adolescence. Such factors include social

isolation, child abuse, and use of cannabis or other street drugs (Murray et al., 2008). In sum, genetic factors may predispose an individual to SZ, and such individual may become schizophrenic upon encountering environmental stress factors at various stages of his life.

1.1.3.1 Genetic Etiology of Schizophrenia

According to Kendler et al. (1993), what is transmitted through genes include predisposition to minor cognitive deficits, poor psychosocial functioning, suspiciousness and oddness and also to psychotic illness, schizotypal personality and paranoid personality disorder. In addition, not only the liability to SZ but also the specific clinical character of the disorder has been attributed to familial influence (Fanous & Kendler, 2005). The major focus of the genetics research on SZ has been on identifying the genes that are associated with SZ. The genetics research on SZ has come a long way since its embarkation in 1916, and more than 2000 association studies involving 500 genes have been reported for either positive or negative associations with SZ (Sun et al. 2008). Although the genetic etiology and pathogenesis of SZ remains largely elusive to date, linkage studies suggest that no one gene exists that increases the risk of SZ by more than three-fold, meaning that there may be a number of susceptibility genes (Murray et al., 2008). In fact, multiple genes and alleles in various combinations may contribute to the genetic background of the disorder, with a proportion of the transmitted genotypes remaining clinically unexpressed (Gottesman & Bertelsen, 1989).

Based on the polygenic multifactorial model of schizophrenia, many putative susceptibility genes have been found, including: DRD3, 5HT2a, DISC1, DISC2, COMT,

ProDH, RGS4, DTNBP1 (dysbindin), NRG1 (neuregulin 1), G72 and DAAO. The list of possible susceptibility genes is ever expanding, but the results of the studies showing positive associations with SZ are not well-replicated (Murray et al., 2008). It should also be considered that not all identified candidate genes may be susceptibility genes, meaning that studies suggest some genes may influence clinical features of SZ without changing the susceptibility to SZ. Such "modifier genes" have been demonstrated to be present in other diseases as well. Another challenge of unraveling the genetic etiology of SZ stems from possible genetic mechanisms that may alter the expression of correctly identified genes such as epistatsis, pleiotropy, imprinting, genetic heterogeneity, and phenocopies (Fanous & Kendler, 2005). Detailed discussion of all such genetic models of SZ is beyond the scope of this paper, but copy number variation (CNV) in particular, is an area of investigation for its undisputed pathogenic role in SZ (Williams et al., 2008).

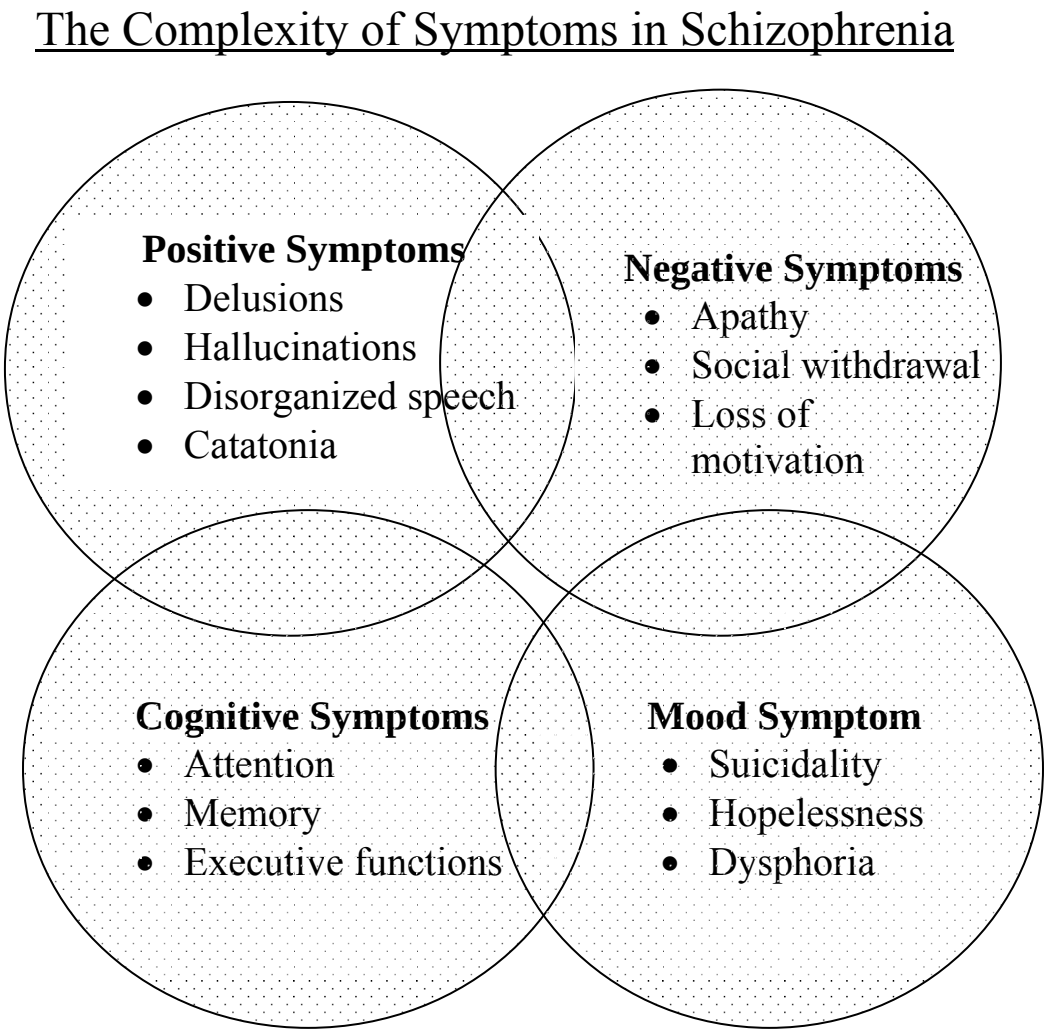


Figure 1. The Complexity of Symptoms in Schizophrenia. Schizophrenia is a multi-dimensional disorder with variable phenotype expressions. It is believed that a variety of molecular pathways involving many different susceptibility genes contributes to this heterogeneity.

1. At least two of A (only one of A required if delusions are bizarre or hallucinations consist of running commentary or discussing voices)
2. Continuous signs of disturbance for at least 6 months with at least 1 month of acute phase symptoms (A)
3. Exclusions: the disorder must not be attributable to schizoaffective or mood disorder with psychotic features, the direct effects of a substance or a general medical condition
4. Social/occupational dysfunction (below the level of expected functioning)

A

- Delusions
- Hallucinations
- Disorganized speech
- Grossly disorganized or catatonic behavior
- Negative symptoms

Figure 2. Diagnostic criteria for schizophrenia from DSM-IV-TR.

Etiologic Pathway of Schizophrenia

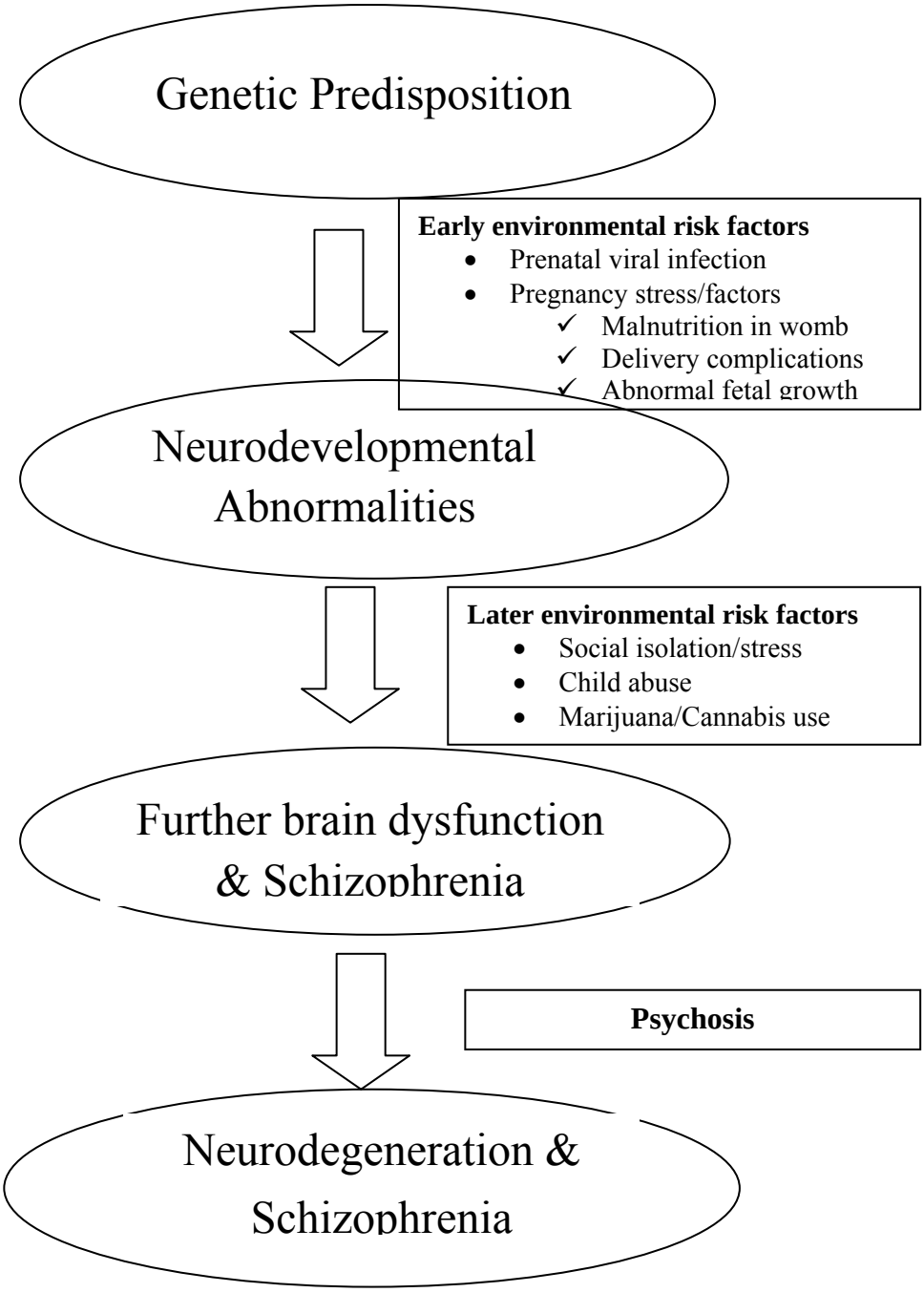


Figure 3. Etiologic Pathway of Schizophrenia. Schizophrenia is caused by the interaction of genetic and environmental factors.

1.2. Genome-wide Association Studies

1.2.1. What is a Genome-wide Association Study?

A genome-wide association study, GWAS, emerged as the completion of the Human Genome Project (2003) and the International HapMap Project (2005). These projects provided computerized databases containing the reference human genome sequence, a map of human genetic variation, and new technologies that enable analyses of whole-genome samples for genetic variations (NHGRI, 2009).

The GWAS and association studies in general are conducted using two groups of participants: people with the disease being studied, and similar people without the disease. Once the participants' blood or cells are obtained, each person's genome is purified and it is surveyed for selected markers of genetic variation called single nucleotide polymorphisms, or SNPs. If a certain genetic variation is observed significantly more frequently in people with the disease, it is said to be associated with the disease. The associated variants have their value in locating the regions of the human genome in which disease-causing problems reside (National Human Genome Research Institute [NHGRI], 2009).

What distinguishes the GWAS from other association studies is that the GWAS allows finding of genetic variations associated with particular diseases by rapidly scanning markers across the genome – the complete set of DNA –, of many individuals. This allows researchers to sample 500,000 or more SNPs from each subject, enabling them to capture variations across the genome at uniform distances between SNPs. The

ability to scan the entire genome, in turn, makes the GWAS a hypothesis-free study, in which candidates identified are more objective.

In the early stages, genetic association studies identified genes responsible for numerous monogenic disorders such as CFTR (cystic fibrosis), Huntingtin (Huntington's disease), BRCA1 (breast and ovarian cancer) and so forth. However, such approach to identify genes in multigenic diseases – diseases caused by the combined effect of multiple polymorphisms in a number of genes – has faced more challenges (Barnes, 2007). The GWAS is considered to have the potential to yield substantial insights into various disorders including psychiatric disorders. Its power to detect small effects without specific knowledge of pathogenesis has especially been useful for researching the genetics of complex disorders (Williams, Owen & O'Donovan, 2009).

