12-2011

Genome-Wide Association Analysis of Major Depressive Disorder and Its Related Phenotypes.

Nagesh Ramarao Aragam
East Tennessee State University

Genome-Wide Association Analysis of Major Depressive Disorder and Its Related Phenotypes

---

A dissertation
presented to
the faculty of College of Public Health
East Tennessee State University

In partial fulfillment
of the requirements for the degree
Doctor of Public Health

---

by
Nagesh R. Aragam
December 2011

Dr. James L. Anderson, Committee Chair
Dr. Kesheng Wang, Research Mentor
Dr. Xuefeng Liu, Committee Member

Key Words: Major Depression Disorder, Genome-Wide Association Study, Gene x Gender Interactions, Novel Genetic Variants, Neuroticism, Age at Onset

ABSTRACT

Genome-Wide Association Analysis of Major Depressive Disorder and Its Related Phenotypes

by

Nagesh Aragam

Major Depressive Disorder (MDD) is a complex and chronic disease that ranks fourth as cause of disability worldwide. Thirteen to 14 million adults in the U.S. are believed to have MDD and an estimated 75% attempt suicide making MDD a major public health problem. Recently several genome-wide association (GWA) studies of MDD have been reported; however, few GWA studies focus on the analysis for MDD related phenotypes such as neuroticism and age at onset of MDD. The purpose of this study is to determine risk factors for MDD, identify genome-wide genetic variants affecting neuroticism and age at onset as quantitative traits, and detect gender differences influencing neuroticism.

Bivariate and multiple logistic regression analyses were performed on 1,738 MDD cases and 1,618 non-MDD controls to determine phenotypic risk factors for MDD. Multiple linear regression analyses in PLINK software were used for GWA analyses for neuroticism and age at onset of MDD with 437,547 Single Nucleotide Polymorphisms (SNPs).

Gender (OR: 1.43; 95% CI: 1.24 – 1.64) and a family history (OR: 2.88; 95% CI: 2.48 – 3.35) were significantly associated with an increased risk of MDD, which supports the findings of prior studies. Through GWA analysis 34 SNPs were identified to be associated with neuroticism ($p < 10^{-4}$). The best SNP was rs4806846 within the TMPRSS9 gene ($p = 7.79 \times 10^{-6}$). Furthermore, 46 SNPs were found showing significant gene x gender interactions for

neuroticism with $p<10^{-4}$. The best SNP showing gene x gender interaction was rs2430132 ($p = 5.37x10^{-6}$) in HMCN1 gene. In addition, GWA analysis showed that several SNPs within 4 genes (GPR143, ASS1P4, MXRA5 and MAGEC1/2) were significantly associated with age at onset of MDD ($p < 5x10^{-7}$).

This study confirmed previous findings that MDD is associated with an increased prevalence in women (about 43% more compared to men) and is highly heritable among first degree relatives. Several novel genetic loci were identified to be associated with neuroticism and age at onset. Gender differences were found in genetic influence of neuroticism. These findings offer the potential for new insights into the pathogenesis of MDD.

DEDICATION

I dedicate this work to my dearest wife Cindy, whose encouragement, sacrifice, and support made my huge career leap from engineering and computer science to public health and medicine possible.

ACKNOWLEDGEMENTS

Funding support for Major Depression: Stage 1 Genomewide Association in Population-Based Samples was provided by NWO: genetic basis of anxiety and depression *(904-61-090)*; resolving cause and effect in the association between exercise and well-being *(904-61-193)*; twin-family database for behavior genomics studies *(480-04-004)*; twin research focusing on behavior (*400-05-717*), Center for Medical Systems Biology *(NWO Genomics)*; Spinozapremie *(SPI 56-464-14192)*; Centre for Neurogenomics and Cognitive Research (*CNCR-VU*); genomewide analyses of European twin and population cohorts *(EU/QLRT-2001-01254)*; genome scan for neuroticism *(NIMH R01 MH059160)*; Geestkracht program of ZonMW (*10-000-1002*); matching funds from universities and mental health care institutes involved in NESDA (GGZ Buitenamstel-Geestgronden, Rivierduinen, University Medical Center Groningen, GGZ Lentis, GGZ Friesland, GGZ Drenthe). Major funding for this project is from the Genetic Association Information Network of the Foundation for the US National Institutes of Health, a public–private partnership between the NIH and Pfizer Inc., Affymetrix Inc. and Abbott Laboratories. The genotyping of samples was provided through the Genetic Association Information Network (GAIN). The dataset(s) used for the analyses described in this manuscript were obtained from the database of Genotype and Phenotype (dbGaP) found at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000020.v2.p1. Samples and associated phenotype data for the Whole Genome Association Study of Bipolar Disorder were provided by Dr**.** Patrick F. Sullivan.

Dr. Wang, my research mentor, has been a constant source of ideas, encouragement, and thoughtful criticisms and a patient teacher of complex issues in population genetics. He has been a great resource for me to delve into the area of population genetics from the very first day I met

him. He has given me confidence to enter an exciting area to conduct research for my dissertation and has provided me with all the resources, tools, and guidance for not only this dissertation work, but also my attempt at F31 grant writing and coauthoring many of the research papers he has published in the past two years in genetic epidemiology. Thank you very much!

I am very thankful to my dissertation committee chair Dr. James Anderson who introduced me to Epidemiology. He has guided my progress of understanding the subject matter with patience and persistent encouragement despite his very busy work schedules, so many other graduate students to guide and department commitments. I will always be grateful and indebted to Dr. Anderson.

Dr. Xuefeng Liu has been my thesis committee member and my teacher of Biostatistics with a cheerful attitude and an inviting openness to answer my technical questions, clarify my understanding of complex facts, and support my dissertation research work throughout. Thank you very much!

I am indebted to Dr. Anderson, Dr. Wang, and Dr. Liu for reviewing many drafts of this dissertation and offering numerous constructive criticisms to make it better.

From the very first idea of switching from my fairly successful career of engineering and computer science into public health and medicine, my family has been solidly supportive and my eternal gratitude goes out to my parents-in-law Dr. Dennis Hamm and Joyce Hamm for opening up their hearts and home for my pursuit of the unknown. This dissertation work and all the other training and education I have received in public health, hospital work experience, and years of heartfelt and heart-warming family life at home have been their greatest gift to me. Thank you two very much from the bottom of my heart.

My wife Cindy, my sons Nevin and Bryon, and my two wonderful step-daughters April and Kiri have shared and endured my journey with laughter, apprehension, confusion, and perhaps even some embarrassment ("Will dad ever get out of school?"). They deserve my utmost thanks for their constant support, love and understanding.

I would like to thank many of my peers and friends (too many names to mention) at ETSU, Johnson City Medical Center, and Mountain States Health Alliance for offering me help, teaching me public health, medical and nursing concepts, sharing many hours of classroom learning and camaraderie. I would also like to thank many of my students in Research Methods and Biostatistics II for letting me solidify my understanding of the subject material through teaching and tutoring. Thank you all very much for your friendship, love, and support.

I am grateful to the College of Public Health and the School of Graduate Studies at the East Tennessee State University for providing me the financial assistance through Graduate Assistantship throughout this research work. My special thanks go out to Dr. James Florence (for having the confidence in an engineer to be admitted into the DrPH program), Dr. Robert Pack to encourage and work with me for my NIH F31 grant application, and Rickie Carter for being my best friend and catering to all my administrative needs. Many thanks to all of you!

Finally, I would like to acknowledge the leadership of Dr. Randy Wykoff, Dean of the College of Public Health for fostering an excellent research environment to learn and work in.

CONTENTS

Page

LIST OF TABLES

Tables Page

LIST OF FIGURES

Figures Page

CHAPTER 1

INTRODUCTION

Major depressive disorder (MDD) is a chronic disease whose epidemiology has been studied worldwide and clinically researched to understand its etiology, distribution, development, and available treatments. MDD is characterized by its defining features of marked and persistent depressive moods, appetite and weight changes, insomnia or hyperinsomnia, fatigue or loss of energy, feelings of excessive guilt or worthlessness, poor concentration and/or indecisiveness, recurrent thoughts of suicide or death, and psychomotor agitation or retardation (Boomsma et al., 2008). MDD is distinct from normal sadness by its persistence for longer than 2 weeks, drug or alcohol dependence, and somatic diseases (Sullivan et al., 2009). The definition of MDD excludes other mental disorders such as, bipolar disorder, schizophrenia, and other schizoaffective disorders. MDD has been found to be comorbid with nicotine dependence (Cardenas et al., 2002), alcoholism (Miguel-Hidalgo et al., 2010), anxiety (Boomsma et al., 2000), and other psychological problems. Those afflicted with MDD may suffer from the disease throughout their lifetime. An estimated 75% of MDD afflicted people suffer from its chronic status and have an increased risk of suicide.

MDD is a complex disease that affects the lives of both the patients and their families. It is a major cause of disability worldwide, ranking 4th according to a recent survey of global burden of disease; and estimated to become 2nd by 2030 (Mathers & Loncar, 2006). The prevalence estimate for MDD in the United States National Comorbidity Survey - Replication (NCS-R) sample is 6.7% for a 12-month period which is an estimated projection of 13.1 to 14.2 million U.S. adults suffering from MDD during a 12-month period (Kessler et al., 2005). It has been firmly established that the risk for MDD is partly genetic and family studies have found a

significantly higher prevalence of MDD among biological relatives of MDD cases (Sullivan, Neale, & Kendler, 2000). Despite the existence of a considerable amount of research on the epidemiology and the biological correlates of MDD, the etiology of MDD remains unknown. An important clue has been the familial tendency and the heritability of MDD which has led to a number of genome-wide linkage studies with the identification of over 100 candidate genes (Sullivan et al., 2008).

Significance

Genetic factors are important in the risk of MDD and heritability has been estimated at 35%-40% (Kendler & Prescott 1999, Sullivan et al., 2000). MDD is frequently associated with smoking (Husky et al., 2008), anxiety (Boomsma et al., 2000), marital problems (South & Krueger 2008), and other psychological problems (Wade, Bulik, Neale, & Kendler, 2000). Recent studies of MDD at the molecular and genetic levels provide increasing evidence linking disease phenotypes with both genes and environment factors. Jabbi and coworkers have reviewed data that supports a role of monoaminergic and other related genes in environmental adaptation to conclude that convergent approaches may be useful in the examination of genetic modulation of disease phenotypes (Jabbi, Korf, Ormel, Kema, & den Boer, 2008). The current literature reveals that more sophisticated tools and methods are needed to understand the etiology, prevalence, and genetic significance of MDD (The Psychiatric GWAS Consortium Steering Committee, 2009).

Morbidity and mortality from MDD are significant public health problems as highlighted by two Surgeon General Reports (U. S. Department of Health and Human Services, 1999; U. S. Public Health Service, 1999). The Surgeon General's Reports point out that suicide, a tragic

consequence of MDD occurs in 10%-15% of patients previously hospitalized for depression (Angst, Angst, & Stassen, 1999). This mortality is three times greater than that reported for the general United States population (U. S. Public Health Service, 1999).

MDD is a highly heritable medical disorder and it is clear that the genetic etiology of MDD is complex and multi-factorial. Studies have shown that gender is important to the course and outcome of MDD. Silverstein (1999) described gender differences in the prevalence of clinical depression and Lavretsky et al. (2004) reported on sex differences in brain structure in geriatric depression. The gender-specific rate of MDD found in community samples agrees with the 1.7:1 prevalence of MDD reported for women vs. men (Marcus et al., 2005). Recently, Essau, Lewinsohn, Seeley, and Sasagawa (2010) studied gender differences in the developmental course of depression and found strong evidence for a female preponderance of MDD in adolescence and adulthood. Immunoreactivity of cortical receptor proteins (NUDR and 5-HT_{1A}) found in female subjects with MDD by Szewczyk et al. (2009) was lower compared to their male cohorts.

Genetic and environmental factors may interact in their contribution to the risk of clinical depression such as Major Depressive Disorder in adolescents and adults. As stated above, MDD frequently coexists with smoking, anxiety, and other psychological problems. There has been little effort to examine gender and genetic risks together. Genetic and gender interaction may explain previously observed discrepancies between studies looking at candidate environmental agents and genes for MDD. Without accounting for these interactions the true main effects of either the gender factor or the gene will not be identified. One strategy for increasing the statistical power to detect MDD genes is to divide the phenotype into genetically meaningful subtypes (also called endophenotypes) and use a sufficiently large sample size of the population under study to decrease heterogeneity.