1.2.2. Genome-wide Association Studies of Schizophrenia

Association studies in SZ have evolved in parallel with GWAS technology, and there have been a number of findings based on positional approaches with compelling evidence. Such findings are likely to include true susceptibility genes, but evidence has not been unequivocal, especially in terms of specific alleles or haplotypes. There have been six published GWASs of schizophrenia up to date, of which three have been based on DNA pooling and one has been limited to non-synonymous SNPs (Williams et al., 2009). These studies reported a number of significant associations with various genes including CSF2RA and SHOX (Mah et al., 2006), ZNF804A (O'Donovan et al., 2008), and reelin (Shiftman et al., 2008). And yet, no findings from the published GWAS studies were able to report a locus that reaches genome-wide levels of significance in any single

or combined study until recently (Dudbridge & Gusnanto, 2008). In 2009, however, The International Schizophrenia Consortium (2009) reported the major histocompatibility complex on chromosome 6 to have reached GWAS significance.

Williams et al. (2009) states that more robust results for other disorders have come from GWAS applied to large patient and control samples and also from follow-up analyses in even larger samples. Such approaches are being applied to SZ in recent years and have yielded some promising results. The success of the future SZ GWAS studies depends largely on assembly of large, well-phenotyped patient samples, effective collaboration and sharing of patient resources, and the ability to handle and analyze increasingly large and complex data sets.

1.3. Hypothesis of the Study

GWAS has the potential to identify genes involved schizophrenia. Although most published data sets did not identify candidate genes in these individual studies, these datasets contain valuable information that can be extracted to identify promising candidates. We hypothesized that each individual study of GWAS had useful information and that by combining multiple datasets we could increase our probability of successfully identifying promising candidate genes. To test this hypothesis, we selected all markers with p-values <= 0.05 from both GAIN and CATIE studies, and matched them against each other. We further refined our selection using linkage disequilibrium data from neighboring markers, the function of the markers and relevance to schizophrenia in PubMed databases. Using these processes, we selected a total of 9 markers in 4 genes and tested their association in our independent Irish samples.

CHAPTER 2: MATERIALS & METHODS

2.1. Subjects

All samples used were obtained and organized by other researchers prior to the current study.

2.1.1. The Irish Study of High-Density Schizophrenia Families (ISHDSF) sample

The ISHDSF sample was collected in Northern Ireland, the United Kingdom, and the Republic of Ireland. Phenotypes were assessed using DSM-III-R. The diagnoses were formed into a hierarchy of 10 categories reflecting the probable genetic relationship of these syndromes to classic SZ. This hierarchy consisted of three definitions of affection: 1) Narrow – categories D1 and D2, or "core schizophrenia" – schizophrenia, poor-outcome schizoaffective disorder and simple schizophrenia; 2) Intermediate – categories D1-D5, or a narrow definition of the schizophrenia spectrum, adding to the narrow definition schizotypal personality disorder, schizophreniform disorder, delusional disorder, atypical psychosis and good-outcome schizoaffective disorder; 3) Broad – categories D1-D8, including all disorders that significantly aggregated in relatives of schizophrenic probands in the Roscommon Family Study (Kendler et al., 1993) and adding to the intermediate definition mood incongruent and mood congruent psychotic affective illness, and paranoid, avoidant and schizoid personality disorder. The final inclusion criteria for pedigrees in the ISHDSF sample required two or more first, second, or third-degree relatives with a diagnosis of D1-D5, one or more of whom had a D1-D2

diagnosis. The sample contained 273 pedigrees and about 1350 subjects had DNA sample for genotyping. Of them, 515 were diagnosed with the narrow definition (351 males and 164 females), 634 were diagnosed with the intermediate definition, and 686 were diagnosed with the broad definition. Of these, 522 were used as cases in this study, and 869 were used as controls. Detailed descriptions of the sample were published previously (Kendler et al., 2000).

2.1.2. The Irish Case-Control Study of Schizophrenia (ICCSS) sample

The ICCSS sample was collected in Northern Ireland, the United Kingdom and the Republic of Ireland. In this study, we used 657 (436 males and 221 females) affected subjects and 411 (233 males and 178 females) control subjects. The affected subjects were selected from in-patient and out-patient psychiatric facilities in the Republic of Ireland and Northern Ireland. Subjects were eligible for inclusion if they had a diagnosis of schizophrenia or poor-outcome schizoaffective disorder by DSM-III-R criteria, and the diagnosis was confirmed by a blind expert diagnostic review. Control subjects, selected from several sources, including blood donation centers, were included if they denied a lifetime history of schizophrenia. However, the fact that control subjects were not screened by clinicians is a potential weakness in the study design. Both case and control subjects were included only if they reported all four grandparents as being born in Ireland or the United Kingdom. Family history (FH), based on the family-history research diagnostic criteria (Endicott, Andreasen & Spitzer, 1978), was assessed by clinical interview of probands and their relatives. Subjects having a first or second-degree relative

diagnosed with schizophrenia were classified as FH positive. In the ICCSS sample, there were 117 subjects (71 males and 46 females) classified with positive FH of schizophrenia.

2.1.3. The Irish Trio Study of Schizophrenia (ITRIO) sample

The ITRIO samples were collected in the same period as the ICCSS sample. The inclusion criteria and symptom ascertainments were identical to those of ICCSS samples. ITRIO samples included probands and their fraternal and maternal parents. Of the 187 families from whom samples were collected, 26 families had two affected subjects; two families had three affected subjects; and the remainder had a single affected subject. In addition, there were 29 subjects who had another first or second-degree relative diagnosed with schizophrenia. DNA samples from a total number of 216 affected subjects and 372 unaffected subjects were used for the current study.

2.2. Genome-wide Association Study Datasets

2.2.1. The Genetic Association Information Network Study

The Genetic Association Information Network (GAIN) is a public-private partnership funding genome-wide association studies established by the Foundation for the National Institute of Health. The GAIN's initial study to genotype existing research studies in six major common diseases was completed in 2007, and the resulting data are being deposited in a database that is available to the research community upon approved request. The network's initial efforts focused on Attention Deficit Hyperactivity Disorder, diabetic neuropathy in type I diabetes, major depression, psoriasis, schizophrenia, and bipolar disorder (Foundation for the National Institute of Health, 2008).

For the current study, the GAIN study data was obtained. The data included information for 906600 markers for 4505 individual samples. Only the Caucasian subset of the GAIN's samples was used for this study. There were 2601 subjects (1172 cases, 1378 controls, 51 missing phenotype; 1485 males and 1115 females) in this subset. Along with the CATIE study data, the GAIN study data was used extensively in the current study as a method of selecting candidate genes and markers.

2.2.2. The Clinical Antipsychotic Trials of Intervention Effectiveness Study

The Clinical Antipsychotic Trials of Intervention Effectiveness Study, funded by the National Institutes of Health's National Institute of Mental Health, is a clinical trial comparing the effectiveness of older (first available in the 1950s) and newer (available since the 1990s) antipsychotic medications that are used to treat schizophrenia (National Institute of Mental Health, 2009). The study's data is available to the research community upon approved request. The CATIE study's 1492 participants (741 cases and 751 controls) included SZ patients across the United States being treated in a variety of settings. For the current study, the Caucasian subset of the CATIE's samples was used. There were 492 cases and 523 controls (771 males and 244 females) in this subset, totaling 1015 subjects. The CATIE study data obtained included information for 495172 markers. The CATIE study dataset was used in the current study for selecting candidate genes and markers.

Participants for the CATIE study were eligible to be included in the study if all of the following criteria were met: 1) 18-65 years old; 2) DSM-IV diagnosis of schizophrenia; 3) adequate capacity to consent. They were excluded in the following

cases: 1) intolerance of failure to respond to one of the treatments; 2) diagnoses of schizoaffective disorder, mental retardation, pervasive developmental disorder, delirium, dementia, amnesia; 3) first episode of schizophrenia; 4) women currently pregnant or breast-feeding; 5) have a serious and unstable medical condition (Lieberman et al., 2005). The details of ascertainment and demographics of the CATIE subjects have been published by Lieberman et al. in 2005.

2.3. Management of GWAS Data

For the purpose of managing the contents of the GAIN and CATIE datasets to fit the needs of the current study, a genomic analysis software, PLINK (v. 1.05) (Purcell, 2007), was used. Using PLINK, each of the two GWAS datasets were compartmentalized into two different datasets according to the subject's race (white or black). This was done as ISHDSF, ICCSS, and ITRIO samples are all of Caucasian descent. The CATIE study data yielded 1015 white subjects and 477 non-white subjects; the GAIN study data yielded 2601 white subjects and 1904 black subjects. The CATIE study included samples whose races are neither Caucasian nor black, and thus, the term "non-white" was used.

2.4. Selection of Genes of Interest

In this study, an approach was taken in which a range of candidate genetic markers were pre-selected if they met the following criteria: 1) markers are included in both the GAIN study dataset and also in the CATIE study dataset, and 2) allelic p-values are less than 0.05 in both datasets. Allelic p-values were calculated from only Caucasian samples in both GAIN and CATIE datasets. From this process, 1229 markers covering

315 genes were pre-selected. Out of 1229 markers, 682 markers were not a part of known genes; they were precluded from further analyses.

The preceding part of the study confided in subjectivity, as no systematic scoring strategy was used. The following were considered to choose candidate genes from the 315 pre-selected genes:

1) Current knowledge regarding location of gene expression & functions of the gene: Upon literature search of individual genes, genes with known biological functions in the nervous system were given more significant consideration for biological plausibility.

2) Physical proximity to other likely candidate genes: Significant consideration was given also to genes that are physically close to other pre-select genes.

3) Number of markers from a particular gene that are included the pre-selected markers: Genes that contain numerous (4 or more) markers from the pre-select marker list were placed in a higher priority list.

4) Gene region covered by the markers in the pre-selected list (i.e. intron, UTR, near gene, etc.): A special interest was given to genes that contain markers with missense mutations. Genes whose pre-select markers are confined to the untranslated region (UTR) or near-gene regions were given relatively less consideration.

5) Previously reported association with schizophrenia or other mental disorders: Genes that have been studied related to schizophrenia or other mental disorders –bipolar disorder, Alzheimer's disease, and Parkinson's disease in particular- were given a special consideration.

2.5. Marker Selection and Genotyping

A total of nine markers from four genes that had shown significant associations ($p < 0.05$) with SZ from GAIN and CATIE white samples were selected. These markers covered the following genes: Protein Tyrosine Phosphate Non-receptor Type 21 (PTPN21), Echinoderm Microtubule-Associated Protein-Like 5 (EML5), Cardiomyopathy Associated 5 (CMYA5), and Nitric Oxide Synthase 1 Adaptor Protein (NOS1AP). All markers contained in the GWAS datasets from each selected candidate gene were extracted using PLINK (v. 1.05) (Purcell, 2007). PLINK (v.1.05) was also used to analyze the associations of these markers in samples from the GWAS datasets. Analyses were done via linear regression with the CATIE dataset, and via allelic associations for the GAIN dataset. Using the linear regression with the CATIE dataset allowed for covariates to be accounted for. The linkage disequilibrium (LD) for GAIN and CATIE datasets was examined for each candidate gene using the HAPLOVIEW program. Multiple factors were taken into consideration in electing SNPs for individual candidate genes. These include the number of markers showing nominal significance, the LD of these markers and their frequencies, and the location and nature of the polymorphisms (i.e. whether these markers change an amino acid, splicing site, transcription factor binding site etc).

All markers were typed using the TaqMan method. In the TaqMan method, the TaqMan probe consists of two types of fluorophores, which are the fluorescent parts or reporter proteins. While the probe is attached or unattached to the template DNA and before the polymerase acts, the quencher fluorophore reduces the fluorescence from the reporter fluorophore. After the denaturation of template DNA reaction cools, the TaqMan

probe binds to its specific piece of the template DNA, and the primers anneal to the DNA. Taq polymerase adds nucleotides to remove the TaqMan probe from the template DNA, which separates the quencher (at 3'-end) from the reporter (at 5'-end), and the reporter emits its energy which is quantified using a computer (Livak, 1999). SNPs typed by this method were either validated assays or custom designed assays developed by Applied BioSystems Corporation (Foster City, CA). All markers were genotyped for the ICCSS and ISHDSF samples, and only the markers for the CMYA5 gene were genotyped additionally for the ITRIO sample. All genotypes were scored using a semi-automated Excel template developed in our lab (van den Oorde, et al., 2003). All markers typed were checked for deviation from the Hardy-Weinberg Equilibrium (HWE) and Mendelian errors by the PEDSTATS program (Wigginton et al., 2005).

2.6. Statistical Analyses

The ISHDSF sample was analyzed using the pedigree disequilibrium test (PDT) (Martin et al., 2000) as implemented in the UNPHASED program (version 2.4, PDTPHASE module) (Dudbridge, 2003) for single marker associations. Single marker association tests compare frequencies of particular alleles, or genotypes, in affected and unaffected subjects. In these analyses, both vertical and horizontal transmissions were included. The p-values reported were based on weighing all families equally (the ave option in the program). For the ICCSS sample, the new version (v. 3.1) of the UNPHASED program (Dudbridge, 2008) was used to analyze single marker associations. For the ITRIO sample, since all samples were from nuclear families with no unaffected siblings, transmission disequilibrium test (TDT) as implemented in the TDTPHASE

module of the UNPHASED program was used. In the combined analyses of discovery samples (GAIN & CATIE), single marker associations were obtained using the linear/logistic regression p-value platform of the PLINK (v.1.05) program. For the combined analyses of replication samples (ICCSS + ISHDSF + ITRIO), single marker associations were performed using the UNPHASED program (v. 3.10).

Odds ratio and confidence interval were computed for all samples using version 3.1 of the UNPHASED program, as version 2.4 did not provide the information needed for the computation in the ISHDSF and ITRIO samples. In all samples, the odds ratio compared case to control allele frequencies. The test statistic is $\chi^2 = [(a/(a+c))/(b/(b+d))]$, where a = number of major alleles present in cases, c = number of minor alleles present in cases, b = number of major alleles present in controls and d = number of minor alleles present in controls. The odds ratio was constructed as (a/c)/(b/d).

CHAPTER 3: RESULTS

3.1. Selected Candidate Genes

3.1.1. Protein Tyrosine Phosphate Non-receptor Type 21 (PTPN21) Gene

Seven markers from the PTPN21 gene on Chromosome 14q31.3 were present in the pre-selected marker list (Table 1). PTPN21 and EML5 genes, also one of the selected candidate genes, are a relatively small distance away in chromosomal position. The two genes are a 60 kb away, having ZC3H14 as the only gene between them (Figure 4). PTPN21 and EML5 were the only genes aggregated in the same region of the genome with 4 or more markers for each gene present in the pre-selected marker list. A number of SZ studies to date have uncovered specific regions of the genome spanning multiple genes to be liable for the disease. Such specific regions of the genome include 5q21-31 (Pimm et al., 2005; Petryshen et al., 2005; Chen et al., 2006; Chen et al., 2007), 22q12-q13, 8p22-p21, 6p24-p22, 13q14-q32, and 6q21-q22 (Riley & Kendler, 2006). Thus, the physical proximity of the two genes, PTPN21 and EML5, presented a possibility of discovering a novel region of the genome implicated with SZ.

PTPN21 is a member of the protein tyrosine phosphatase (PTP) family. PTP proteins are known to be signaling molecules regulating various cellular processes such as cell growth, mitotic cycle, differentiation, and oncogenic transformation. PTPN21, specifically, has been shown to interact with a member of Tec tyrosine kinase, BMX/ETK. PTPN21 has also shown evidence for playing a role in liver regeneration and spermatogenesis (Zeitlin et al., 2007). Although biological roles played by PTPN21 known to date do not appear to provide obvious links to SZ, many cellular processes that

PTP family of proteins regulate have been shown to be irregular in SZ. For instance, McCurdy et al. (2006) reported significantly more mitosis in SZ patients compared to controls. Their microarray data also suggested alterations to the cell cycle in SZ. This suggests a possibility that PTPN21 may have an effect on the cell cycle in SZ.

3.1.2. Echinoderm Microtubule-Associated Protein-Like 5 (EML5) Gene

The EML5 gene, located on Chromosome 14q31.3, contained six markers in the pre-selected marker list (Table 1). This gene has been shown to be expressed in the rat brain, especially at high levels in the hippocampus, cerebellum, and olfactory bulb. EML5 has homology to EMAP, the major microtubule-associated protein in dividing sea urchin embryos. EML5 contains WD40 and HELP domains that have been suggested to be involved in microtubule binding. It has been suggested that like other EMAP-like proteins, EML5 plays a role in the regulation of cytoskeletal rearrangements during neuronal development and in adult brain (O'Conner et al., 2004). The gene being expressed in the brain made it a more attractive candidate gene. Also, studies have implied involvement of other genes regulating cytoskeletal arrangements with SZ (Hennah & Porteous, 2009; Kleppisch & Feil, 2009) which provides further evidence for the relevance of EML5 to SZ. Along with EML5's proximity to PTPN21, as discussed above, its biological significance contributed to its potential to be a gene that may be implicated with SZ.

3.1.3. Cardiomyopathy Associated 5 (CMYA5) Gene

The CMYA5 gene, located on Chromosome 5q14.1, contained three markers in the pre-selected marker list (Table 1). The CMYA5 gene is a muscle-specific member of

the TRIM superfamily. Studies have shown its expression in cardiac and skeletal muscle. It is involved in protein kinase A signaling and vesicular trafficking. So far, the gene has been reported in studies involving Duchenne muscular dystrophy and cardiac disease (Sarparanta, 2008). CMYA5 was also shown to be a binding partner for dysbindin protein in muscle, and it is intriguing to note that dysbindin, involved in the biogenesis of lysosome-related organelles, has been receiving much attention as a key SZ susceptibility factor (Benson, Tinsley, & Blake, 2004). The dysbindin gene may influence glutamatergic neurotransmission, which has long been suspected of being a mechanism by which the pathophysiology of SZ is manifest (Sodhi, Wood, & Meador-Woodruff, 2008). As previously identified SZ susceptibility genes have shown biological pathways involving multiple genes to be engaged in pathogenic mechanisms of SZ, CMYA5 gene's relation to dysbindin gene offers a motive for further investigation of CMYA5 as a candidate gene.

Out of the three markers of the CMYA5 gene in the pre-selected marker list, two of them consisted of missense mutations. This is significant because the likely outcome of proteins encoded with missense mutations can be expected to be non-functional. In fact, such mutations are responsible for a number of diseases such as sickle-cell disease and amyotrophic lateral sclerosis. In SZ research, a widely studied candidate gene DISC1 contains missense mutations that give rise to phenotypes in SZ (Millar et al., 2007). Out of the 1229 total markers in the pre-selected marker list, there were 11 missense mutations total. There were only two genes that contained 2 or more markers with missense mutations, and they were PTPN21 and CMYA5. The prevalence of missense

mutations in pre-selected markers presented a case for possible protein dysfunctions that has an effect on SZ.

3.1.4. Nitric Oxide Synthase 1 Adaptor Protein (NOS1AP) Gene

The NOS1AP gene, located on Chromosome 1q22, contained six markers in the pre-selected marker list (Table 1). This gene encodes a cytosolic protein that binds to the signaling molecule, neuronal nitric oxide synthase (nNOS), which regulates neuronal nitric-oxide (NO) synthesis (Jaffrey et al., 1998). Chromosome 1q22 is a locus initially identified as of interest from linkage studies of SZ. Recent studies have implicated the NOS1AP gene in susceptibility to SZ in various ethnic samples (Bruzustowicz et al., 2004; Zheng et al., 2005; Xu et al., 2005). However, replication studies have been inconsistent (Fang et al., 2008; Puri et al., 2006). After the completion of the current study, more studies linking NOS1AP with SZ susceptibility were published; Kremeyer et al. (2009) found eight SNPs in the NOS1AP gene to be statistically significant in SZ samples from South America. Also, Wratten et al. (2009) found the A allele of rs12742393 to be a risk allele associated with SZ via enhancing transcription factor binding and gene expression. The presence of NOS1AP gene markers with significant p-values in the pre-select marker list made the gene an appealing choice of a candidate gene supported by previous findings on the gene's association with SZ.

PTPN21							
Marker	Chr	Location (bp)	Gene Region	Polymorphism	GAIN p-value	GAIN MAF	CATIE p-value
rs2274736	14	88008405	MM	G/A	0.00106	0.354	0.01705
rs2401751	14	88016375	MM	A/G	0.00084	0.354	0.01883
rs1864744	14	88020759	Intron	C/G	0.00101	0.354	0.02501
rs7160647	14	88043437	Intron	T/C	0.00106	0.318	0.03853
rs10138002	14	88046168	Intron	A/G	0.00092	0.318	0.03834
rs10150311	14	88046225	Intron	G/A	0.00133	0.317	0.04164
rs11847417	14	88046703	intron	A/G	0.00104	0.318	0.03618

EML5							
Marker	Chr	Location (bp)	Gene Region	Polymorphism	GAIN p-value	GAIN MAF	CATIE p-value
rs10132509	14	88273534	Intron	T/C	0.0183	0.448	0.02896
rs17260415	14	88281726	Intron	G/C	0.00650	0.280	0.04058
rs10140896	14	88288291	Intron	G/C	0.00892	0.464	0.02532
rs12880096	14	88288568	Intron	T/C	0.02032	0.463	0.02744
rs7147796	14	88298322	Intron	C/G	0.03648	0.469	0.01437
rs7157149	14	88301598	Intron	G/A	0.00812	0.273	0.01

CMYA5							
Marker	Chr	Location (bp)	Gene Region	Polymorphism	GAIN p-value	GAIN MAF	CATIE p-value
rs6880680	5	79058467	Intron	C/G	0.01146	0.075	0.01209
rs3828611	5	79070418	MM	G/C	0.02248	0.058	0.01704
rs10043986	5	79131173	MM	T/C	0.04148	0.135	0.02514
rs4704591	5	79139217	UTR	G/C	0.00036	0.385	0.03813

NOS1AP							
Marker	Chr	Location (bp)	Gene Region	Polymorph -ism	GAIN p-value	GAIN MAF	CATIE p-value
rs1123217	1	160307219	Intron	G/C	0.00028	0.030	0.01887
rs4656349	1	160316448	Intron	G/A	0.01439	0.295	0.01278
rs1337062	1	160329647	Intron	C/T	0.01036	0.298	0.03275
rs1337061	1	160329876	Intron	A/G	0.00064	0.299	0.0346
rs4657150	1	160368688	Intron	C/T	0.02003	0.349	0.00914
rs164151	1	160605586	Near gene	C/T	0.02792	0.153	0.01037

Table 1. All pre-select SNPs (GAIN & CATIE p<0.05) on candidate genes. Denotation: MAF, Minor allele frequency; MM, Missense Mutation; UTR, Untranslated region.

Genomic Context of PTPN21 & EML5 Genes

Chromosome: 14; Location: 14q31.3

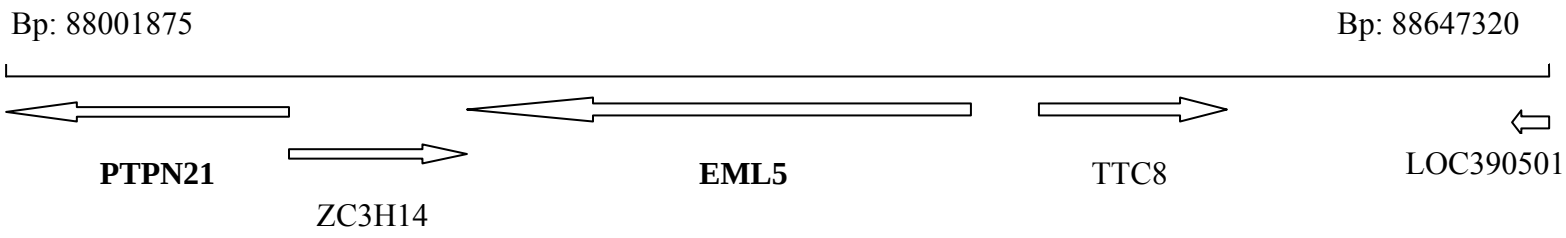


Figure 4. Genomic Context of PTPN21 & EML5 Genes. PTPN21 and EML5 genes are 60 kb away, having ZC3H14 as the only gene between them.

3.3. Linkage Disequilibrium Structures & Selected SNPs

Detailed descriptions of the final list of markers selected and genotyped are listed in Table 3. The linkage disequilibrium (LD) structures spanning the candidate genes are indicated by the pseudo-color matrix, with darker shades indicating strong LD between a pair of chromosomal markers. Markers in strong LD are transmitted together on a chromosome, and thus can be used as proxies when searching for markers associated with disease risk (Petryshen et al., 2005).

3.3.1. PTPN21

From the GAIN study, 9 markers were included for PTPN21 with the p-value of less than 0.05. In the CATIE study data, 10 markers were included for PTPN21 with the same criteria. One marker from the CATIE study was expelled as it did not meet the minimum HWE p-value (HW p-value > 0.01). There were 7 common markers between the two datasets that meet all requirements (Table 1). Out of the 7 common markers, there were 2 markers that cause missense mutations and they were included in the replication study; rs2274736 and rs2401751. The LD structure of the 7 common markers revealed that GAIN study (white samples only) grouped all 7 markers in one block, whereas the CATIE study grouped them into two separate blocks (Figure 5). Since the 2 missense mutation causing markers belong to only one of the two LD blocks in CATIE, one additional marker was selected from the LD block from which no marker has been chosen; rs10150311. Out of the 4 markers this LD block contained, rs10150311 was chosen as the tag SNP from the HaploView program. In sum, rs2274736, rs2401751, and rs10150311 were chosen to be tested for the PTPN21 gene (Table 3, Figure 10).