Another strategy was proposed by Risch and Merikangas (1996) in their analysis of the statistical power of linkage and association studies. By having a large number of genome-wide Single Nucleotide Polymorphisms (SNPs) a case-control association study may be conducted to associate specific SNPs and genomic regions with complex disease traits. Back in 1996, the human genome had not yet been sequenced, but the hypothesis of Risch and Merikangas laid the foundation for the genome-wide association studies that became a reality after 2004 when the functional human genome was fully sequenced.

Identification of specific genes and gene x gender interactions may point to identifying specific population subgroups at increased risks for MDD. The findings have the potential to help with understanding the genetic associations, predicting the risk, and developing treatments for MDD. Increased understanding of how genetic factors and gender interact to alter the risks for depression may allow us to provide new targets for therapy. Identifying genes that may control the AAO of MDD would contribute to the understanding of MDD development and progression and allow therapeutic interventions for delaying the onset of MDD for specific families and population subgroups.

Research Purpose

The purpose of this research is to conduct Genome-Wide Association (GWA) analyses of MDD affection, related phenotypes (e.g., neuroticism, anxiety), and age at onset (AAO) of MDD.

Research Questions

This study addresses the following three research questions:

1. Are gender and family history significantly associated with an increased risk of major depressive disorder (MDD)?
2. Is neuroticism (a measurable, quantitative trait) a useful indicator of major depressive disorder and are there any gene x gender interactions present?
3. Do the genetic components of age at onset help us to understand development of major depressive disorder?

Research Hypotheses

Hypothesis 1

Being female, smoking, and family history of MDD are three of the strongest risk factors for the development of MDD in the sample of European population from the Netherlands.

Rationale. With the availability of phenotypic information from a sizeable number of MDD cases and non-MDD controls with 1,150 men and 2,206 women, carrying out statistically meaningful linear and logistic regression analyses with a binary outcome as MDD affection for multiple logistic regression analysis and a continuous quantitative outcome for multiple linear regression analysis must be straightforward. In addition, with data available for a total of 10 predictors for depression, meaningful and strong asymptotic associations from the linear and logistic regression would pare down the number of possible risk factors for MDD.

Hypothesis 2

Neuroticism as a continuous variable may increase statistical power of genome-wide association analysis, and gender differences may exist in the genome-wide analysis of neuroticism.

Rationale. Neuroticism as an endophenotype for Major Depressive Disorder (MDD) is a well-known, measurable, genetic, and environment-dependent mental condition. While neuroticism has been recognized as one of the endophenotypes of MDD, few genome-wide analyses of neuroticism as a quantitative trait have been reported to date. Neuroticism, as has been conceptualized by Eysenck and Eysenck (1975), Hirschfield et al. (1983), and Pervin and John (1990), is a personality trait that reflects emotional instability, vulnerability to stress, and being anxiety prone. Neuroticism is considered as a moderate risk factor for major depression disorder (MDD) and as a quantitative personality trait, neuroticism is moderately heritable (Calboli et al., 2010). According to Bienvenu et al. (2001) neuroticism is the strongest predictor of comorbidity of MDD.

Hypothesis 3

Age at onset as a quantitative trait may aid the identification of novel genetic variants for the development of MDD.

Rationale. Although some genome-wide association studies of major depression disorder and related comorbidity traits such as smoking, alcoholism, and neuroticism have been reported, few studies have focused on the age at onset (AAO) of MDD. Luby (2009) asserts that empirical evidence for clinical depression in children as young as age 3 has recently been validated. Other studies have reported Traumatic Brain Injury (TBI) at an early age has led to higher rates of major depression within 3 months of injury (Rao et al., 2010). Kovacs and Lopez-Duran (2010) report that a compelling body of literature indicates that depressive symptoms in youngsters predict subsequent MDD. Also, a Turkish study (Bilgi et al., 2010) suggests small frontal gray matter volume leading to first-episode depression. However, few genome-wide studies in

literature have reported on the analysis of age at onset of major depression.

CHAPTER 2

LITERATURE REVIEW

Major depressive disorder (MDD) is a widespread and common mental condition that affects persons of all cultures, races, genders, and ages (Millon, 2004). Both genetic and environmental factors influence the occurrence of MDD. MDD has been diagnosed using several approaches and instruments, and these are constantly undergoing revisions and refinements (Kessler et al., 2009).

The Diagnostic Interview Schedule (DIS) was one of the first instruments used to epidemiologically study MDD (epidemiology catchment area (ECA) study by Kessler et al., 1994). Currently, a standard instrument for measuring MDD is the World Mental Health (WMH) Survey Initiative Version of the World Health Organization's (WHO) Composite International Diagnostic Interview (CIDI) (Kessler & Ustun, 2000). The first nationally representative survey in the U.S. was conducted from 1990 to 1992 using a modified version of CIDI and a survey method (using face-to-face interviews) similar to ECA. This survey is known as the National Comorbidity Survey (NCS). A revised version of this survey (NCS-R) based on International Classification of Diseases, 10th Revision (ICD-10) and Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) was reported by Kessler et al. using data from face-to-face interviews by professional interviewers from the Institute for Social Research at the University of Michigan, Ann Arbor, between February 2001 and April 2003 (Kessler et al., 2005). The diagnostic criteria used by instruments like CIDI are based on the DSM-IV published by the American Psychiatric Association (APA). DSM diagnostic criteria are actively undergoing revisions (an updated DSM-V was recently announced by APA) (Regier, 2009) and is expanding to include newer diagnostic elements for such conditions as autism. In their study,

"Epidemiology of major depressive disorder: Results from the NCS-R", Kessler et al. report that the prevalence estimates for CIDI/DSM-IV MDD in the total NCS-R sample (N = 9282) are 16.2% (95% CI, 15.1 – 17.3) for lifetime and 6.6% (95% CI, 5.9 – 7.3) for the 12 months before the interview (Kessler et al., 2003).

Further details of lifetime prevalence of Major Depressive Disorder and Age at Onset distributions as determined by the Kessler et al. (2003) study are given in Appendix A1 for the interested reader.

Major Depressive Disorder and Related Phenotypes

Kessler and Ustun (2000) classify mental disorders under five categories as follows –

1. Mood, which includes major depression, neuroticism, bipolar disorder, and mania.
2. Anxiety, which includes panic disorder, phobias like agoraphobia, generalized anxiety disorder, posttraumatic stress disorder (PTSD), and obsessive-compulsive disorder (OCD).
3. Substance abuse, which includes alcohol abuse and dependence and nicotine addiction.
4. Childhood-related, which includes attention-deficit hyperactivity disorder (ADHD), conduct disorder, separation anxiety disorder, and oppositional-defiant disorder.
5. Others, which include eating disorders, premenstrual disorder, intermittent explosive disorder, pathological gambling, etc.

Specific details regarding the diagnosis of the above mentioned mental disorders including screening for signs and symptoms, discussion of lifetime reviews, risk factors, treatment, etc. are discussed further in Kessler and Ustun (2000). The complete WMH-CIDI includes a screening

module, 40 modules focusing on diagnoses, functioning, treatment, risk factors, socio-demographic correlates, and methodological factors. In addition, elaborate CDROM-based training materials are available to teach interviewers to administer the tests and to teach the administrators how to ensure the quality of data collected. An overview of the WMH-CIDI development is presented in Appendix A2.

As is consistent for other complex disease traits, MDD shows strong evidence of heritability; however, identification of the causal genes for MDD has not yet been successful. Results from genome-wide linkage scans for MDD and related quantifiable personality traits show a relationship with "harm avoidance" and "neuroticism". Regions on chromosomes 1, 4, 5, 7, 8, 11, 12, and 13 show significant linkage signals, but, these findings have not been replicated among other populations (Boomsma et al., 2008). Neuroticism, Index of Depressive Symptomatology, and Beck's Anxiety Disorder Inventory have all been found to be quantitative traits for MDD. Neuroticism is a personality trait that reflects emotional instability, vulnerability to stress, and being anxiety prone (Eysenck & Eysenck, 1975; Hirschfield, Klerman, Clayton, & Keller, 1983; Pervin & John, 1990). Neuroticism is moderately heritable and is considered to be a moderate risk factor for MDD (Calboli et al. 2010) and is the strongest comorbid predictor of MDD (Bienvenu et al., 2001). Fanous, Gardner, Prescott, Cancro, and Kendler (2002) have examined the genetic and environmental sources of covariation between neuroticism and MDD but were not able to identify a significant gender difference. Hettema et al. (2006) have sought to identify additional etiologic factors that contribute to the comorbidity of neuroticism with MDD. Their findings show substantial, if not complete, overlap between the genetic factors that influence individual variation in neuroticism and those that increase liability across the internalizing disorders.

Major Depressive Disorder Affection as a Binary Outcome

In this study, MDD is considered as a binary outcome (i.e., Affected or Not Affected) that has related phenotypes (or risk factors) that are both categorical (e.g., married or single or widowed; smoking or non-smoking) and continuous (e.g., neuroticism and age at onset).

Risk Factors of Major Depressive Disorder

The phenotypes, gender, marital status (living with a partner is considered as being married), smoking status, alcohol use, and family history of depression are either binary or categorical and require little further explanation. Age and Age at Onset are continuous variables taking values of 10-90 years depending on the specific population being used in a study.

The phenotypes determining the traits of neuroticism, anxiety severity, depression symptomatology, and age at onset are further described as follows. Neuroticism is considered as a personality trait associated with several mental disorders and is considered as a risk factor for the development of MDD, anxiety disorders, and dementia (Calboli et al., 2010). A widely used measure for neuroticism is the NEO personality inventory (Costa & McCrae, 1992) which measures the following five dimensions of personality – Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness (also collectively known as NEO-Five Factor Inventory or NEO-FFI). A score based on the subject's response to a 5-point scale ('strongly disagree' to 'strongly agree') for each of the dimensions of personality in a 60-item questionnaire is totaled to give a total score called the NEO-FFI total score (Guerrera et al., 2005). This continuous variable is then used to determine along with other predictor variables (e.g., Inventory of Depressive Symptomatology and Beck's Anxiety Inventory scores) a binary predictor variable that is used as

the outcome - the MDD affliction status (i.e., affected / not affected) for use in logistic regression analyses of MDD.

The inventory of depressive symptomatology (IDS) is a 30-item questionnaire designed to assess (and often quantify the severity of) depressive symptoms. The assessment uses all the criterion symptom domains designated by the American Psychiatry Association (APA) Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (APA, 1994). The time frame for assessing the symptom severity is 7 days prior to the day of assessment. The Beck Anxiety Inventory (BAI) is a 21-question multiple choice inventory created by Aaron T. Beck used for measuring the severity of an individual's anxiety level (Leyfer, Woodruff-Borden, Klein-Tasman, Fricke, & Mervis, 2006). Each question in the inventory has one answer out of four ('not at all', 'mildly', 'moderately', and 'severely') with a point system coding each category. For example, the answer 'not at all' scores 0-7 points, while the answer 'severely' scores 26-63.

Other MDD related phenotypes such as Age at Onset (AAO) of MDD, Anxiety Severity, Depression Severity, and Neuroticism are defined below and their descriptive characteristics are described later in chapter 3.

- *Age at onset* is defined by Composite Interview Diagnostic Interview (CIDI) (Kessler & Ustun, 2000) as the age when the first of the 2 years or longer a subject felt sad or depressed.
- *Depression Severity* is defined by the Inventory of Depressive Symptoms as a total score computed by summing the 28 items symptom list in the CIDI interview.
- *Anxiety Severity* is defined by the Beck Anxiety Inventory which is a total score computed by summing the 21 items symptom list in the CIDI interview.

Brief Survey of Major Depressive Disorder Phenotypic Studies

The psychiatric and psychopathologic literature is abundant with phenotypic studies of MDD focusing on gender difference (e.g., Goodwin & Gotlib, 2004; Hakamata et al., 2009; Hyde, Mezulis, and Abramson, 2008; Parker & Hadzi-Pavlovic, 2004; Silverstein, 1999), marital problems and/or partner violence (e.g., Shorey et al., 2011; South & Krueger, 2008), smoking status (e.g., Cardenas et al., 2002; Gulec et al., 2005), family history (Lazary, Gonda, Benko, Gacser, & Bagdy, 2009), and alcohol use (e.g., Miguel-Hidalgo et al., 2010). Each of these MDD phenotypes has been researched well, and the following subsections briefly summarize the previous work.