3.3.2. EML5

From the GAIN study white samples, 7 markers were included for EML5 with the p-value of less than 0.05. In the CATIE study data, 8 markers were included for EML5 with the same criteria. There were 6 markers from each dataset that were in common with the other (Table 1), and all of these markers were shown to be form a tight LD block (Figure 6). However, out of the 6 markers, 4 did not meet the HWE cutoff (HW p-value > 0.01) for CATIE, and these markers were omitted from the study. This yielded 2 markers to be tested for the EML5 gene; rs17260415 and rs7147796 (Table 3, Figure 11).

3.3.3. CMYA5

Two of the three markers from the pre-select list of common significant markers between GAIN (white sample) and CATIE were chosen for the study, since they cause missense mutations (rs3828611 & rs10043986). The LD maps of all common markers between GAIN and CATIE datasets, regardless of their p-values, are shown in Figure 7 and 8. Rs4704591 was later added to the current study upon comparative analyses of all markers in the gene plus 20 KB upstream and downstream sequences in the GAIN and CATIE datasets. Table 2 shows rs4704591 as the marker with the lowest p-value from the combined analyses of GAIN and CATIE Caucasian datasets. In sum, rs3828611, rs10043986, and rs4704591 were chosen to be tested for the CMYA5 gene (Table 3, Figure 12).

3.3.4. NOS1AP

Using the HaploView program, a list of four tag SNPs were made for common markers between GAIN (white sample) and CATIE. However, two of the markers did not meet the HWE criteria for the CATIE study. Therefore, rs1123217 and rs1337062 were chosen to be tested. However, the company from which the markers were ordered from failed to manufacture

rs1337062. In sum, only rs1123217 was tested in the current study (Table 3, Figure 13). Figure 9 shows LD maps of NOS1AP gene.

Marker	GAIN Allelic P-value	CATIE Allelic P-value	GAIN & CATIE Combined Allelic P-value
rs3087813	0.04322	0.31680	0.02482
rs11960229	0.16513	0.28656	0.54371
rs16877060	0.97432	0.78186	0.89937
rs6870619	0.01205	0.06333	0.00161
rs6880680	0.00942	0.02118	0.00068
rs3828611	**0.01799**	**0.06267**	**0.00310**
rs1991483	0.04335	0.08867	0.00904
rs1428227	0.35889	0.25859	0.18088
rs12655366	0.02697	0.61408	0.03383
rs259124	0.30637	0.57222	0.25216
rs259127	0.00633	0.30154	0.00445
rs259129	0.00550	0.40468	0.00547
rs259130	0.56299	0.60695	0.44294
rs12657828	0.80540	0.00307	0.06365
rs1129770	0.41612	0.33013	0.89374
rs66682	0.16659	0.65761	0.16700
rs3749683	0.20457	0.06374	0.89035
rs10043986	**0.04579**	**0.00411**	**0.00126**
rs7343	0.29541	0.30116	0.73649
rs16877214	0.17563	0.09249	0.77931
rs735639	0.02712	0.27799	0.19888
rs259067	0.27719	0.31467	0.14943
rs259066	0.21252	0.24477	0.09726
rs259064	0.04508	0.55884	0.04532
rs4704591	**0.00036**	**0.03813**	**0.00004**
rs430866	0.07734	0.55305	0.06905
rs17471700	0.12049	0.08197	0.70503

Table 2. Comparative analyses of all markers of CMYA5 gene in GAIN & CATIE datasets. rs4704591 showed the lowest p-value. P-values were computed using the UNPHASED program and are slightly different from the p-values in Table 3, which were computed using the pLink program.

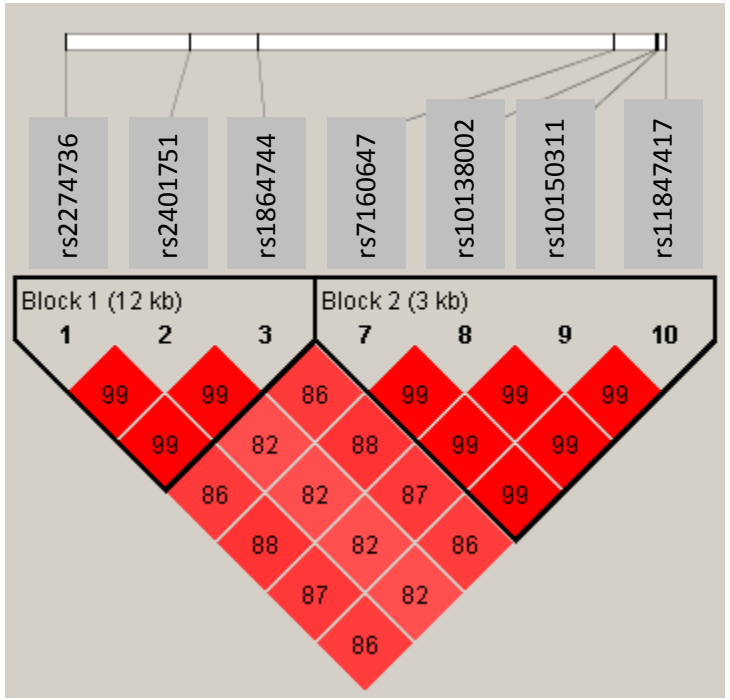
Marker ID	Marker	Gene	Polymorphism	GAIN MAF	GAIN p-value	GAIN OR	CATIE MAF	CATIE p-value	CATIE OR
1	rs2274736	PTPN21	G/A	0.354	0.00106	0.8385 [0.7455, 0.9431]	0.390	0.01705	0.8368 [0.7222, 0.9696]
2	rs2401751	PTPN21	A/G	0.354	0.00084	0.8256 [0.7335, 0.9292]	0.347	0.01883	0.8157 [0.7013, 0.9487
3	rs10150311	PTPN21	G/A	0.317	0.00133	0.8341 [0.7389, 0.9416]	0.371	0.04164	0.8811 [0.7588, 1.0232]
4	rs17260415	EML5	G/C	0.280	0.00650	0.859 [0.757, 0.9747]	0.238	0.04058	0.8195 [0.692, 0.9705]
5	rs7147796	EML5	C/G	0.469	0.03648	1.1179 [1, 1.2498]	0.367	0.01437	1.0302 [0.8866, 1.197]
6	rs3828611	CMYA5	G/C	0.058	0.02248	1.3121 [1.0476, 1.6434]	0.101	0.01704	1.4255 [0.9796, 2.074
7	rs10043986	CMYA5	T/C	0.135	0.04148	0.8435 [0.7133, 0.9973]	0.091	0.02514	0.6718 [0.5107, 0.8835]
8	rs4704591	CMYA5	C/G	0.386	0.00036	0.813 [0.7255, 0.9111]	0.385	0.03812	0.8273 [0.6915, 0.9898]
9	rs1123217	NOS1AP	G/C	0.030	0.00028	0.593 [0.4458, 0.7889]	0.124	0.01887	0.8293 [0.6408, 1.073]

MAF denotes minor allele frequency; OR denotes odds ratio.

Table 3: Final list of SNPs. Marker characteristics in GAIN & CATIE datasets.

LD Block Structures of Common Markers between GAIN & CATIE: PTPN21

1) CATIE



2) GAIN

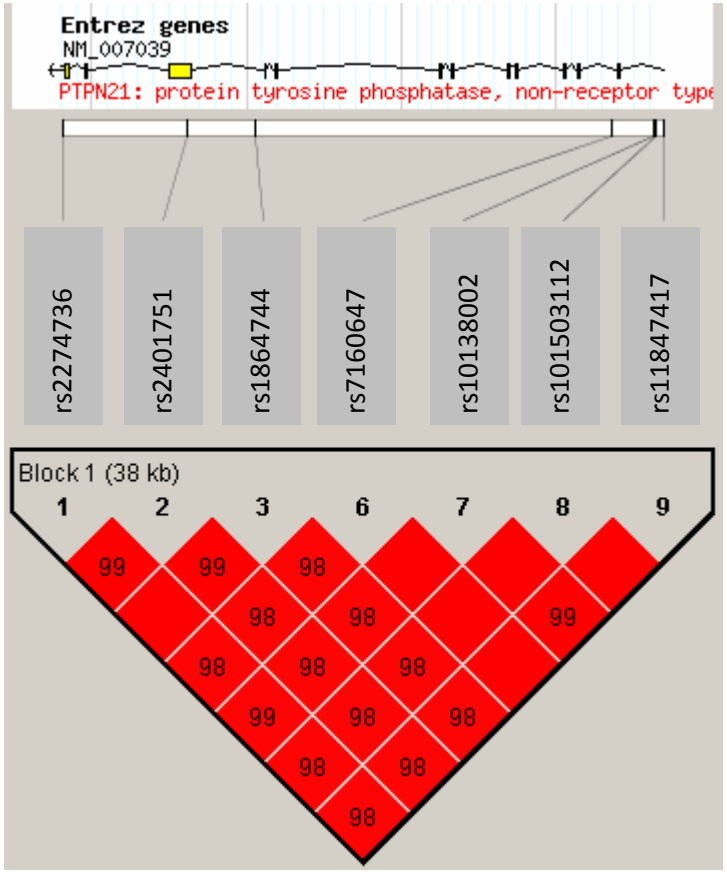
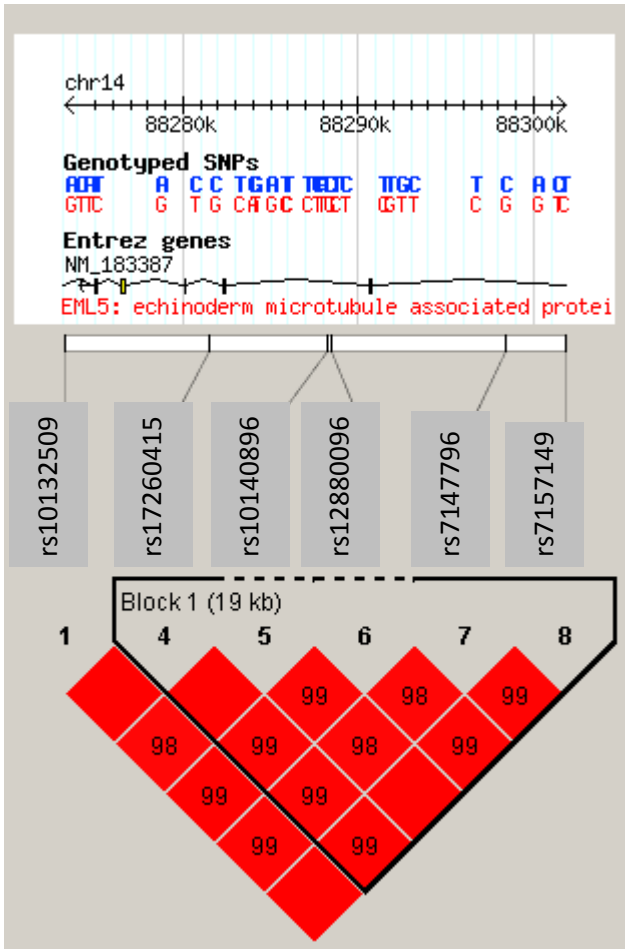


Figure 5. LD Block Structures of Common Markers between GAIN & CATIE: PTPN21

LD Block Structures of Common Markers between GAIN & CATIE: EML5

1) CATIE



2) GAIN

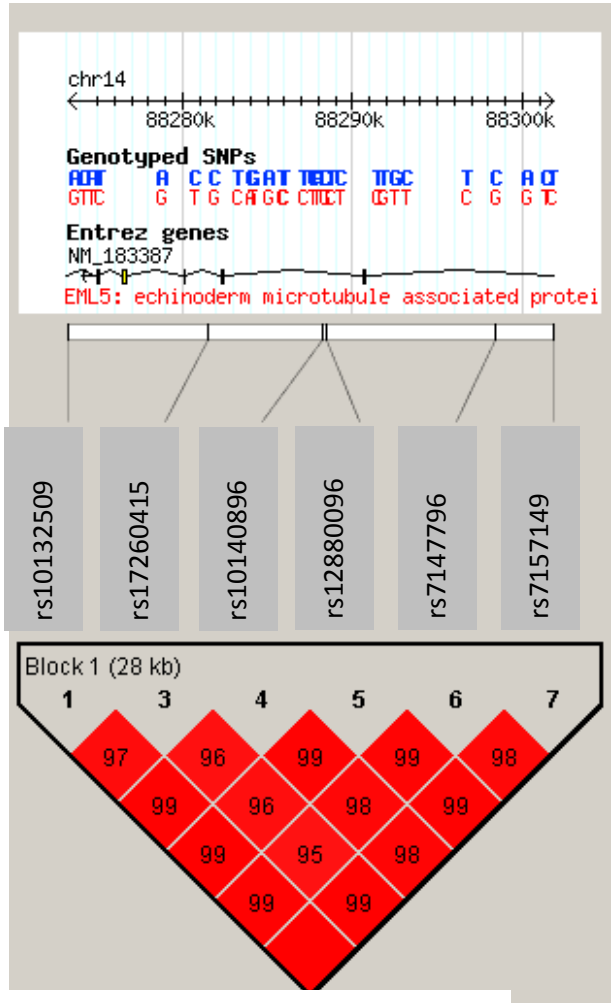


Figure 6. LD Block Structures of Common Markers between GAIN & CATIE: EML5

LD Block Structures of Common Markers between GAIN & CATIE: CMYA5

1) CATIE

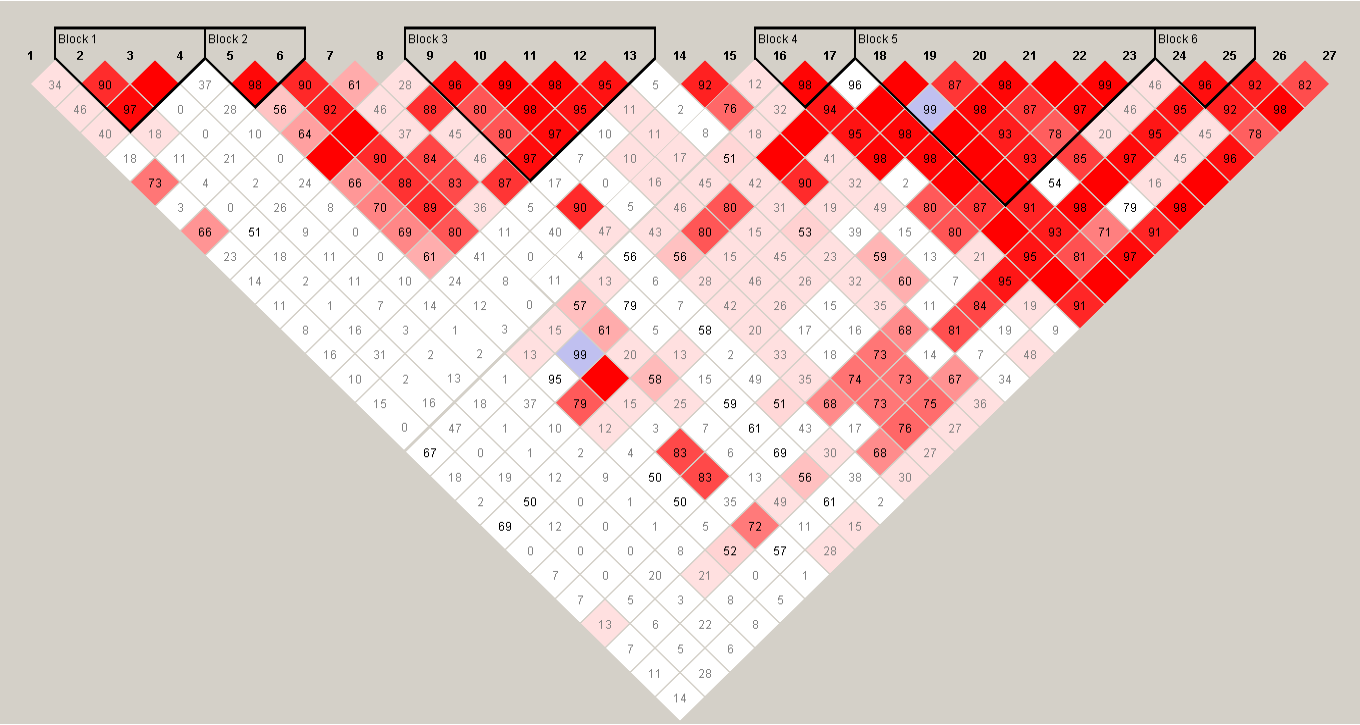
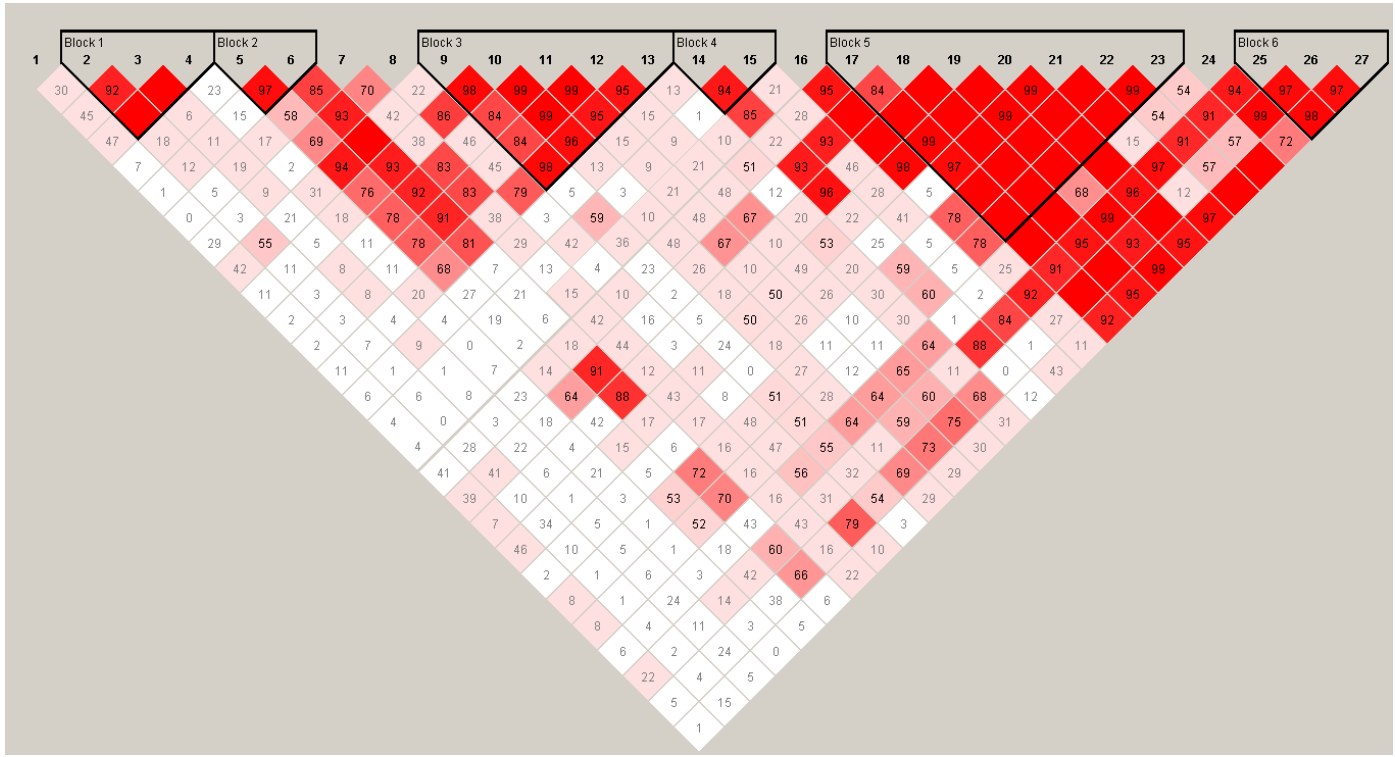


Figure 7. LD Block Structures of Common Markers between GAIN & CATIE: CMYA5 on CATIE platform. Common markers throughout the CMYA5 gene regardless of their p-values are shown here.

Denotations:
1) rs3087813
2) rs11960229
3) rs16877060
4) rs6870619
5) rs6880680
6) rs3828611
7) rs1991483
8) rs1428227
9) rs12655366
10) rs259124
11) rs259127
12) rs259129
13) rs259130
14) rs12547828
15) rs1129770
16) rs66682
17) rs3749683
18) rs10043986
19) rs7343
20) rs16877214
21) rs735639
22) rs259067
23) rs259066
24) rs259053
25) rs4704591
26) rs430866
27) rs17471700

LD Block Structures of Common Markers between GAIN & CATIE: CMYA5

2) GAIN

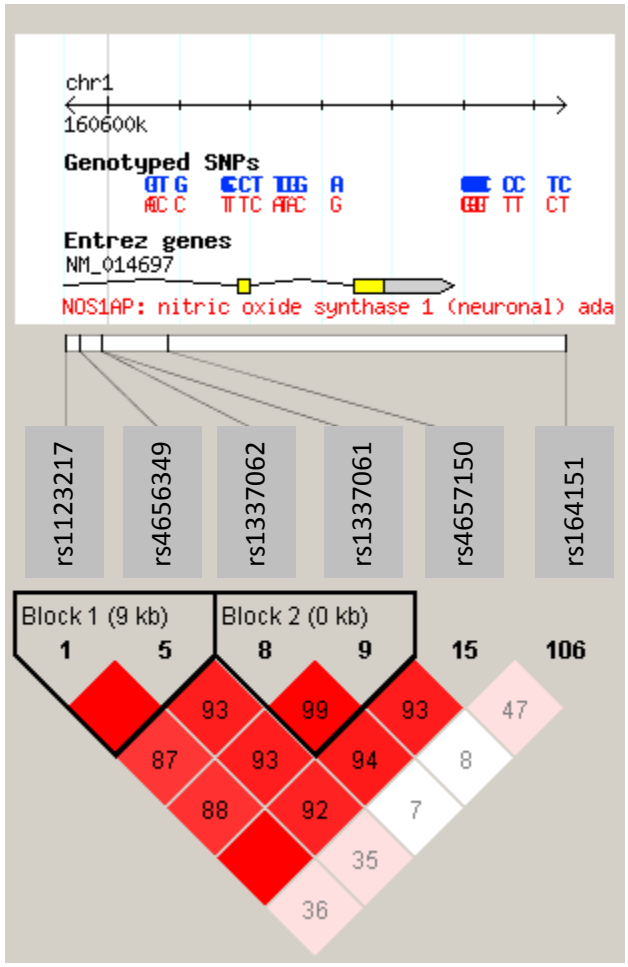


Denotations:
1) rs3087813
2) rs11960229
3) rs16877060
4) rs6870619
5) rs6880680
6) rs3828611
7) rs1991483
8) rs1428227
9) rs12655366
10) rs259124
11) rs259127
12) rs259129
13) rs259130
14) rs12547828
15) rs1129770
16) rs66682
17) rs3749683
18) rs10043986
19) rs7343
20) rs16877214
21) rs735639
22) rs259067
23) rs259066
24) rs259053
25) rs4704591
26) rs430866
27) rs17471700

Figure 8. LD Block Structures of Common Markers between GAIN & CATIE: CMYA5 on GAIN platform. Common markers throughout the CMYA5 gene regardless of their p-values are shown here.