The reported heritability of neuroticism is equal or greater than heritability estimates for MDD (Calboli et al., 2010). The heritability of neuroticism has been estimated at 0.30 – 0.50 based on twin studies (Birley et al., 2006; Bouchard & Loehlin, 2001; Floderus-Myrhed, Pederson and Rasmuson, 1980; Jang, Livesley & Vernon, 1996; Martin et al., 2000) along with genetic covariance with depression and anxiety (Jardine, Martin, & Henderson, 1984; Fanous et al., 2002; Hettema, Prescott, & Kendler, 2004; Hettema, Neale, Myers, Prescott, & Kendler, 2006; Huezo-Diaz, Tandon, & Aitchison, 2005; Kendler, Neale, Kessler, Heath, & Eaves, 1992, 1993). A link between genotype and personality has been tried to be established by several researchers using NEO PI-R facets. For example, Persson et al. (2000) have shown an association between a polymorphism in the tyrosine hydroxylase gene and personality traits, but, Tochigi et al. (2006) report that they could not confirm the association with the tyrosine hydroxylase gene. However, in a classic study by Lesch et al. (1996) a relationship between the serotonin transporter gene regulatory region (5-HTTLPR) and neuroticism has been found. Furthermore, it was found that individuals with a shorter allele version of the gene had higher

neuroticism scores and was significant for both heterozygous and homozygous versions of the allele. The authors further predict that approximately 10 to 15 genes in the 5-HTTLPR region might be implicated in the moderation of neuroticism scores (Lesch et al., 1996).

Gender Difference and Major Depressive Disorder

The importance of gender on the course and outcome in MDD has been widely acknowledged. For example, Silverstein (1999) described the gender difference in the prevalence of clinical depression, while Lavretsky et al. (2004) reported sex differences in brain structure in geriatric depression. The gender-specific rate of MDD is proportional to rates found in community samples with a 1.7:1 prevalence of MDD in women vs. men (Marcus et al., 2005). Recently Esau et al. (2010) studied the gender differences in the developmental course of depression. Especially, gender-specific associations of NUDR and 5-HT_{1A} receptor proteins with MDD were reported by Szewczyk et al. (2009).

Hyde et al. (2008) in trying to find an integrated model of affective (emotional reactivity), biological (genetic vulnerability, pubertal hormone, timing, and development) and cognitive (cognitive style, objectified body consciousness, and rumination) have summarized the state-of-the-art in gender differences for MDD. Accordingly, several factors can be attributed to the higher prevalence of MDD in girls than boys and in women than men as follows –

- Greater ruminative coping (Nolen-Hoeksema & Girgus, 1994)
- Dependence on relationships or affiliative needs (Cyranowski, Frank, Young, & Shear, 2000)
- Ovarian and adrenal hormonal changes in puberty (Goodyer, Herbert, Tamplin, & Altham, 2000)
- Genetic factors (Kendler et al., 1993; Zubenko et al., 2002)

- Body dissatisfaction (Nolen-Hoeksema & Girgus, 1994)
- Greater cognitive vulnerability (Hankin & Abramson, 2001)
- Exposure to negative life events like rape and child sexual abuse (Kendler, Gardner, & Prescott, 2002)
- Gender intensification and adherence to traditional gender roles (Aube, Fichman, Saltaris, & Koestner, 2000)
- Interactions among the above listed factors (Hankin & Abramson, 2001).

While an in-depth summary of each of the above study is beyond the scope of this report, suffice is it to say that the female gender has specific biological and cognitive factors that are unique to the gender and thus the manifestation of higher prevalence of MDD in females does not seem to be unnatural. In particular, the social factors like body image, traditional role playing (this would especially be true in Asian and Mid-eastern cultures), and lack of access to equal rights (e.g., to education) may breed gender differences right from the moment of birth for girls and women.

According to Essau et al. (2010) the greatest increase in gender difference occurs between the ages of 15 and 18 and the underlying mechanisms for this are not clear. It seems to generally reflect the interplay of gender socialization, hormonal changes, and stressful life events associated with adolescence (Cyranowski et al., 2000). Thus Essau and coworkers who studied the developmental course of MDD in a project (Oregon Adolescent Depression Project) with 773 participants concluded that childhood depression may be a more serious risk factor for girls than boys.

More interestingly however, Szewczyk et al. (2009) point to some basic biological reasoning in terms of the gender-specific alterations in cortical NUDR protein receptor gene. Female

subjects with MDD in their study had reduced protein expression of NUDR and 5-HT_{1A} receptors in the prefrontal cortex. Thus, they attribute biological factors for higher prevalence of MDD in females.

Thus it seems that both social and biological factors play a big role in understanding the gender differences and prevalence of MDD in girls and women. This area of understanding the 1.7:1 prevalence of MDD in women has become a prime target with researchers in psychology, psychiatry, molecular biology, and epidemiology.

Smoking Status and Major Depressive Disorder

Husky et al. (2008) conducted an epidemiological study to determine whether smoking behavior was associated with MDD, and further if the association was greater in women. They used data from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). The details of data collection and the sample are published elsewhere (Grant et al., 2003). Relationships between smoking status (categorized as daily, occasional, prior) and DSM-IV MDD by gender were assessed in terms of odds ratios using logistic regressions. The results showed that women with prior smoking were at significantly higher risk of current and past MDD than men (OR: 1.53 vs 1.36; 1.72 vs 1.36). Similarly, the results for current occasional – the ORs were 1.92 vs 1.39; 1.9 vs 1.3; and for daily smoking – the ORs were 2.52 vs 1.95; 1.84 vs 1.48.

MDD and nicotine dependence are highly comorbid (Cardenas et al., 2002). It seems that the MDD patients may use nicotine to ameliorate their depressive symptoms. Cardenas et al. hypothesized that a dysfunctional brain reward system (BRS) might link the use of nicotine and MDD neurobiologically to enhance the dopaminergic activity. So they conducted a double-blind case-control study with 34 subjects (18 nicotine-dependent cases and 16 nicotine-independent

controls) to assess their hypothesis. The results were mixed in the sense that the nicotine (or smoking) did not modify the response to d-amphetamine in either cases or controls, but it decreased the overall negative mood state during placebo sessions. The researchers concluded that although the BRS may be dysfunctional in MDD cases, chronic nicotine use does not modify response to d-amphetamine. In my personal opinion, the sample size (34) was too small to be able to conclude anything definitively.

In another cross-sectional study of association between smoking and depressive symptoms in Turkey the researchers found the smokers among medical students were 2.2 times more likely to have depressive symptoms than nonsmokers (Gulec et al., 2003).

Family History and Major Depressive Disorder

As stated earlier, MDD has been shown to have a heritability of almost 40% (Kendler & Prescott, 1999). MDD as an affective disorder has been shown to be both inherited and influenced by environmental factors such as living in a household with diagnosed MDD cases (Lazary et al., 2009). However, sufficient knowledge concerning inherited background and how the environmental factors moderate the onset of MDD is still unknown. Lazary and coworkers studied how affective temperaments (AT) and affective family history (AFH) relate to depressive symptoms in a general population. They used 501 Hungarian adults to measure and analyze the mediation of AT in AFH groups. Not surprisingly, a critical part of inherited factors of depression is mediated by affective temperaments. The probability of having dominant temperament was more than two-fold in the group with AFH than one without.

ATs are hypothesized to be associated with phenotypic expression of affective disorders (Whittle, Allen, Lubman, & Yucel, 2006) and their relevance to endophenotype studies has also

been proposed (Gonda et al., 2006). But much remains to be done in this area to fully understand the role of family history in the development of MDD and other depressive disorders.

Marital Problems and Major Depressive Disorder

South and Krueger (2008) opine that marital distress may exert influence on MDD by acting as a stressor. They conducted a study of marital quality on a range of mental disorders (including symptoms for MDD) in 379 twin pairs concordant for marriage. A phenotypic factor analysis was first conducted to confirm that one factor best accounted for the variance shared between MDD and other mental disorders. Then they investigated the overlap between genetic and environmental influences on both marital quality and internalizing spectrum to find genetic influences common to both phenotypes. Their conclusion was that those with a genetic predisposition to internalizing syndromes may be more likely to express the predisposition in context of a dissatisfying marriage.

Alcohol Use and Major Depressive Disorder

Miguel-Hidalgo et al. (2010) studied MDD and alcohol use in a more clinical context by measuring the levels of excitatory amino acid transporters (EAATs) 1 and 2 in a case-control experiment. Their results show that the EAAT2 immunoreactivity was significantly lower in MDD cases than in controls as were the levels of EAAT1. Their clinical conclusion is that there are differential changes in the expression of glial glutamatergic markers in depression and alcoholism suggesting a depletion of certain aspects of glutamatergic processing in MDD cases. Thus alcohol may influence MDD through complex processes in the prefrontal cortex that are still unknown.

Interestingly enough, Nurnberger, Jr. et al. (2001) have found evidence using genome-wide sibling-pair linkage analysis of comorbid alcoholism and MDD to conclude that certain genes

may influence both predisposition to alcoholism as well as detrimental diseases like MDD. They evidence a specific locus (or a genetic marker) on chromosome 1 that is responsible to this dual vulnerability. This was a part of the collaborative study on the genetics of alcoholism from the National Institute on Alcohol Abuse and Alcoholism and the data from this study were used to test three phenotypes – comorbid alcoholism and depression, alcoholism or depression, and just depression. It is also worth noting in the Nurnberger et al. study that a majority of subjects with alcoholism problem in this data set were men and a majority of subjects with depression were women. Thus, the study may have limitations with confounding effects along with interactions between alcoholism and depression.

However, some epidemiological studies have also shown researchers the positive effects of coping with MDD by using moderate alcohol consumption and the comorbidity of alcohol use and MDD needs to be researched further to elucidate our understanding of alcoholism as a predictor or risk factor of MDD.

Genetic Studies of Major Depressive Disorder and Related Phenotypes

Before the human genome was completely sequenced, genetic studies took the form of family, twin, linkage, and association studies that dealt with how genetic information was transmitted and inherited down through generations of affected patients of a disease to be studied. Most often these diseases were non-Mendelian, complex, and chronic diseases like Alzheimer's, Parkinson's, Major Depression, etc. The main thrust of these genetic studies was to figure out how diseases are carried through genetic markers and how studying the genes of immediate family members like twins, parents and children, and siblings might show patterns of disease linked genes in specific genetic regions and how they get inherited via strong linkages on

specific chromosomes. An excellent overview of these genetic studies is given in "Statistical Genetics: Gene Mapping through Linkage and Association" by Neale, Ferreira, Medland, and Posthuma (2009).

A brief review of genetic linkage and association studies of MDD and its phenotypes in the literature is given below. For detailed discussion of any one method or one phenotype linkage mapping across an individual family genome or inheritance patterns, the reader is referred to a plethora of information available in the PubMed. We discuss a sample of representative family studies, linkage studies, and association studies of MDD below.

Genetic Family studies of Major Depressive Disorder

To explore the continuities and discontinuities between MDD in children and adolescents and MDD in adults, Klein, Lewinsohn, Seeley, and Rhode (2001) use family studies. In their study of 268 adolescents with a history of MDD, 110 adolescents with non-MDD disorders, and 291 adolescents with no history of mental disorders they found evidence of familial aggregation of adolescent MDD. They also found considerable specificity in the pattern of familial transmission.

They used hazard ratios (HRs) to measure the increased risk for family members of a MDD afflicted patient and they found the following –

- Elevated rates of MDD (HR: 1.77, 95% CI: 1.46 – 2.31)
- Dysthymia (HR: 1.79, 95% CI: 1.11 – 2.87)
- Alcohol abuse or dependence (HR: 1.29, 95% CI: 1.05 – 1.53)

This pattern of MDD transmission within families was also found for other mental disorders like anxiety disorder, substance abuse, and antisocial behavior.

According to the literature available on family studies and MDD, it seems that the heritability poses almost a three-fold increase in risk (2.84) according Sullivan et al. (2000). However, it has also been established that families share MDD along with other phenotypic manifestations such as mood disorders (esp. bipolar), schizophrenia, and neuroticism. Golster-Dubner, Galili-Weistubb, and Segman (2010) criticized that a family-based association study of MDD may not pick up large genetic effects for MDD. They make the point that MDD coexists in families with other mental disorders like schizophrenia, bipolar disorder, and neuroticism, and thus any family association of MDD may be only small and overlapping with other mental disorder effects. In addition, two separate studies (Liu et al., 2009; Serretti et al., 2003) trying to link specific genes (BDNF) and specific enzymes (e.g. tyrosine hydroxylase) that are known to have a role in depression within a family failed to produce family-based associations. This might be due to the small genetic effects mentioned earlier for MDD and the fact that the specific etiology of family inheritance of MDD is still largely unknown.

Genetic Linkage Studies of Major Depressive Disorder

Genetic risk factors are well established for MDD and a twin study has indicated >70% heritability in twins (McGuffin, Katz, Watkins, & Rutherford, 1996). Genome-Wide linkage analysis was carried out by McGuffin et al. (2005) in a sample of 497 sib pairs concordant for recurrent MDD. The study found linkages on chromosomes 1, 12, and 13 at 1p36, 12q23.3 – q24.11, and 13q31.1 – q31.3 respectively, while implicating genes MTHFR and DAO. The 12q region has already been implicated with other mental disorders like schizophrenia and bipolar disorders and the 13q region has been linked to panic disorders. A previous report of a locus on 15q also showed genome-wide significance for recurrent depression and the current 12q findings are also genome-wide significant.

In 2003 Zubenko and his team of coworkers used the genome-wide linkage studies to report genetic loci influencing the recurrent early-onset MDD (RE-MDD) (Zubenko et al., 2003). This study found 392 highly informative polymorphisms with an average spacing of 9 cM in the region of the CREB1 gene. Nineteen chromosomal regions contained linkage peaks and 10 reached an adjusted p-value of < 0.001. Five loci showed evidence of interaction with the CREB1 locus, and the authors of the report concluded that genes whose products participate in cellular signaling pathways that converge on CREB region harbor alleles that affect the development of MDD.

Holmans et al. (2004) conducted a study of genome-wide significant linkage to RE-MDD on the region of chromosome 15q using 297 informative families containing 415 independent affected sibling pairs and 685 informative relative pairs. All affected cases were diagnosed before the person turned 31, and the affected relatives had a mean AAO of 41 years. Genome-Wide significant linkage was observed on chromosome 15q25.3-26.2 with an empirical genome-wide p-value of 0.023. These findings indicate chromosome 15 as a strong candidate for further studies of MDD susceptibility.

In a replication study Camp et al. (2005) used Utah pedigrees to identify loci that influenced RE-MDD and anxiety disorders. They used 87 large extended Utah pedigrees to investigate three phenotypes: RE-MDD and anxiety, RE-MDD or anxiety, and RE-MDD and anxiety. They replicated the earlier loci on 12q, 7p, and 18q and identified further interesting regions on 4q and 15q. Their study suggests some overlapping genetic etiologies between MDD and anxiety disorder.

Candidate Gene Association Studies of Major Depressive Disorder

Several candidate genes such as the CFH gene, the FTO gene, the TCF7L2 gene, and the IL23R gene have been shown to be associated with MDD (e.g., Boomsma et al., 2008). Increased risk for MDD has also been established in a study by Traks et al. (2008) which relates the MDD risk to IL20 and IL24 genes and suggest that cytokines may contribute to the pathogenesis of MDD. A recent study by Rietschel et al. (2008) has implicated the G72 gene with MDD and neuroticism in large population-based groups in Germany. More recently, Shyn and Hamilton (2010) have presented an intensive review.

In a classic paper by Cordell and Clayton (2005), the authors discuss the rationale behind genetic association studies. Traditional epidemiologic studies of environmental risk factors and genetic studies have a lot of similarities, but there are issues specific to the studies of genetic risk factors (e.g. use of specific family-based designs), population history, and underlying genetic mechanisms. Genetic association differs from genetic linkage in that the alleles of interest will be the same across the whole study population while linkage allows different alleles to be associated with the disease trait in different families.

Genetic associations can be direct or indirect, confounded or with interactions, consider direct relationship between genotype and phenotype, or it could be indirect relationship with the consideration of linkage disequilibrium. A brief literature review of the genetic associations of two of the MDD traits we have proposed to study (i.e., neuroticism and age at onset) is given below.

Calboli et al. (2010) have found that neuroticism is a moderately heritable MDD trait and a risk factor for developing MDD. They performed a genome-wide association analysis of 2,235 participants drawn from a large population-based study of neuroticism using 430,000 autosomal

SNPs together with 1.3 million imputed SNPs from the HapMap CEU samples (Gibbs et al., 2003). They found that the gene NKAIN2 showed suggestive association ($p < 10^{-6}$) with neuroticism as a main effect and the gene GPC6 showing evidence of interaction ($p \approx 10^{-7}$) with age. They found support for one previously associated association with the gene PDE4D but failed to replicate other recent finds. They concluded that common SNP variation does not strongly influence neuroticism.

Brummett, Boyle, Kuhn, Siegler, and Williams (2008) examined prolactin responses to a tryptophan challenge as they relate to the Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness (NEO Five Factor Inventory or NEO-FFI used to measure neuroticism quantitatively; described more fully in 2.1) using 67 volunteers. Their findings indicate that the gender moderated the association between neuroticism and prolactin level ($p < 0.03$) and higher levels of neuroticism were associated with decreased levels of prolactin responses in females; males responded exactly opposite, i.e., higher levels of prolactin produced higher levels of neuroticism.

Lake, Eaves, Maes, Heath, and Martin (2000) examined the hypothesis that environmental transmission is a significant factor in individual differences for neuroticism among 45,850 members of extended twin kinships from Australia (N=20,945) and the U.S. (N=24,905). They found no such evidence of environmental transmission influencing neuroticism levels in replicated samples from two different continents, concluding that a simple genetic structure underlies familial resemblance for the neuroticism trait.

In a study of gene x environment interactions in mental disorders Tsuang, Bar, Stone, and Faraone (2004) state that association studies provide potentially useful approach to the detection of gene-environment interactions in mental disorders. They point to depression-like behaviors in

rats and various nonhuman protocols when they are exposed to environmental stress such as maternal separation, neglect, and social deprivation. However, the behavioral responses differ depending on the individual genetics and the interaction effects of the environment on the genes.

Unlike neuroticism, the age at onset of MDD has not been explored under the umbrella of genetic association studies, especially in cases of early childhood depression in children as young as age 3 (Luby, 2009). However, the literature points to several linkage studies of recurrent, early-onset major depression (Camp et al., 2005; Holmans et al., 2004; Zubenko et al., 2003). These were discussed earlier under genetic linkage studies. Other genetic association studies of MDD involving specific genes such as 5-HTTLPR and IL-6 are discussed next.

Myung et al. (2010) studied the serotonin transporter gene polymorphisms and chronic depression recurrence in Korean subjects using 252 patients with MDD. The patients were genotyped for s/l polymorphisms in 5-HTT promoter and s/l variation in the second intron of the 5-HTT gene (5-HTT VNTR intron 2). Chronicity was associated with 5-HTTLPR where the l/l allele showed higher rate of chronicity than s/l or s/s varieties (OR: 4.45; 95% CI: 1.59 – 12.46; $p = 0.005$). Thus 5-HTTLPR was implicated in the association with MDD, while the 5-HTT VNTR intron2 was not.

Uddin et al. (2010) studied epigenetic differences between patients with lifetime MDD and those without. They used a community-based sample in a Detroit neighborhood taken from a sample of participants in the Detroit neighborhood Health Study (DNHS). Their sample contained 33 persons with lifetime MDD and 67 persons without lifetime MDD. Bioinformatic analyses were performed on the genes uniquely methylated and unmethylated in each group and inflammatory biomarkers Interleukin 6 (IL-6) and C-reactive protein (CRP) were measured to investigate the possible significance of the methylation profile. Their conclusion was that

examining epigenetic mechanisms in concert with other dynamic markers of physiologic functioning could improve our understanding of neurobiology of depression.

Lesch (2004) comprehensively studied MDD and the gene x environment interactions for MDD in a paper describing how starting from family studies and linkage studies the research has moved on to the genetic association studies of MDD. Investigation of subtle alterations in gene expression, correlations between genotype and brain activity and of environmental variables interacting with genetic variants have advanced research into the genetics of depression. He adds that gene-phenotype correlations have been substantiated by functional neuroimaging and the notion of gene networks that control brain development is increasingly recognized. However, given the etiologic and psychobiologic complexities of mental disorders, identification of specific genetic factors is extremely difficult and gene hunting continues.

MDD like other complex genetic diseases is heterogeneous in origin and is polygenic where different susceptibility genes may be operating in different families. Thus the gene x environment interactions need to be fully evaluated before understanding the etiology of MDD and there is a need to investigate the molecular basis of MDD. This is discussed in the next subsection briefly.

Molecular Basis of Major Depressive Disorder

Jabbi et al. (2008) point out that there is growing evidence linking disease phenotypes with genes on one hand and the genesis of stress related disorders like MDD as a result of exposure to environmental pathogens on the other. They conclude that advocating the use of convergent approaches in examining the genetic modulation of disease phenotypes might be useful. Taking the example of MDD and a few well-known MDD associated genes (e.g., 5-HTT, MAOA, and COMT) the researchers explored the gene x gene interactions to find out how certain genetic

markers may result to endophenotypes like HPA axis dysfunction leading to susceptibility for some of the co-occurring neuropsychiatric disorders like MDD and neuroticism. They further point to such results as HPA-axis involvement in the maintenance of homeostasis during stress and a speculation of interaction between COMT and MAOA on peripheral endocrine responsivity to stress may be related to a combined influence of these genes affecting the CRH functioning. Thus understanding genetic associations at the molecular functioning level are needed to further understand MDD and the modulation of complex behavioral adaptations relevant for a disease phenotype. Further discussion of this topic is beyond the scope of this report.

Genome-Wide Association Studies of Major Depressive Disorder

Large-scale studies are needed with multiple phenotypes, DNA, and ideally, biological material that enables gene expression and other genomic and proteomic analyses. Genome-Wide analysis study is one such large-scale undertaking with data collected over several years in order to identify potential genes and genomic regions that are significant in patients diagnosed with cases of MDD.

The conventional genome-wide association study (GWAS) approach is a hypothesis-free, systematic search of tagging Single Nucleotide Polymorphisms (SNPs) across the human genome to identify novel gene associations with common diseases and has emerged as a powerful tool to identify disease-related genes for many common human disorders and other phenotypes (Guessous, Gwinn, & Khoury, 2009; McCarthy et al. 2008; Wellcome Trust Case Control Consortium 2007). Recently, several GWAS of MDD have been reported. For example, Sullivan et al. (2009) reported the possible role for the presynaptic protein piccolo (gene PCLO);

Shyn et al. (2009) found strong associations for genes ATP6V1B2, SP4, and GRM7 from a meta-analysis; Garriock et al. (2010) found three SNPs within genes UBE3C, BMP7, and RORA associated with MDD with p-values less than 1×10^{-5}; and Shi et al. (2010) found SNPs within the gene SP4. Also, shared genetic risks for MDD and bipolar disorder have been reported (McMahon et al., 2010).

One of the first Genome-Wide Association (GWA) studies that examined gender differences for MDD is now published online (Aragam, Wang, & Pan 2011). That study identified 40 male-specific and 56 female-specific MDD associated Single Nucleotide Polymorphisms (SNPs) with P-values less than 10^{-4}. In addition, 38 SNPs showing gene x gender interactions influencing MDD ($P < 10^{-4}$) were also found. The implicated genes in this study were LGSN, PCLO, FIGN, and OR4B1.

A brief summary of Genomics and GWAS literature is provided in Appendix B for interested readers.

Genome-Wide Association Studies of Neuroticism

In earlier GWA studies, van den Oord et al. (2009) have reported finding potential association between the MAMDC1 gene and neuroticism, while Shifman et al. (2008) have found suggestive association between the PDE4D gene and neuroticism in one sample and replicated in another. However, replication in some other samples by Shifman et al. was not possible. Terraciano et al. (2010) found potential association signals for all five neuroticism scales, but the effect sizes were small and most associations failed to replicate in other samples. Several researchers have also reported an association of Neuroticism and Alzheimer's disease in literature (Wang et al., 2009; Wilson et al., 2003).

The differences in the quantitative trait scores for neuroticism in men and women have been found to moderate the prevalence of MDD in females. Recently some studies of neuroticism have been reported to identify gender differences in MDD severity as a function of neuroticism total score (Eaves, Heath, Neale, Hewitt, & Martin, 1998; Fanous et al., 2002; Lake et al., 2000; Liu et al., 2006) with the males showing consistently less severity than females (Jardine et al., 1984; Jorm, 1987). Other population and gender studies on neuroticism in the literature include, a twin study of the relationship between internalizing disorders and neuroticism (Hettema et al., 2006), association of serotonergic function with neuroticism being moderated by the gender (Brummett et al., 2008), neuroticism and dementia (Wang et al., 2009), and prospectively assessed neuroticism association with anorexia nervosa among Swedish twins (Bulik et al., 2006).

Summary

Literature about MDD as a psychiatric trait, or a disease phenotype, is well-known and has been studied for a number of years. MDD is heritable, runs in families, and is found to be comorbid with smoking and is more prevalent in women than in men. An elaborate international instrument (WMH-CIDI) has been developed to study MDD epidemiologically and is in wide use. MDD is a chronic disease by its own accord and should not be confused with other mental disorders like bipolar disorder and schizophrenia. Furthermore, it is conceptualized as a binary trait in this study although some experts argue otherwise.