LD Block Structures of Common Markers between GAIN & CATIE: NOS1AP

1) CATIE



2) GAIN

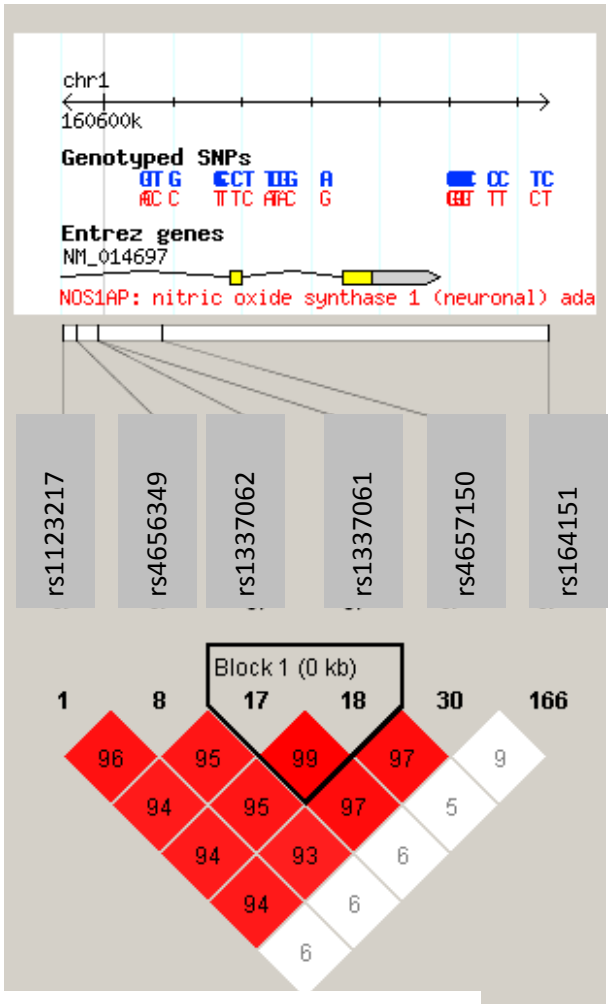


Figure 9. LD Block Structures of Common Markers between GAIN & CATIE: NOS1AP

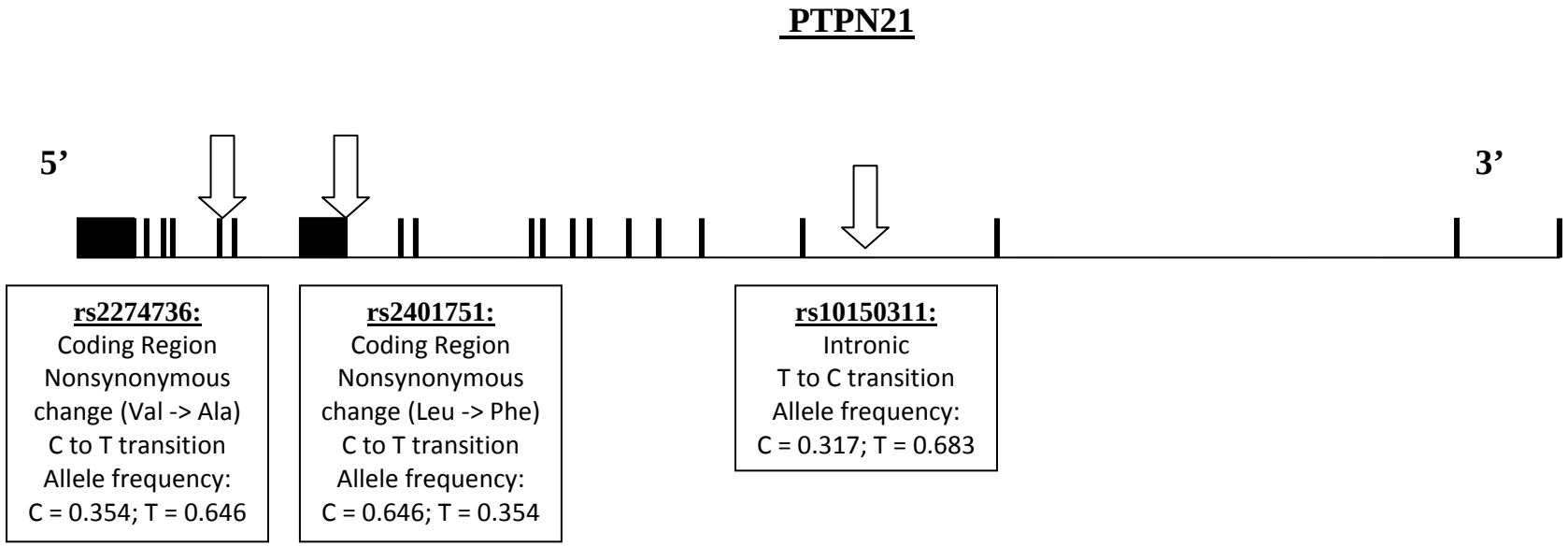


Figure 10. SNP localization along the PTPN21 gene. Location of each investigated SNP within the PTPN21 gene with their respective allelic frequencies.

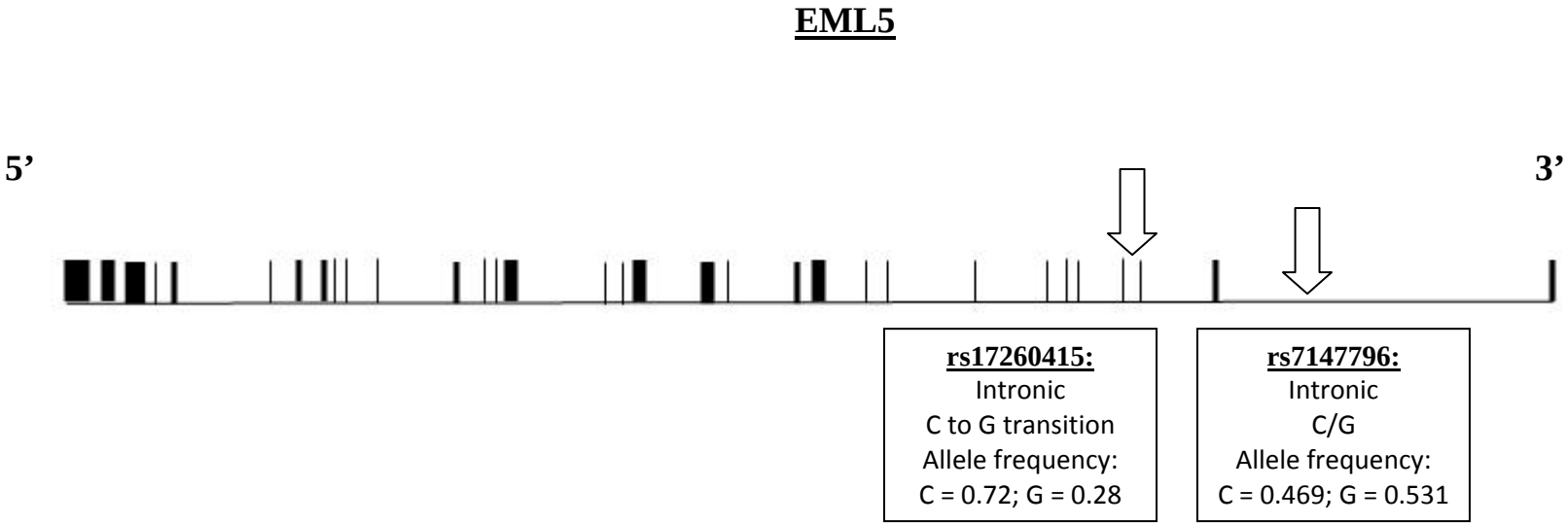


Figure 11. SNP localization along the EML5 gene. Location of each investigated SNP within the EML5 gene with their respective allelic frequencies.

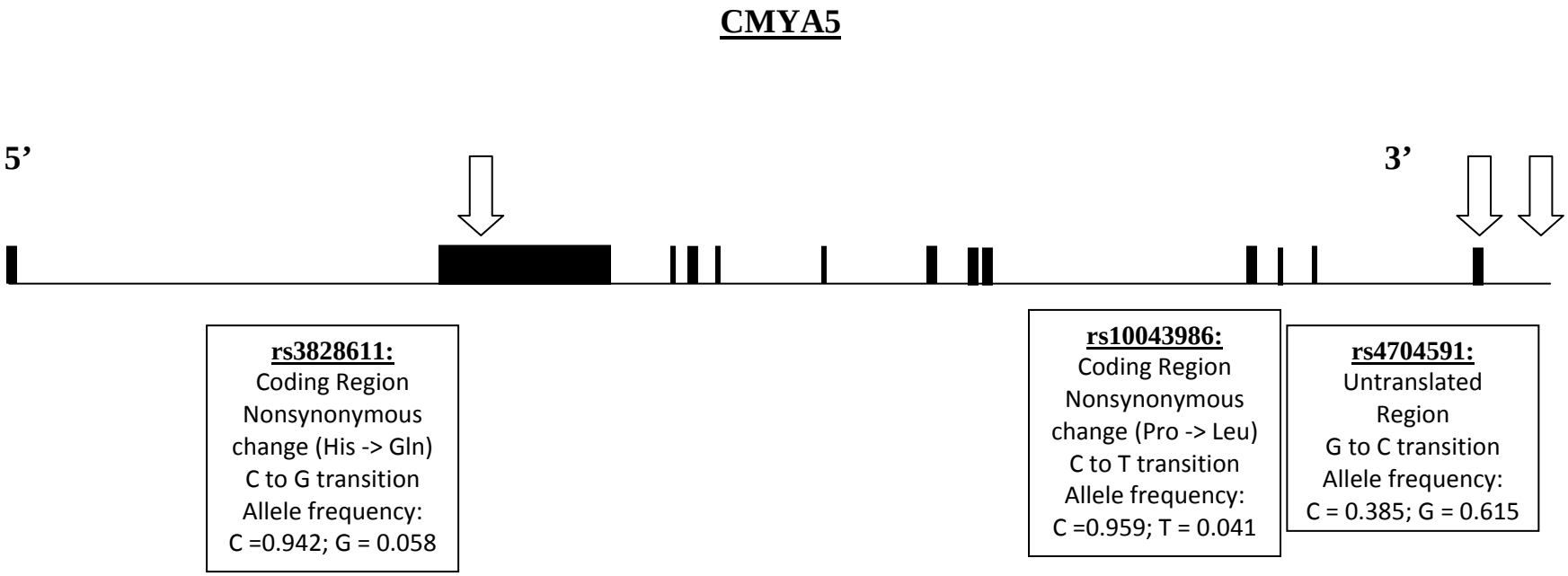


Figure 12. SNP localization along the CMYA5 gene. Location of each investigated SNP within the CMYA5 gene with their respective allelic frequencies.

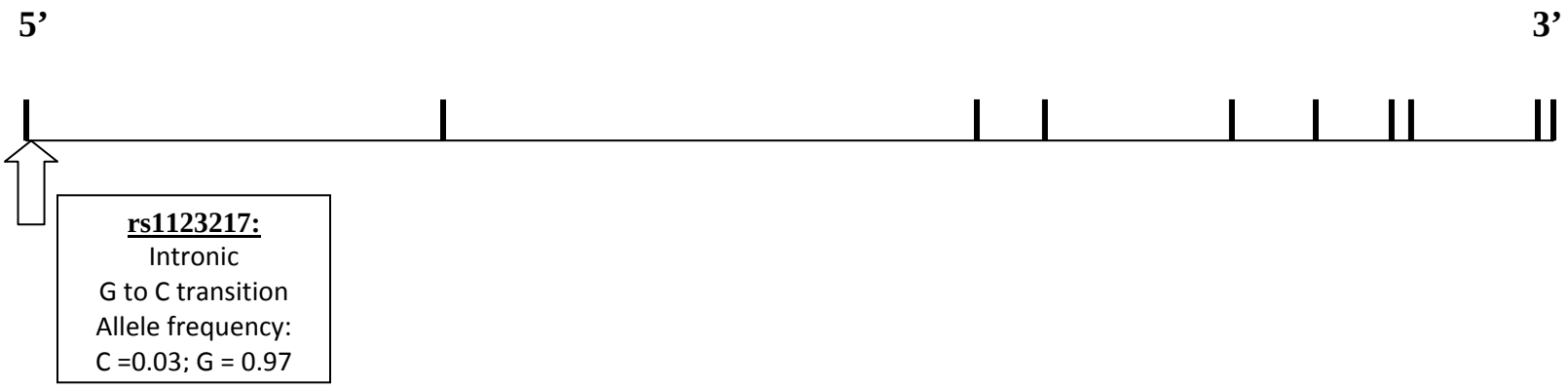


Figure 13. SNP localization along the NOS1AP gene. Location of each investigated SNP within the NOS1AP gene with their respective allelic frequencies

3.4. Statistical Results

Table 4 shows single-locus association results from the ISHDSF sample grouped by gene, and Table 5 shows those from the ICCSS sample. No marker was shown to be statistically significant ($p < 0.05$) for ISHDSF and ICCSS samples. The lowest p-value achieved was 0.0883 with rs10043986 for the CMYA5 gene in the ICCSS sample, but it is not statistically significant.

Three markers, rs3828611, rs10043986, and rs4704591, of the CMYA5 gene were additionally tested with the ITRIO sample. Marker rs4704591 from the CMYA5 gene was statistically significant in this sample ($p = 0.04479$) (Table 8). When the three replication samples (ICCSS + ISHDSF + ITRIO) are combined and analyzed, its significance improved ($p = 0.00388$), as alleles are in the same direction for the samples (Table 8 & 9). On the other hand, the combined analysis of rs3828611 resulted in a less significant p-value, as the direction of the allele was flipped for the ITRIO sample, indicated by the value of odds ratio (Table 9).

Marker ID	Marker	Gene	Polymorphism	MAF	p-value	OR
1	rs2274736	PTPN21	G/A	0.355	0.8848	0.740 [0.451, 1.212]
2	rs2401751	PTPN21	A/G	0.349	0.8108	0.775 [0.468, 1.283]
3	rs10150311	PTPN21	G/A	0.314	0.8446	0.739 [0.451, 1.212]
4	rs17260415	EML5	G/C	0.282	0.6766	0.686 [0.411, 1.142]
5	rs7147796	EML5	C/G	0.472	0.44	1.358 [0.865, 2.131]
6	rs3828611	CMYA5	G/C	0.061	0.3359	1.111 [0.451, 2.741]
7	rs10043986	CMYA5	T/C	0.123	0.7467	1 [0.554, 1.806]
8	rs4704591	CMYA5	C/G	0.376	0.0936	0.659 [0.418, 1.037]
9	rs1123217	NOS1AP	G/C	0.055	0.8657	1 [0.351, 2.851]

MAF denotes minor allele frequency; OR denotes odds ratio.