Of late MDD is being studied as a complex genetic disease, and researchers are looking to conduct genome-wide studies to understand and elucidate the genetic variants of this chronic disease. Although strong correlations exist between neuroticism and MDD, few studies have

looked at neuroticism as an endophenotype of MDD. Linkage and family studies have been conducted widely in the past to unearth several genetic variants of MDD, but the implicated genes have only been attributed to have small effects. Lately genome-wide studies have begun with the availability of large amounts of genotype information and SNP data from selected samples of MDD patients.

Genome-Wide association analysis is gaining ground as an attractive hypothesis-free approach to study complex genetic diseases such as MDD. Although neuroticism and family risks (i.e., being a first degree relative of someone who has been diagnosed with MDD) are known predictors, few genome-wide studies using neuroticism and age at onset have been found in the literature. In this proposed research study, both neuroticism and age at onset via inheriting specific genes are studied to explore genome-wide association of MDD in a specific subset of human population representative of the people with a European origin.

Determining the use of endophenotypes of MDD (e.g. neuroticism) as quantitative traits to find the genetic variants of MDD and to find out how the analysis of age at onset of MDD as a quantitative trait may help to understand the prevalence and severity of MDD lays the foundation for this research work. A major goal of our study is that the study results will provide a genetic basis for the elucidation of MDD and its phenotypes for future studies that examine other populations.

CHAPTER 3

MATERIALS AND METHODS

The overall goal of this research study is to understand and investigate the epidemiology and underlying genetic basis for the onset, prevalence, and the chronic suffering (e.g., 12-month disease burden, being disposed to such comorbidity as neuroticism) of those with major depressive disorder (MDD). The following sections describe details of data sets and study participants used for this research. A detailed study plan is also presented.

Datasets

The two data sets used in this research are briefly described below -

#1 GAIN Sample

This dataset is a combination of samples from the Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Registry (NTR). These data are available from the public database of National Center for Biotechnology Information (NCBI), known as dbGaP. Briefly, the data were collected from two longitudinal studies between 1991-2004 that used clinical interviews, demographic questions, and biobanking home visits. Data collection for the NESDA study was from 1996–2004 and for the NTR was from 1991-2004. There are 3,741 records in the combined NESDA and NTR data of which 1,991 records show a diagnosis of MDD.

In 2006 a consortium of researchers from the University of North Carolina at Chapel Hill (UNC-CH), Virje University in Amsterdam, and the Universities in Groningen and Leiden were selected to genotype the combined NESDA and NTR samples in order to conduct a GWAS of major depression. This study of MDD by the consortium of researchers came to be known as one

of the six Genetic Association Information Network (GAIN) studies. The GAIN initiative is a part of a public-private partnership of the Foundation for the National Institutes of Health Inc. The funding for this initiative came from three pharmaceutical companies – Pfizer, Affymetrix, and Abbott Laboratories. Not all 1991 NESDA cases were selected for the GAIN study and a few of the non-MDD controls were added from the NESDA sample. Thus the combined NESDA and NTR sample consists of 1,738 MDD cases and 1,773 controls with 1,213 males and 2,298 females.

Genotyping data using the Perlegen 600K SNP chips (total 437,547 filtered SNPs) are available for 1,738 cases and 1,773 controls in the GAIN database. GWA genotyping for the selected candidates (see Chapter 3 for details) was done by Perlegen Sciences and the Stage 1 results became available in 2007 via the National Center for Biotechnology Information (NCBI). The complete database of candidates' phenotype and genotype information is known as dbGaP and it has the following web portal (http://www.ncbi.nlm.nih.gov). Access to this public database requires preapproved applications for research in MDD with the restriction that use of the genotype and phenotype data is limited to psychiatric health and related somatic conditions.

#2 OZALC Sample

The data for the OZALC study were obtained from telephone diagnostic interviews of two general population volunteer cohorts consisting of Australian twins and their spouses, – a total of 11,000 families. The data used come from the publicly available data from the Genome wide Association Study of Alcohol Use and Alcohol Use Disorder in Australian Twin-Families (OZALC GWAS) – Study Accession: phs000181.v1.p1. The details about these subjects are described elsewhere (Grant et al., 2009, Nelson et al., 2004). Genotyping data using the Illumina

Human CNV370v1 (total of 343,955 SNPs) are available for 4,119 individuals in this data set. After merging with pedigree and phenotype of Anti-Social Personality Disorder (ASPD), we removed one from each of 44 Monozygotic (MZ) twins and 72 outliers based on the data description. Consequently, 103 cases of ASPD were left. This database is used for replication study only and for finding common genetic variants in both Neuroticism and ASPD. This data set is used in the study of gene x gender interaction of neuroticism.

Overview of Sample Selection, Data Collection, and Data Characteristics

Sample Selection

The defining features of MDD used for this study are those used for NESDA study and are marked by persistent depressive moods associated with physical and cognitive signs and symptoms (e.g., insomnia, anhedonia, appetite and weight changes, psychomotor agitation or retardation, etc.) as described in Boomsma et al. (2008). The combined dataset also contain severity indexes for mental disorders such as neuroticism and generalized anxiety disorder for MDD cases.

Major Depressive Disorder Cases - Inclusions and Exclusions

The inclusion criteria for the MDD cases were:

- WMH-CIDI diagnosis of MDD
- Age between 18-65 years
- Knowledge of Dutch language
- North-European ancestry.

The exclusion criteria for the MDD cases were:

- People with primary diagnoses of psychosis, bipolar disorder, obsessive compulsion disorder, severe addiction disorder
- Insufficient knowledge of the Dutch language

Major Depressive Disorder Controls - Inclusions and Exclusions

The inclusion criteria for MDD controls were:

- Never scoring high (> 0.65) on a general factor score for anxious depression
- Never reported any history of MDD
- Age between 18-65
- Knowledge of Dutch language
- North-European ancestry

The exclusion criteria for controls arose when there were multiple eligible controls in a family. In that case, gender and age were matched with cases first and then the control that completed the highest number of questions in the questionnaire was selected.

The selected cases and controls in the GAIN database were then genotyped by Perlegen Sciences (Mountain View, CA) that was blinded as to the case or control status and the corresponding SNP data were stored in the database of Genotypes and Phenotypes (dbGaP) as a part of the GAIN initiative.

Data Collection

The NESDA cases were selected from households in 90 Dutch municipalities and had been diagnosed with lifetime MDD or anxiety disorder during one of the CIDI interviews in 1996, 1997, or 1999. The total catchment area for NESDA had a population of 1,175,000 people and selection was done using stratified random sampling to select potential study candidates and

7,076 people responded. The sample also included a subgroup of 18-25 year olds that participated in for the Adolescent at Risk for Anxiety and Depression.

The collaborative NESDA study involved four academic and two nonacademic centers. NESDA was a longitudinal cohort study (1996 – 2004) that followed 2,850 persons, aged 18-65 years, and had five assessments, one at the baseline and after 1, 2, 4, and 8 years of follow-up. The clinical diagnosis of MDD was based on a detailed survey. The criterion for a case selection was the diagnosis of a depressive or anxiety disorder during a standardized intake assessment. The process of recruitment consisted of three phases –

1. A preliminary screening questionnaire filled out at the GP's office where the subject consulted for depressive symptoms.
2. Those who screened positive were phone interviewed with the short form of the CIDI.
3. Those with current diagnosed MDD status were asked to participate in NESDA and were further invited for a baseline assessment that included the full WMH-CIDI interview.

The final inclusion criterion was a diagnosis of DSM-IV MDD, an age between 18 and 65 years and self-reported European ancestry (Boomsma et al., 2008).

The controls data from the NTR were collected longitudinally by mail surveys every 2-3 years since 1991. There are about 22,000 participants from 5,546 families who were all assessed for depressive symptoms, anxiety, neuroticism, and other personality disorders. The controls never scored high on a general factor score (mean = 0.0 and SD = 0.7; a score of > 0.65 constitutes depressive disorder) to qualify for anxious depression. The factor score was a combined measure of neuroticism, depressive symptoms, and anxiety assessed via longitudinal questionnaires. These control subjects never reported a history of MDD in any survey or the

blood sampling visits. The controls were similar with cases in demographics (i.e., of Dutch origin), gender distribution (40% men and 60% women), and age (18-65). But they did not have a lifetime diagnosis of MDD or anxiety disorder as assessed by the WMH-CIDI. There were also a few (167) healthy controls selected from NESDA study in this sample. Further details of selection objectives, recruitment, and methods of NESDA have been described elsewhere (Boomsma et al., 2008; Kessler, et al. 2003; Penninx, et al., 2008).

Data Characteristics

Baseline characteristics of cases and controls are summarized in Tables 1 and 2 given on the website (www.ncbi.nlm.nih.gov/dbgap) on October 9, 2007. Accordingly, the ages of participants (both cases and controls) were not significantly different, but the number of women in cases was higher than in controls. More controls were married or living with a partner (20% more) than cases and there were almost double the number of cases who smoked (42% vs 20%) compared to controls. Alcohol use was higher in controls than in cases (80% vs. 66%) and the NEO neuroticism score had a higher mean in cases than in controls (39.3 vs. 28.2). More complete details of how the data were collected, sampled, and the establishment of the Genetic Association Information Network (GAIN) are described by Boomsma (Boomsma et al., 2008).

Table 1

Characteristics of GAIN MDD Cases

Characteristic	Predictor Type	MDD Cases; N = 1821
Age (years)	Continuous	42.1 ± 12.7
Female (% Yes)	Categorical	69.7
With Partner / Married (% Yes)	Categorical	68.5
Smoking (current in % Yes)	Categorical	42.1
Alcohol use (last year in % Yes)	Categorical	66.3
NEO[a] neuroticism score (0 – 60)	Continuous	39.3 ± 8.0

Table 1 (Continued)

Family history of MDD (% Yes)	Categorical	85.4
Depression; Age at Onset (years)	Continuous	27.5 ± 12.3
IDS[b] depression (0 – 69)	Continuous	25.8 ± 13.5
BAI[c] anxiety (0 -62)	Continuous	14.3 ± 10.6

For continuous traits, the mean score and SD are listed.

[a]NEO = Neuroticism, Extrovertness, Openness.

[b]IDS = Index of Depressive Symptomatology.

[c]BAI = Beck's Anxiety Inventory.

Table 2

Characteristics of GAIN MDD Controls

Characteristic	Predictor Type	MDD Controls; N = 1822
Age (years)	Continuous	45.1 ± 14.1
Female (% Yes)	Categorical	61.8
With Partner / Married (% Yes)	Categorical	87.0
Smoking (current in % Yes)	Categorical	20.4
Alcohol use (last year in % Yes)	Categorical	80.0
NEO[a] neuroticism score (0 - 60)	Continuous	28.2 ± 5.5
ABV[b] neuroticism score (11 -124)	Continuous	36.7 ± 16.0
ABV somatic complaints (12 – 46)	Continuous	15.9 ± 3.8
STAI[c] trait anxiety (20 – 79)	Continuous	30.1 ± 5.6
YASR[d] anxious depression (0 – 21)	Continuous	3.22 ± 2.7
Beck depression (0 – 23)	Continuous	1.11 ± 1.4

For continuous traits, the mean score and SD are listed.

[a]NEO = Neuroticism, Extrovertness, Openness.

[b]ABV = Amsterdam Biographic Survey.

[c]STAI = Spielberger Trait Anxiety Inventory.

[d]YASR = Young Adult Self Report.

Formation of dbGaP

From the NESDA participants the cases for the GAIN MDD study and genotyping were selected from September 2004 through February 2007. Although an initial 1,821 cases and 1,822 controls were available for analyses, final selection in the GAIN sample excluded certain subjects (because of genotyping and sample problems) and included some others (e.g., parents of some controls) to form trios (the subject and his/her parents). After the above adjustments a total of 1,860 cases and 1,857 controls were included in the GAIN database.

Figure 1 below illustrates the make-up of the GAIN database of genotypes and phenotypes (dbGaP). The phenotype data for NESDA consisted of responses from the WMH-CIDI forms (described in Appendix A2) and represented MDD, depressive symptoms, and other psychopathology indicators. In addition, the CIDI interview also provided information on age at onset, number of episodes of MDD, and specific symptoms of depression along with quantitative scores for NEO-FFI, IDS and BAI.

The biobank samples from NESDA and NTR participants were collected and stored at the time of their baseline visit for CIDI interviews (for NESDA participants) and a separate blood or urine collection project (for NTR participants). All necessary protocols were followed for collection, storage, and genotyping of the biosamples. The dbGaP holds both phenotype and genotype information for cases and controls. The individual researcher or the team of researchers who gets access to the data in dbGaP might be required to reformat the data tables to suit the needs of a particular study.

For example, a data item in a specific column of a table in the database might not be useful for a specific study and may have to be deleted; or, one may want to categorize a continuous data item (e.g., age of 18-65 may be categorized into 18-30, 31-60, 61+), etc. In case of genotypes

data one may have to filter the data for conformity with Hardy-Weinberg proportion and Minimum Allele Frequency (usually > 1%) and so one might eliminate certain SNPs and individual data row.

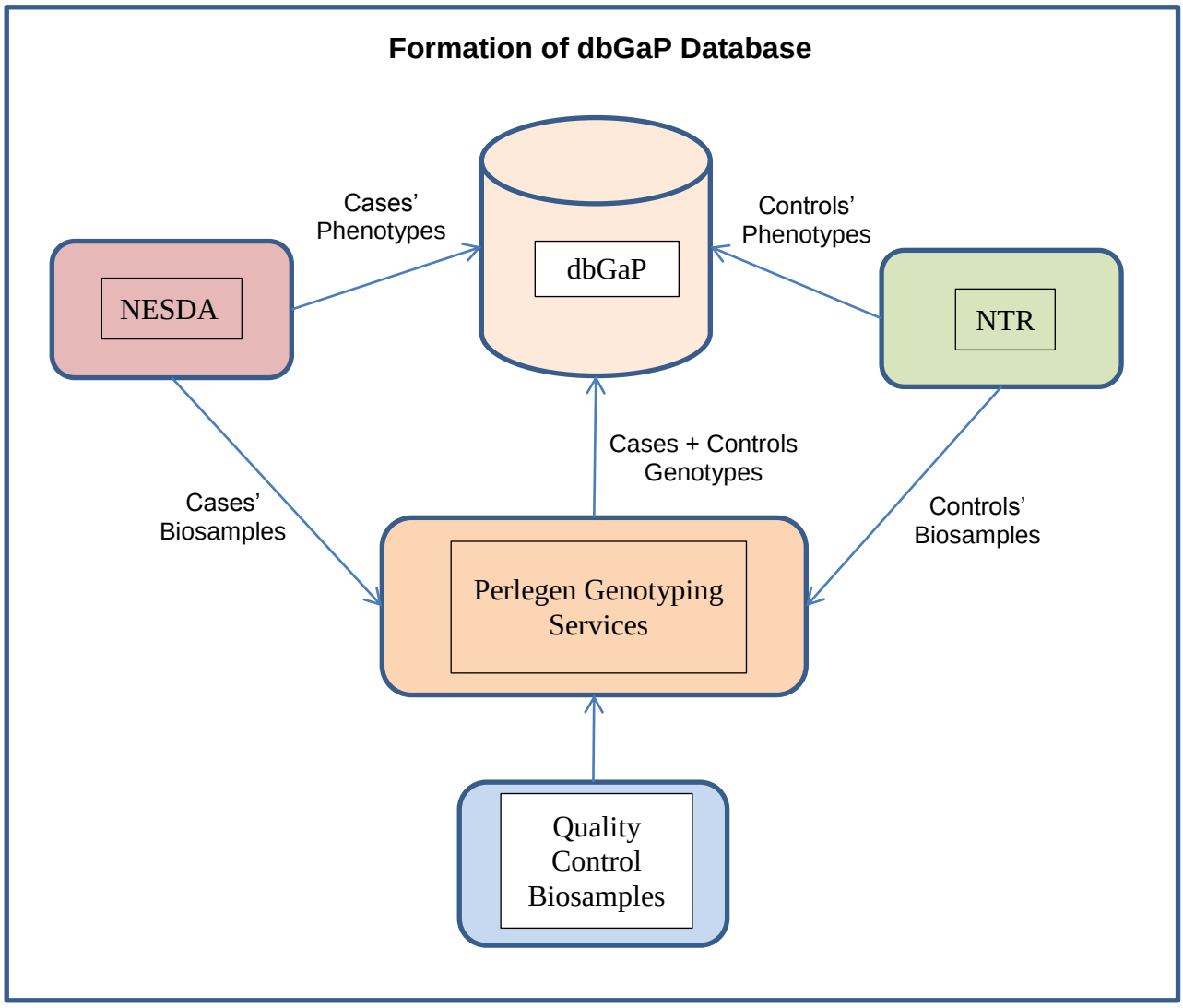


***Figure 1*. Formation of dbGaP**

In addition, one may want to eliminate outliers in the data and so might need to use other software tools to run the data required first before making it ready for the required analyses. This type of preliminary processing for needed quality control of the data is the responsibility of the researcher who is given access to the dbGaP data.

Study Designs for Data Analyses

Using the two data sets described above, three study designs are developed below for addressing the three research questions we identified earlier in chapter 1. These are given below:

Study Design for Research Goal #1: Risk Factors for Major Depressive Disorder

Only the primary data set #1 GAIN dbGaP described earlier was used to address this research goal.

Study population. Data from 3,356 individual records sampled from the NESDA/NTR phenotypes dataset and stored in GAIN dbGaP was used to conduct descriptive and regression analyses.

Study Variables. Four continuous and five categorical predictor variables are used in order to determine the various quantitative relationships. These are shown in Table 3.

Table 3

Study Variables Used for Research Goal #1

Outcome Variable	Predictor Variables
MDD Affection Status (Binary)	Age (continuous)
	Gender (categorical)
	Marital status (categorical)
	Smoking Status (categorical)
	Alcohol Use (categorical)
	Family Risk (categorical)
	Neuroticism (NEO) (continuous)
	Index of Depressive Symptomatology (IDS) (continuous)
	Beck's Anxiety Index (BAI) (continuous)
Neuroticism (Continuous)	Age (continuous)
	Gender (categorical)
	Marital status (categorical)

Table 3 (Continued)

	Smoking Status (categorical)
	Alcohol Use (categorical)
	Family Risk (categorical)
	Index of Depressive Symptomatology (IDS) (continuous)
	Beck's Anxiety Index (BAI) (continuous)
Age at Onset (Continuous)	Gender (categorical)
	Marital status (categorical)
	Smoking Status (categorical)
	Alcohol Use (categorical)
	Family Risk (categorical)
	Index of Depressive Symptomatology (IDS) (continuous)
	Beck's Anxiety Index (BAI) (continuous)
	Neuroticism (NEO) (continuous)

The potential confounders are Age, Gender, Marital Status, Smoking Status, Alcohol Use, and Family Risk. The potential effect modifiers are Gender, Smoking Status, and Family Risk.

Statistical Methods

Descriptive Analyses. SAS 9.2 and SPSS 17 were used for descriptive analyses. Univariate analysis was conducted to obtain baseline information prior to conducting more advanced bivariate and multiple variable analyses. The univariate analysis of MDD phenotypes such as NEO, IDS, and BAI provide us an estimate of distributions of phenotypes in the study population. Histograms and box plots were used to graphically show and compare the prevalence of specific MDD phenotypes for men and women, married and unmarried, smokers and nonsmokers, and so forth. We hope to replicate prior established facts such as a higher

prevalence of MDD in females as opposed to males, comorbidity smoking and MDD, and described familial risks for MDD such as an affected first-degree relative increasing the risk of MDD in individuals. In addition, the statistical distributions of continuous and categorical predictors are plotted using SPSS software.

Bivariate Analyses. Simple logistic regression methods were used to analyze the data in order to evaluate the association of the covariates and MDD. Bivariate logistic regressions of the primary outcome variable (MDD affection) were carried out for age, neuroticism score, gender, marital status, smoking status, alcohol use, and family risk. The results were used to examine the strength of the potential risk association of the different predictors for MDD and its related phenotypes.

Multiple Variable Analyses. Multiple logistic regressions were used to measure the relationship between the outcome variable MDD affection and the different predictors of interest. Multiple logistic regressions use an odds ratio (OR) to assess the association and potential risk of MDD affection for one predictor while adjusting for all other predictors or covariates. The covariates that were used for multiple logistic regression of MDD affection are – age, gender, marital status, smoking status, alcohol use, neuroticism score, and family risk. Multiple linear regressions of continuous predictors like NEO, IDS, and BAI assess the relationship between the continuous outcome variable (NEO, IDS, BAI) and a predictor like age, gender, marital status, smoking status, etc. while adjusting for all other covariates.

The SAS procedures used for analyses are “proc reg”, “proc logistic”, and Generalized Linear Model (“proc glm”). Odds ratios along with Wald confidence intervals for MDD with predictor variables are computed using the multiple logistic regression procedure. The multiple regression procedures used ‘backward’ selection of variables and included only those predictors

that were significant at the level of 0.05 alpha values. In addition, interactions were tested for the three leading predictor variables - gender, smoking status, and family risk. The "glm" procedure is used with the predictors gender, marital status, smoking status, alcohol use, and family risk - as class variables to discover potential confounding effects. After eliminating the confounding effects of predictors using multiple variable regressions a reasonable relationship between the primary outcome variable (MDD affection) and the predictor variables (MDD phenotypes) were assessed.

Study Design for Research Goal #2: Genome-Wide Association Analysis of Neuroticism

Two data sets are used in this study – one primary (#1 GAIN dbGaP) and one secondary (#2 OZALC) for replication purposes. The data sets #1 and #2 were described earlier in the sections on data sources and participants.

Study Population. For this Research Goal, 2,748 individuals (902 males and 1,846 females) were analyzed. There were 437,547 SNPs available for the data. These managed data were the result of the data quality control and stratification steps taken in the previous phase. Any additional data management required before analyzing the continuous severity variables selected using PLINK and/or other software was accomplished first. The second data set is used for comparing the results of significant SNPs obtained in the association studies with the significant SNPs obtained from a family-based association analysis of Australian twins for Anti-Social Personality Disorder (Wang et al., 2011).

Study Variables. Both phenotypic and genotypic variables were used in this study. The outcome variable Neuroticism (NEO) and the predictor Age were continuous variables and the covariates Gender and Genotype were categorical. The genotypes of 902 males and 1,846

females available from the study data set #1 were used for both neuroticism association and gene x gender interaction studies.

Statistical Methods.

Quality Control. For the initial GWA analysis of MDD, HelixTree Software (http://www.goldenhelix.com/SNP_Variation/HelixTree/index.html) was used to assess genotype data for conformity with Hardy-Weinberg equilibrium (HWE). To deal with population stratification, the principal-component analysis approach (Price et al., 2006) in HelixTree was used to identify outlier individuals.

Linear Model. To test for association with a quantitative trait, linear regression was performed by PLINK to obtain the regression coefficient and Wald test asymptotic p-value. In addition to obtaining nominal P-values, empirical P-values are generated by 100,000 permutation tests using Max (T) permutation procedure implemented in PLINK. In this procedure, the corrected values for multiple testing (corrected empirical P-values) are calculated.

1. Gender-specific association study with separate male and female samples

(a) QT = b0 + b1*ADD + b2*Gender + b3*Age + e

where, QT is a quantitative trait (neuroticism), ADD is a dummy variable for coding the possible genotypes AA, Aa, and aa with the additive model; Gender is the variable for coding male or female sample data in the additive model; b0 is the intercept, b1, b2, b3 are the corresponding regression coefficients, and e is the error term.

For neuroticism, we used (a) to perform association analysis for MDD case samples controlling age.

2. Gene x gender interactions of neuroticism

(b) QT = b0 + b1*ADD + b2*Gender + b3*ADDxGender + b4*Age + e

where, QT is the quantitative trait (Neuroticism) being analyzed, ADD is the dummy variable for coding three genotypes AA, Aa, and aa with additive model, b0, b1, b2, b3, and b4 are the corresponding intercept, individual regression coefficients, Age and Gender are the covariates being used and e is the error term.

For neuroticism, we used (b) to test gene x gender interactions controlling age.

Statistical Significance. We used a conservative per test significance level of $\alpha=5x10^{-7}$ (Wellcome Trust Case Control Consortium, 2007). At the same time, we also had a less stringent criterion of "suggestive association" with a cut-off of $\alpha=10^{-4}$.

Study Design for Research Goal #3: Genome-Wide Association Analysis of Age at Onset

Only the primary data set #1 GAIN dbGaP as described earlier was used for this analysis.

Study Population. Sixteen hundred three subjects (481 males and 1,122 females) were used for this analysis. There were 437,547 SNPs available for the data. This managed data were the result of the data quality control and stratification steps taken in the previous phase. Any additional data management required before analyzing the continuous severity variables selected using PLINK and/or other software were accomplished first.

Study Variables. Both phenotypic and genotypic variables were used in this study. The outcome variable Age at Onset (AAO) was a continuous variable and the covariates Gender and Genotype were categorical. The genotypes of 481 males and 1,122 females available from the study data set #1 were used for AAO association with MDD.