Table 4. Single marker associations (*p* values) of the ISHDSF sample

Marker ID	Marker	Gene	Polymorphism	MAF	p-value	OR
1	rs2274736	PTPN21	G/A	0.35	0.52585	0.936 [0.7911, 1.107]
2	rs2401751	PTPN21	A/G	0.345	0.67394	0.9503 [0.804, 1.123]
3	rs10150311	PTPN21	G/A	0.317	0.57809	0.9496 [0.8006, 1.126]
4	rs17260415	EML5	G/C	0.286	0.23790	0.9952 [0.8349, 1.186]
5	rs7147796	EML5	C/G	0.478	0.6808	1.044 [0.8895, 1.225]
6	rs3828611	CMYA5	G/C	0.060	0.67949	1.059 [0.7579, 1.479]
7	rs10043986	CMYA5	T/C	0.130	0.08830	0.8149 [0.6439, 1.031]
8	rs4704591	CMYA5	C/G	0.355	0.19820	0.9008 [0.7632, 1.063]
9	rs1123217	NOS1AP	G/C	0.038	0.73411	0.8977 [0.5944, 1.356]

MAF denotes minor allele frequency; OR denotes odds ratio.

Table 5. Single marker associations (p values) of the ICCSS sample

Marker ID	Marker	Gene	Polymorphism	p-value
1	rs2274736	PTPN21	G/A	0.35410
2	rs2401751	PTPN21	A/G	0.23447
3	rs10150311	PTPN21	G/A	0.13893
4	rs17260415	EML5	G/C	0.19597
5	rs7147796	EML5	C/G	0.08274
6	rs3828611	CMYA5	G/C	0.87345
7	rs10043986	CMYA5	T/C	0.42478
8	rs4704591	CMYA5	C/G	0.00388
9	rs1123217	NOS1AP	G/C	0.41842

MAF denotes minor allele frequency.

Table 6. Single marker associations (p values) of the combined replication for all markers. For CMYA5 gene, ICCSS + ISHDSF + ITRIO; for all other genes, ICCSS + ISHDSF.

			Discovery 1 (GAIN -Caucasian)			**Discovery 2 (CATIE - Caucasian)**			**Discovery Combined (1+2)**		
			Cases n = 1172 Controls n = 1378 Missing phenotype n = 51			Cases n = 492 Controls n = 523 Missing phenotype n = 0			Cases n = 1686 Controls n = 1901 Missing phenotype n = 51		
Chr./Mb	**SNP**	**Minor Allele**	SZ	CON	Allelic P-value	SZ	CON	Allelic P-value	SZ	CON	Allelic P-value
5/79070418	rs3828611	G	0.074	0.058	0.02248	0.068	0.049	0.01704	0.074	0.056	0.00310
5/79131173	rs10043986	T	0.116	0.135	0.04148	0.098	0.139	0.02514	0.110	0.136	0.00126
5/79139227	rs4704591	C	0.359	0.408	0.00036	0.362	0.406	0.03813	0.360	0.408	0.00004

SZ and CON; allele frequency in schizophrenia and controls.

Table 7. CMYA5 Gene Combined Discovery Sample (GAIN & CATIE) Analyses.

		Replication 1 (ICCSS)			**Replication 2 (ISHDSF)**			**Replication 3 (ITRIO)**			**Replications Combined (1+2 +3)**		
		Cases n = 657 Controls n = 411			Cases n = 522 Controls n = 869			Cases n = 216 Controls n = 372			Cases n = 1395 Controls n = 1652		
Chr./Mb	**SNP**	MA SZ	MA CON	P-value	MA SZ	MA CON	P-value	MA SZ	MA CON	P-value	MA SZ	MA CON	P-value
5/79070418	rs3828611	0.061	0.058	0.67949	N/A	N/A	0.3359	0.041	0.052	0.48179	0.058	0.057	0.87345
5/79131173	rs10043986	0.119	0.142	0.08830	N/A	N/A	0.7467	0.094	0.010	0.79855	0.119	0.142	0.42478
5/79139227	rs4704591	0.343	0.368	0.19820	N/A	N/A	0.0936	0.346	0.419	0.04479	0.343	0.367	0.00388

MA SZ and MA CON; minor allele frequency in schizophrenia and controls.

Table 8. CMYA5 Gene Combined Replication Sample (ICCSS + ISHDSF + ITRIO) Analyses. Single marker associations were calculated for ICCSS, ISHDSF, and ITRIO samples respectively and also for the combined samples (ICCSS + ISHDSF + ITRIO).

Marker ID	Marker	GAIN OR	CATIE OR	ICCSS OR	TRIADS OR	ISHDSF OR
1	rs2274736	0.8385 [0.7455, 0.9431]	0.8368 [0.7222, 0.9696]	0.936 [0.7911, 1.107]	N/A	0.7396 [0.4514, 1.212]
2	rs2401751	0.8256 [0.7335, 0.9292]	0.8157 [0.7013, 0.9487]	0.9503 [0.804, 1.123]	N/A	0.775 [0.4681, 1.283]
3	rs10150311	0.8341 [0.7389, 0.9416]	0.8811 [0.7588, 1.0232]	0.9496 [0.8006, 1.126]	N/A	0.7393 [0.4507, 1.212]
4	rs17260415	0.859 [0.757, 0.9747]	0.8195 [0.692, 0.9705]	0.9952 [0.8349, 1.186]	N/A	0.6855 [0.4114, 1.142]
5	rs7147796	1.1179 [1, 1.2498]	1.0302 [0.8866, 1.197]	1.044 [0.8895, 1.225]	N/A	1.358 [0.8649, 2.131]
6	rs3828611	1.3121 [1.0476, 1.6434]	1.4255 [0.9796, 2.0743]	1.059 [0.7579, 1.479]	0.7778 [0.3868, 1.564]	1.111 [0.4505, 2.741]
7	rs10043986	0.8435 [0.7133, 0.9973]	0.6718 [0.5107, 0.8835]	0.8149 [0.6439, 1.031]	0.9355 [0.5637, 1.552]	1 [0.5538, 1.806]
8	rs4704591	0.813 [0.7255, 0.9111]	0.8273 [0.6915, 0.9898]	0.9008 [0.7632, 1.063]	0.747 [0.5553, 1.005]	0.6585 [0.4182, 1.037]
9	rs1123217	0.593 [0.4458, 0.7889]	0.8293 [0.6408, 1.073]	0.8977 [0.5944, 1.356]	N/A	1 [0.3508, 2.851]

OR denotes odds ratio. Brackets indicate 95% confidence intervals for the OR.

Table 9. Odds Ratio Analyses of All Samples. All samples were analyzed using UNPHASED v.3.1.

CHAPTER 4: DISCUSSION

The SZ research community has produced a plethora of research data involving a large number of candidate genes and various ethnic groups as samples. As this trend continues, it has become increasingly important to devise methods to effectively and systematically integrate accumulated data in order to select a list of "prioritized" candidate genes (Le-Niculescu et al., 2007). Publicly available genome-wide association studies such as GAIN and CATIE introduced in this study not only cover markers across the genome, but also provide information for a large number of samples. The current study takes an approach of selecting candidate genes by relying on statistical evidence from not just one set of GWAS data, but on two. This study design assumed the hypothesis that p-values from GWAS datasets suggest an increased likelihood of such markers being replicated in other populations. Although each of these large GWAS datasets is able to provide immensely useful information on its own, a successful integration of the datasets increases the chance of selecting promising candidate genes for further examinations by basing the selection upon evidence from independent samples. This is likely to improve the reliability for candidate selection.

From the results of the current study, we successfully replicated one of the 9 markers tested, or one of 4 candidate genes. Our performance is slightly better than random selection. However, it is difficult to attribute the better performance strictly to statistical evidence (p-value) from GWAS datasets, because, judging from the p-values of the markers selected from both GAIN and CATIE, rs4704591, the one marker successfully confirmed, is not the best one among the 9 selected markers (see data in Table 3).

Although the correctness of our hypothesis remains uncertain, the scarcity of statistically significant findings from the current study may be partially accounted for by the lack of power due to small replication sample sizes. A review of odds ratios (Table 9) reveals that for some of the markers genotyped, the sample sizes were not efficient to detect their effects. In addition to the statistical significance found with rs4704591 (CMYA5 gene) in the ITRIO sample, modest effects observed in other samples for the same marker were seen to be in the same direction on the same allele, which led to the p-value of the combined analyses of all replication samples to be relatively small (p-value = 0.00388). Thus, more replications are needed to verify CMYA5 as a true susceptibility gene for Schizophrenia and also for other genes in the study. If additional replications still produce negative associations of such genes with SZ, such reports provide useful information so that the next generation of research studies can be appropriately focused.

Schizophrenia is thought to be caused by a large number of contributing loci with subtle effects, interactions (gene-gene, gene-environment, intralocus), complex single locus effects, or parent-of-origin or epigenetic effects (The Psychiatric GWAS Consortium Steering Committee, 2009). This fact points us to areas in which future studies can improve on. One could argue that selecting significant p-values from a simple additive genetics model of GWAS may not be sufficient for discovering candidate genes and SNPs for SZ, which is thought to not conform to this model. In addition, whereas GWAS assess only a subset of genetic variants (SNPs and CNVs), there are many other types of genetic variants that may be important and that could be systematically incorporated into candidate selection. Another possible area of error may be found in analyzing the combined CATIE and GAIN datasets. A direct overlap between different platforms is often modest at best, and these platforms must be made comparable in terms of

SNPs included, subjects, and other factors of bias such as population stratification, cryptic relatedness or genotyping batch effects (The Psychiatric GWAS Consortium Steering Committee, 2009). In the current study, efforts were made to eliminate ethnic heterogeneity of subjects present in the CATIE study that is not comparable with the GAIN study by employing only Caucasian samples for our analyses. However, other factors of platform compatibility were not resolved, and future studies may benefit from considering the issue.

Out of the four genes investigated for association with schizophrenia in the current study, a promising result emerged from the CMYA5 gene. As for the markers that failed to show significance, the knowledge of what a disorder is not, has its value in appropriately focusing the next generation of research. The positive association found with CMYA5 is exciting, as the discovery was made with the presence of a compelling association without an obvious biological function of the gene in the nervous system. CMYA5 being a binding partner for dysbindin in muscles sheds light on the effects of glutamatergic neurotransmission on the pathophysiology of schizophrenia in which dysbindin is involved. Glutamatergic neurons innervate the cortex, limbic regions, and striatum as a part of excitatory afferent and efferent systems. A family of glutamate-gated ion channels that depolarize neurons mediate the postsynaptic actions of glutamate. Psychotomimetics such as phencyclidine act on one of these receptors, the N-methyl-D-asparate (NMDA) receptor, and reproduce schizophrenic symptoms in normal individuals. It has been suspected that glutamatergic dysfunction is especially relevant to schizophrenia forms in which negative symptoms, cognitive deficits, and deterioration are prominent. In addition, drugs that enhance NMDA-receptor functions reduce such symptoms in chronic schizophrenia patients (Coyle, 1996). The association made with the CMYA5 gene gives future researchers a

starting point to determine how variations in related glutamatergic pathways or genetic regions affect vulnerability to schizophrenia. Currently, our laboratory is collaborating with other institutions in order to replicate our results on the CMYA5 gene in other samples. Long-term goals of the current study may enable clinicians to prevent schizophrenia, or to identify cases earlier for more efficacious treatments which may lead to decreased mortality.

For the GWAS data based on individual genotyping thus far reported, the total combined sample sizes are still small (fewer than 1500 cases), and the power to identify small genetic effects is limited (Williams et al., 2009). The current study made an improvement on previous GWAS-based studies by increasing the discovery sample size to better approximate marker significances. Although it is inconclusive how beneficial this approach was in the current study, as a result, we identified CMYA5 to be associated with schizophrenia in our Irish samples. The finding provides a cause for optimism that larger scale of future studies will be able to identify additional loci associated with schizophrenia.

List of References

List of References

Benson, M.A., Tinsley, C.L., & Blake, D.J. (2004). Myospryn is a novel binding partner for dysbindin in muscle. *Journal of Biological Chemistry, 279*(11), 10450-10458.

Brzustowicz, L.M. (2008). NOS1AP in schizophrenia. *Current Psychiatry Reports, 10*(2), 158-63.

Bruzustowicz, L.M., Simone, J., Mohseni, P., Hayter, J.E., Hodgkinson, K.A., et al. (2004). Linkage disequilibrium mapping of schizophrenia susceptibility to the CAPON region of chromosome 1q22. *American Journal of Human Genetics, 74*, 1057-1063.