Statistical Methods

Linear Model. Linear model analysis was performed for AAO using PLINK v1.07 (Purcell et al., 2007). To test for association with AAO as a quantitative trait, linear regression was performed by PLINK to obtain the regression coefficient and Wald test asymptotic p-value.

(a) $QT = b0 + b1*ADD + b2*Gender + e$

where, QT is a quantitative trait (Age at Onset), ADD is a dummy variable for coding the possible genotypes AA, Aa, and aa with the additive model; Gender was the variable for coding male or female sample data in the additive model; b0 is the intercept, b1 and b2 are the corresponding regression coefficients, and e is the error term.

For Age at Onset we used (a) to perform association analysis for MDD case samples controlling gender.

Statistical significance. We used a conservative per test significance level of $\alpha=5x10^{-7}$ (Wellcome Trust Case Control Consortium, 2007). At the same time, we also had a less stringent criterion of "suggestive association" with a cut-off of $\alpha=10^{-4}$.

CHAPTER 4

RESULTS

The results from the three studies identified in the previous chapter are presented here. First, some of the results from the descriptive analyses are illustrated as graphical plots showing some of the significant distributions of MDD related phenotypes in the sample population selected for univariate analyses. The histograms illustrated below in Figure 2 correspond to two of the three major predictors of MDD we have found, namely, gender and marital status. Because smoking is both a confounder and a possible effect modifier, and because we do not know whether the smoking data are pre- or postdiagnosis of MDD, we cannot rely on any illustrative interpretations of the raw distribution plots that involve any smoking data.

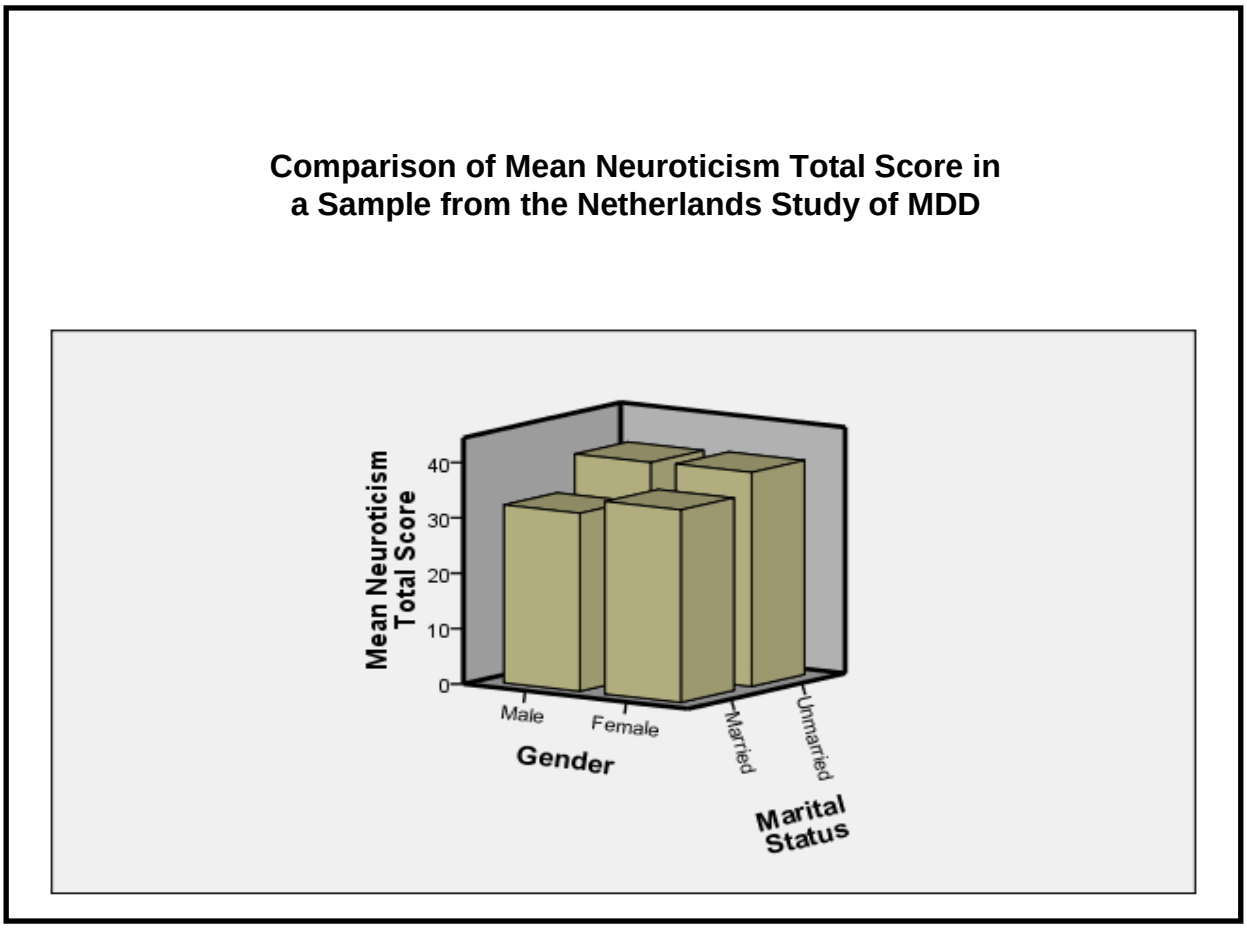


Figure 2. **Comparison of Mean Neuroticism Scores**

The histograms show the pattern of mean frequencies in our NESDA-based sample and should not be generalized for any other population. From Figure 2 we can surmise that the mean score for NEO-FFI in unmarried females seems to be slightly higher (in value) than unmarried males but certainly higher than both married males and females.

In Figure 3 we see that the age at onset is the earliest in unmarried males while not much difference is seen in case of females. The finding that unmarried males are prone to early age at onset of MDD will need to be further evaluated by future studies.

In Appendix C we show some other histograms generated during the descriptive analysis and Gene x Gender interaction analysis of Neuroticism scores in our sample.

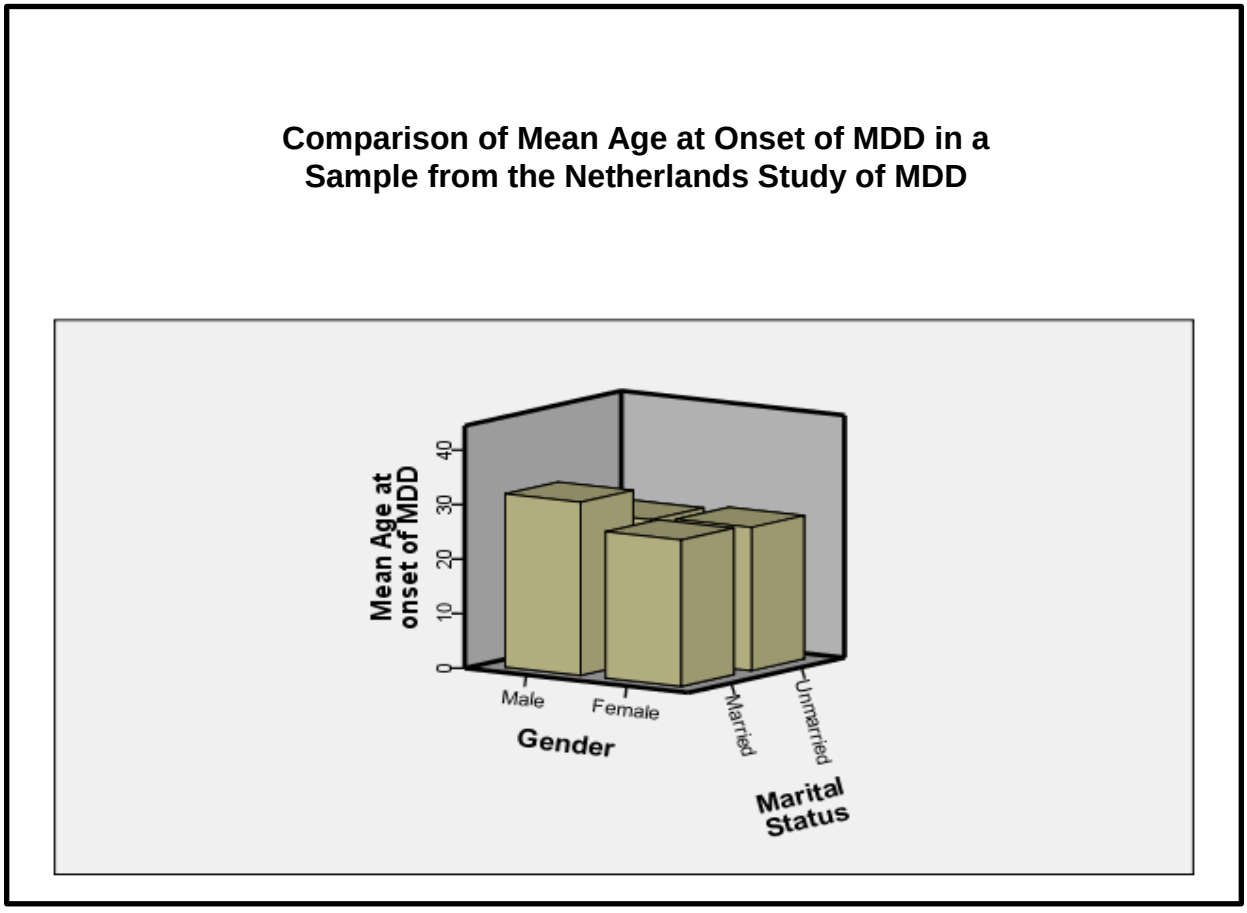


***Figure 3*. Comparison of Mean Age at Onset**

Risk Factors of Major Depressive Disorder

The results from the logistic regression of Affection_Status of MDD for each of the predictor variables – Age, Gender, Marital Status, Smoking Status, Alcohol Use, Family Risk and NEO Score are summarized in Table 4 below. They all show statistically significant relationships with MDD. The OR estimates indicate that Age, Marital Status, and Alcohol Use show a reduction in risk OR < 1); all others show increased risks (OR > 1). The bivariate analyses do not adjust for potential confounding and further evaluations are required. Table 4 shows the following odds ratios and confidence intervals for each of the predictors tested – Age (OR: 0.984, 95%CI: 0.980 – 0.989); Gender (OR: 1.428, 95% CI: 1.242 – 1.643); Marital Status (OR: 0.327, 95% CI: 0.275 – 0.389); Smoking Status (OR: 2.879, 95% CI: 2.476 – 3.347); Alcohol Use (OR: 0.498, 95% CI: 0.426 – 0.583); Family Risk (OR: 3.451, 95% CI: 2.423 – 4.916); NEO Score (OR: 1.243, 95% CI: 1.223 – 1.264).

Table 4 shows the bivariate analyses for Family Risk and NEO Score with reduced sample sizes. The sample sizes for these analyses are reduced because the phenotypes evaluated apply only to cases and some of the data (esp. for family risk) are missing in the dataset #1. Further discussion of these results is deferred until Chapter 5.

Table 4

Bivariate Logistic Regression of Major Depressive Disorder with Predictor Variables

Outcome Variable = Affection_Status (MDD)							
Predictor Variable	**N**	**Model Fit (-2LL)**	**Parameter Estimate**	**Wald χ^2**	**p-value**	**OR**	**95% CI for OR**
Age	3,510	4,827.322	-0.0156	37.8419	< 0.0001	0.984	(0.980 – 0.989)
Gender	3,511	4,841.868	0.3564	24.9257	< 0.0001	1.428	(1.242 – 1.643)
Marital Status	3,457	4,620.548	-1.1183	158.9867	< 0.0001	0.327	(0.275 – 0.389)

Table 4 (Continued)

Smoking Status	3,488	4,636.111	1.0574	189.4331	< 0.0001	2.879	(2.476 – 3.347)
Alcohol Use	3,380	4,606.269	-0.6965	75.7377	< 0.0001	0.498	(0.426 – 0.583)
Family Risk	1,826	1,018.826	1.2387	47.0730	< 0.0001	3.451	(2.423 – 4.916)
NEO Score	2,748	2,529.105	0.2176	666.5801	<0.0001	1.243	(1.223 – 1.264)

Multiple variable regressions with adjustment for confounding must be carried out before specific predictors might be held responsible for possible associations with the outcome of diagnosed MDD affection in a subject from the study sample. Multiple variable logistic regressions were run with the binary outcome variable (i.e., Affection_Status) for MDD and the requisite predictors identified earlier – Age, Gender, Marital Status, Smoking Status, Alcohol Use, and Family Risk. The SAS regression tool eliminated the predictor 'Age' from consideration because of statistical insignificance (P-value > 0.05) and other reasons. We used age as a continuous variable and did not categorize the age into different categories that could have shown significance in certain age groups (e.g. postadolescence age of 17-25). The analysis was carried out starting with the predictor family risk first; other predictors were added one at a time in the following order – NEO Score, Smoking Status, Alcohol Use, Marital Status, and Gender. The results of logistic multiple variable regressions of Affection_Status (for MDD) with individual predictor variables are shown in Table 5 below. Tables 6, 7, and 8 show results from the interaction studies of the significant predictors.