Cardino, A.G., & Gottesman, I.I. (2000). Twin studies of schizophrenia: From bow-and-arrow concordances to star wars Mx and functional genomics. American Journal of Medical Genetics, 97(1), 12-7.

Chen, X., Wang, X., Hossain, S., O'Neill, F.A., Walsh, D., Pless, L., et al. (2006). Haplotypes spanning SPEC2, PDZ-G EF2 and ACSL6 genes are associated with schizophrenia. *Human Molecular Genetics, 15*, 3329-3342.

Chen, X., Wang, X., Hossain, S., O'Neill, F.A., Walsh, D., van den Oord, E.J., et al. (2007). Interleukin 3 and schizophrenia: The impact of sex and family history. *Molecular Psychiatry, 12*, 273-282.

Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE): NIMH Study To Guide Treatment Choices for Schizophrenia. (2005). Retrieved March 10, 2009, from National Institute of Mental Health Web Site: http://www.nimh.nih.gov/health/trials/practical/catie/index.shtml

Coyle, J.T. (1996). The glutamatergic dysfunction hypothesis for schizophrenia. *Harvard Review of Psychiatry, 3*(5), 241-253.

Dudbridge, F. (2003). Pedigree disequilibrium tests for multilocus haplotypes. *Genetic Epidemiology, 25*, 115-121.

Dudbridge, F. (2008). Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Human Heredity, 66,* 87-89.

Dudbridge, F., Gusnanto, A. (2008). Estimation of significance thresholds for genomewide association scans. *Genetic Epidemiology, 32*, 227-234.

Endicott, J., Andreasen, N., Spitzer, R.L. (1978). Family history research diagnostic criteria. New York State Psychiatric Institute: New York, 1978.

Fang, C., Tang, W., Tang, R.Q., Wang, L., Zhou, G.Q., Huang, K., et al. (2008). Family-based association studies of CAPON and schizophrenia in the Chinese Han population. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 32*(5), 1210-1213.

Fanous, A.H., Kendler, K.S. (2005). Genetic heterogeneity, modifier genes, and quantitative phenotypes in psychiatric illness: Searching for a framework. Molecular *Psychiatry, 10*, 6-13.

Foundation for the National Institute of Health. (2008). *Genetic Association Information Network (GAIN).* Retrieved March 10, 2009, from Foundation for the National Institute of Health Web Site: http://www.fnih.org/index.php?option=com_content&task=view&id=338&Itemid=454

Gottesman, I.I. (1991). Schizophrenia genesis: The origins of madness. New York: W.H. Freeman.

Gottesman, I.I., Bertelsen, A. (1989). Confirming unexpressed genotypes for schizophrenia. *Archives of General Psychiatry, 48*, 867-872.

Hennah, W., & Porteous, D. (2009). The DISC1 pathway modulates expression of neurodevelopmental, synaptogenic and sensory perception genes. *PLoS ONE, 4*(3), e4906.

Jablensky, A.V., Kalaydjieva, L.V. (2003). Genetic epidemiology of schizophrenia: Phenotypes, risk factors, and reproductive behavior. *American Journal of Psychiatry, 150*, 425-429.

Jablensky, A., Sartorius, N., Ernberg, G., Anker, M., Korten, A., Cooper, J.E., et al. (1992). Schizophrenia: Manifestations, incidence and course in different cultures: A World Health Organization ten-country study. *Psychological Medicine Monograph Supplement, 20*, 1-97.

Jaffrey, S.R., Snowman, A.M., Eliasson, M.J., Cohen, N.A., Snyder, S.H. (1998). CAPON: a protein associated with neuronal nitric oxide synthase that regulates its interactions with PSD95. *Neuron, 20*(1), 115-124.

Kendler, K.S. (2003). The genetics of schizophrenia: Chromosomal deletions, attentional disturbances, and spectrum boundaries. *American Journal of Psychiatry, 160*(9), 1549-1553

Kendler, K.S., McGuire, M., Gruenberg, A.M., O'Hare, A., Spellman, M., Walsh, D. (1993). The Roscommon Family Study. III. Schizophrenia-related personality disorders in relatives. *Archives of General Psychiatry, 50*, 781-788.

Kendler, K.S., Myers, J.M., O'Neill, F.A., Martin, R., Murphy, B., MacLean, C.J., et al. (2000). Clinical features of schizophrenia and linkage to chromosomes 5q, 6p, 8p, and 10p in the Irish Study of High-Density Schizophrenia Families. *American Journal of Psychiatry, 157*, 402-408.

Kleppisch, T., & Feil, R. (2009). cGMP signaling in the mammalian brain: role in synaptic plasticity and behavior. *Handbook of Experimental Pharmacology, 191*, 549-579.

Kremeyer, B., Garcia, J., Kymalainen, H., Wratten, N., Restrepo, G., Palacio, C., et al. (2009). Evidence for a role of the NOS1AP (CAPON) gene in schizophrenia and its clinical dimensions: an association study in a South American population isolate. *Human Heredity, 67*(3), 163-173.

Le-Niculescu, H., Balaraman, Y., Patel, S., Tan, J., Sidhu, K., Jerome, R.E., et al. (2007). Towards understanding the schizophrenia code: An expanded convergent functional genomics approach. *American Journal of Medical Genetics Part B, 144B*(2), 129-158.

Lieberman J.A. et al. (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine*, *353*, 1209-1223

Mah, S., Nelson, M.R., Delisi, L.E., Reneland, R.H., Markward, N., James, M.R. et al. (2006). Identification of the semaphoring receptor PLXNA2 as a candidate susceptibility to schizophrenia. Molecular Psychiatry, 11, 471-278.

Martin, E.R., Monks, S.A., Warren, L.L., Kaplan, N.L. (2000). A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *American Journal of Human Genetics, 67,* 146-154.

McCurdy, R.D., Feron, F., Perry, C., Chant, D.C., McLean, D., Matigian, N., et al. (2006). Cell cycle alterations in biopsied olfactory neuroepithelium in schizophrenia and bipolar I disorder using cell culture and gene expression analyses. *Schizophrenia Research, 82*(2-3), 163-173.

McGrath, J., Saha, S., Welham, J., El Saadi, O., MacCauley, C., & Chant, D. (2004). A systematic review of the incidence of schizophrenia: The distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Medicine, 2*, 13.

Millar, J.K., Mackie, S., Clapcote, S.J., Murdoch, H., Pickard, B.S., Christie, S. et al. (2007). Disrupted in schizophrenia 1 and phosphodiesterase 4B: Towards an understanding of psychiatric illness. *Journal of Physiology, 584*(Pt 2), 401-405.

Murray, R.M., & Dean, K. (2008). Schizophrenia and related disorders. In Murray, R.M., Kendler, K.S., McGuffin, P., Wessely, S., & Castle, D.J. (Ed.), *Essential psychiatry* (pp. 284-319). New York: Cambridge University Press.

National Human Genome Research Institute. (2009). *Genome-wide Association Studies.* Retrieved March 10, 2009, from National Human Genome Research Institute Web Site: http://www.genome.gov/20019523

O'Conner, V., Houtman, S.H., De Zeeuw, C.I., Bliss, T.V., & French, P.J. (2004). Eml5, a novel WD40 domain protein expressed in rat brain. *Gene, 336*(1), 127-137.

O'Donovan, M.C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V. et al. (2008). Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nature Genetics, 40*, 1053-1055.

Petryshen, T.L., Middleton, F.A., Kirby, A., Aldinger, K.A., Purcell, S., Tahl, A.R. et al. (2005). Support for involvement of neuregulin 1 in schizophrenia pathophysiology. *Molecular Psychiatry, 10*(4), 366-374, 328.

Petryshen, T.L., Middleton, F.A., Tahl, A.R., Rockwell, G.N., Purcell, S., Aldinger, K.A., et al. (2005). Genetic investigation of chromosome 5q GABAA receptor subunit genes in schizophrenia. *Molecular Psychiatry, 10*, 1074-1088.

Pimm, J., McQuillin, A., Thirumalai, S., Lawrence, J., Quested, D., Bass, N., et al. (2005). The epsin 4 gene on chromosome 5q, which encodes the clathrin-associated protein enthoprotin, is involved in the genetic susceptibility to schizophrenia. American Journal of Human Genetics, 76, 902-907.

Pompili, M., Amador, X.F., Girardi, P., Harkavy-Friedman, J., Harrow, M., Kaplan, K. et al. (2007). Suicide risk in schizophrenia: Learning from the past to change the future. *Annals of General Psychiatry, 6*, 10.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., et al. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics, 81*(3), 559-575.

Purcell, S. (2008). PLINK (version 1.05). Retrieved from PLINK Web Site: http://pngu.mgh.harvard.edu/purcell/plink/

Puri, V., McQuillin, A., Thirumalai, S., Lawrence, J., Krasucki, R., Choudhury, K., et al. (2006). Failure to confirm allelic association between markers at the CAPON gene locus and schizophrenia in a British sample. *Biological Psychiatry, 59*(2), 195-197.

Riley, B., & Kendler, K.S. (2006). Molecular genetic studies of schizophrenia. *European Journal of Human Genetics, 14*(6), 669-680.

Sarparanta, J. (2008). Biology of myospryn: what's known? Journal of Muscle Research and Cell Motility, 29(6-8), 177-180.

Shiftman, S., Johannesson, M., Bronstein, M., Chen, S.X., Colllier, D.A., Craddock, N.J. et al. (2008). Genome-wide association identifieds a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLos Genetics, 4,* e28.

Sodhi, M., Wood, K.H., Meador-Woodruff, J. (2008). Role of glutamate in schizophrenia: Integrating excitatory avenues of research. *Expert Review of Neurotherapeutics, 8*(9), 1389-1406.

Spielman, R.S., & Ewens, W.J. (1996). The TDT and other family-based tests for linkage disequilibrium and association. *American Journal of Human Genetics, 59*(5), 983-989.

Sun, J., Kuo, P., Riley, B.P., Kendler, K.S., & Zhao, Z. (2008). Candidate genes for schizophrenia: A survey of association studies and gene ranking. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics), 147B*, 1173-1181.

The International Schizophrenia Consortium (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, Jul 1, Epub ahead of print.

The Psychiatric GWAS Consortium Steering Committee (2009). A framework for interpreting genome-wide association studies of psychiatric disorder. *Molecular Psychiatry, 14*, 10-17.

van den Oorde, E.J., Jiang, Y., Riley, B.P., Kendler, K.S., & Chen, X. (2003). FP-TDI SNP scoring by manual and statistical procedures: a study of error rates and types. *Biotechniques, 34*, 610-620.

Wigginton, J.E., Cutler, D.J., Abecasis, G.R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics,76*, 887-893.

Williams, H.J., Owen, M.J., O'Donovan, M.C. (2009). New findings from genetic association studies of schizophrenia. *Journal of Human Genetics, 54*, 9-14.

Wratten, N.S., Memoli, H., Huang, Y., Dulencin, A.M., Matteson, P.G., Cornacchia, M.A., et al. (2009). Identification of a schizophrenia-associated functional noncoding variant in NOS1AP. *American Journal of Psychiatry, 166*(4), 434-441.

Wu, E.Q., Birnbaum, H.G., Shi, L., Ball, D.E., Kessler, R.C., Moulis, M. et al. (2005). The economic burden of schizophrenia in the United States in 2002. *Journal of Clinical Psychiatry, 66*(9), 1122-1129.

Xu, B., Wratten, N., Charych, E.I., Buyske, S., Firestein, B.L., & Brzustowicz, L.M. (2005). Increased expressionin dorsolateral prefrontal cortex of CAPON in schizophrenia and bipolar disorder. *PLoS Medicine, 2*(10), e263.

Zeitlin, A.A., Heward, J.M., Brand, O.J., et al. (2007). Use of tag single nucleotide polymorphism (SNPs) to screen PTPN21: No association with Grave's disease. *Clinical Endocrinology (Oxf), 65*(3), 380-384.

Zheng, Y., Li, H., Qin, W., Chen, W., Duan, Y., Xiao, Y., et al. (2005). Association of the carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase gene with schizophrenia in the Chinese Han population. *Biochemical and Biophysical Research Communications, 328*(4), 809-815.

VITA

Grace Hyeiwon Lee was born on January 30th, 1984 in Floydada, Texas, the daughter of Sook Ran Lee and Jae Ki Lee. She lived in South Korea, Canada, and the United States during her childhood. After graduating from Curtis Senior High School in University Place, Washington in 2002, she went on to the University of Virginia in Charlottesville, Virginia where she studied biology and psychology and received her Bachelor of Arts in May 2006. After working as a research assistant for a year in an infectious disease lab in the University of Virginia, she enrolled in graduate studies at Virginia Commonwealth University in 2007. She received a Masters in Science in Physiology in August 2009.