Table 5 shows the following odds ratios and confidence intervals for each of the predictors tested – Gender-female (OR: 1.282, 95% CI: 1.078 – 1.602); Marital Status (OR: 0.432, 95% CI:

0.330 – 0.564); Smoking Status (OR: 3.079, 95% CI: 2.447 – 3.874); Alcohol Use (OR: 0.407, 95% CI: 0.318 – 0.521); Family Risk (OR: 3.112, 95% CI: 2.172 – 4.461); NEO Score (OR: 1.231, 95% CI: 1.210 – 1.253). The predictor Age is not at all significant (p = 0.7496) and was eliminated in the analysis as a predictor by the SAS program. The corresponding results for Age are shown as greyed out in Table 5.

Our analyses demonstrate that the major risk factors for MDD are Gender, Smoking Status, Family Risk, and NEO Score. The odds ratios (ORs) for Gender (female), Smoking Status, Family Risk, and NEO Score all show increased risks for MDD. The adjustment for confounding does not remove the higher risk of MDD that was noted for female Gender, Smoking Status and the Family Risk (i.e., a history of depression among the first degree relatives. The interaction between Gender and Smoking Status, Gender and Family Risk, and Smoking Status and Family Risk do not show any significance at a 5% (alpha = 0.05) significance level. Further discussion of these results is given in Chapter 5.

Table 5

Multiple Logistic Regression of Major Depressive Disorder with Predictor Variables

Outcome Variable = Affection_Status (MDD); N = 2730; -2LogL = 2330.167; Wald χ2 =696.0755; p < 0.0000000001					
Predictor Variable	**Parameter Estimate**	**Wald χ^2**	**p-value**	**OR**	**95% CI for OR**
Age	-0.0013	0.1021	0.7493	0.999	(0.991 – 1.006)
Gender	+1.2488	4.8161	0.0282	1.282	(1.078 – 1.602)
Marital Status	-0.8403	38.0185	0.0000000007	0.432	(0.330 – 0.564)
Smoking Status	+1.1245	23.7250	< 0.0000000001	3.079	(2.447 – 3.874)
Family Risk	+1.1354	38.2266	< 0.0000000001	3.112	(2.172 – 4.461)
Alcohol Use	-0.8980	50.9876	< 0.0000000001	0.407	(0.318 – 0.521)
NEO Score	+0.2082	565.9205	< 0.0000000001	1.231	(1.210 – 1.253)

Table 6

Multiple Logistic Regression of Major Depressive Disorder with Interaction Terms Gender*Smoking Status

Outcome Variable = Affection_Status (MDD); N = 3488; Hosmer & Lemeshow Model Fit χ^2 = 0.0000; p = 1.0000					
Predictor Variable	**Parameter Estimate**	**Wald χ^2**	**p-value**	**OR**	**95% CI for OR**
Gender	0.3495	15.3997	< 0.0001	1.501	(1.298 – 1.735)
Smoking Status	0.9678	59.1310	< 0.0001	2.966	(2.546 – 3.454)
Gender*Smoking Status	0.1818	1.2941	0.2553	n/a	n/a

Table 7

Multiple Logistic Regression of Major Depressive Disorder with Interaction Terms Gender*Family Risk

Outcome Variable = Affection_Status (MDD); N = 1826; Hosmer & Lemeshow Model Fit χ^2 = 0.0000; p = 1.0000					
Predictor Variable	**Parameter Estimate**	**Wald χ^2**	**p-value**	**OR**	**95% CI for OR**
Gender	0.3946	1.7784	0.1823	1.351	(0.933 – 1.958)
Family Risk	1.2680	20.8312	< 0.0001	3.293	(2.283 – 4.749)
Gender*Family Risk	-0.1117	0.0922	0.7614	n/a	n/a

Table 8

Multiple Logistic Regression of Major Depressive Disorder with Interaction Terms Smoking Status*Family Risk

Outcome Variable = Affection_Status (MDD); N = 1825; Hosmer & Lemeshow Model Fit χ^2 = 0.0000; p = 1.0000					
Predictor Variable	**Parameter Estimate**	**Wald χ^2**	**p-value**	**OR**	**95% CI for OR**
Smoking Status	1.2949	10.2781	0.0013	2.555	(1.686 – 3.872)
Family Risk	1.2646	38.1491	< 0.0001	2.970	(1.964 – 4.491)
Smoking Status*Family Risk	-0.4264	0.8222	0.3645	n/a	n/a

Notice that the results for OR and 95% CI's have been greyed out in tables 8, 9, and 10 above because the interaction terms are not significant at the $p < 0.05$ level and thus the corresponding values of ORs and 95% CIs need to be recomputed.

Genome-Wide Association Analysis of Neuroticism

First we tested the association of neuroticism as a quantitative trait in PLINK to obtain the Wald test asymptotic p-value. Thirty-four significant ($p < 10^{-4}$) neuroticism (as a quantitative trait) associated SNPs were detected in this analysis. Table 9 shows the top 20 neuroticism associated SNPs along with any associated genes and/or flanking marker genes. Accordingly, the best SNP associated with neuroticism ($p < 10^{-5}$) was rs4806846 ($p = 7.79 \times 10^{-6}$) within TMPRSS9 gene and the next best was rs220549 ($p = 1.05 \times 10^{-5}$) in GRIN2B gene. The SNP rs1046329 ($p = 1.37 \times 10^{-5}$) in the gene SGCA is the third best. In addition, Table D-1 (Appendix D) shows the complete list of asymptotic P-values for the 34 significant SNPs ($p < 10^{-4}$) along with the corresponding P-values in the separated male and female populations.

In Table 9 it is interesting to notice the widely differing p-values for males and females for some significant ($p < 10^{-4}$) SNPs. For example, for the SNP rs1046329 the female p-value is 2.88×10^{-5} whereas the same SNP is insignificant in males with $p = 0.1835$. Similarly, the SNP rs4342432 is significant ($p < 10^{-3}$) in females ($p = 0.0001645$) but not in males ($p = 0.1475$).

Table 9

Genome-Wide Association Study Results of Significant SNPs for Neuroticism in Major Depressive Disorder *

CHR	Reference SNP	Position[a]	Gene Name	Gene Location	P-value Total[b]	P-value Male[c]	P-value Female[d]
19	rs4806846	2361484	TMPRSS9	19p13.3	7.79E-06	0.001109	0.002405
12	rs220549	13828587	GRIN2B	12p12	1.05E-05	0.05337	0.0002437

Table 9 (Continued)

17	rs1046329	45603905	SGCA	17q21	1.37E-05	0.1835	2.88E-05
2	rs6757820	147996163	ACVR2A	2q22.3	1.51E-05	0.003489	0.002758
6	rs4510687	96219582	(MANEA & FUT9)	6q16.1	2.49E-05	0.1229	0.0001593
3	rs985280	64083893	PSMD6	3p14.1	2.61E-05	0.07875	0.0009192
6	rs4342432	96195874			2.96E-05	0.1475	0.0001645
16	rs9807002	26184920			3.05E-05	0.1409	0.000289
14	rs9323020	21228477	Near EDDM3B	14q11.2	3.29E-05	0.06499	0.0001355
17	rs8068962	9254866	STX8	17p12	3.32E-05	0.05502	0.001416
2	rs2344734	152400075	NEB	2q22	3.61E-05	0.1872	0.0005467
2	rs12620464	101396078			4.09E-05	0.04817	0.0006929
8	rs1608361	123732642	LOC100131552	8q24.13	4.13E-05	0.08719	0.004728
2	rs10930046	162846229			4.68E-05	0.06334	0.001034
5	rs6894463	3266546			4.81E-05	0.001539	0.003872
17	rs2301685	32452234			4.94E-05	0.1001	0.002558
20	rs6065392	39870080	ZHX3	20q12	5.02E-05	0.06333	0.001298
11	rs10835855	32165526			5.15E-05	0.06333	0.001298
8	rs16915079	51685259	Near CYCSP22	8q11.22	5.54E-05	0.03473	0.006221
3	rs9834457	69604258	Near LOC642487	3p14.1	6.10E-05	0.02482	0.000309

[*] CHR = Chromosome Number; SNP = Single Nucleotide Polymorphism

[a] Physical position is based on NCBI genome build 36.3.

[b] Wald test asymptotic P-value for the association of neuroticism with MDD in the total sample.

[c] Wald test asymptotic P-value for the association of neuroticism with MDD in the male sample.

[d] Wald test asymptotic P-value for the association of neuroticism with MDD in the female sample.

In addition, those significant SNPs ($p < 10^{-1}$) that are significant for ASPD in the OZALC study that share the same genomic regions as the significant ($p < 10^{-1}$) neuroticism SNPs are

shown in Table 10. Accordingly, 8 SNPs in the region of the significant gene markers (GRIN2B, OR10G2, OR4E2, and RNASE1 on chromosomes 12 and 14) for association of neuroticism in the NESDA sample were found with borderline associations for ASPD in the Australian sample at a P-value of 0.09 or lower. Table D-4 shows the complete list of ASPD P-values for SNPs in the region of implicated genes for neuroticism.

Table 10

Genome-Wide Association Study Results of Significant SNPs for ASPD in the Region of Implicated Genes for Neuroticism [*]

CHR	Reference SNP	Gene Name	P-Neoneu Total[a]	P-Neoneu Male[b]	P-Neoneu Female[c]	P-MDD-Total[d]	P-MDD-Male[e]	P-MDD-Female[f]	OZ-ASPD Sample P-value[g]
12	rs11055608	GRIN2B	0.9412	0.8302	0.5511	0.7588	0.3476	0.2884	0.06706
12	rs2216127	GRIN2B	0.08245	0.4308	0.07982	0.0391	0.1045	0.1781	0.04839
12	rs7974275	GRIN2B	0.865	0.9641	0.4977	0.4612	0.7569	0.2365	0.01938
14	rs10146821	OR10G2	0.08711	0.07005	0.7192	0.5681	0.3188	0.8758	0.04173
14	rs2874103	OR4E2	0.1635	0.8785	0.05961	0.2232	0.3295	0.3568	0.08020
14	rs970382	OR4E2	0.1826	0.8555	0.06383	0.2245	0.3046	0.3741	0.09011
14	rs718433	Near RNASE1	0.7877	0.643	0.5419	0.8413	0.7324	0.5755	0.05733
14	rs2141971	Near RNASE1	0.7169	0.2254	0.3192	0.3974	0.7782	0.1487	0.00448

[*] CHR = Chromosome Number; SNP = Single Nucleotide Polymorphism; MDD = Major Depressive Disorder; ASPD = Anti-Social Personality Disorder.

[a, b, c] Wald test asymptotic P-value for the association of neuroticism with MDD in the total, male, and female sample.

[d, e, f] Wald test asymptotic P-value for the association with MDD in the total, male, and female sample.

[g] Wald test asymptotic P-value for the association with ASPD in the OZALC sample.

Finally, Table 11 shows the top 20 significant ($p < 10^{-4}$) gene x gender interaction SNPs for neuroticism along with the corresponding Wald asymptotic P-values in the total, male-only and female-only population in our sample discussed in Chapter 3. Table D-2 gives the complete list of significant ($p < 10^{-4}$) SNPs. Top three significant ($p < 10^{-5}$) SNPs are rs243012 ($p = 5.37x10^{-6}$) in the gene HMCN1, rs17674783 ($p = 6.92x10^{-6}$) in the gene SNTG1 and rs2612437 ($p = 9.30x10^{-6}$) in the intergenic region between the genes MAP2K4 and MYOCD. These findings are further elaborated in Chapter 5 in terms of any gene functionality and possibly other disease associations available from the literature.

Table 11

Genome-Wide Association Study Results of Significant SNPs for Gene x Gender Interaction of Neuroticism*

CHR	Reference SNP	Gene Name	Gene Location	P-assoc Total[a]	P-assoc Male[b]	P-assoc Female[c]	P_GxE
1	rs2430132	HMCN1	1q25.3-q31.1	0.351	0.4283	0.6643	5.37E-06
8	rs17674783	SNTG1	8q11-q12	0.8544	0.5047	0.999	6.92E-06
17	rs2612437	(MAP2K4 & MYOCD)		0.9471	0.526	0.5177	9.30E-06
12	rs11614157	(PRP1R12A & C12orf64)		0.0612	0.3835	0.2374	1.48E-05
4	rs244052	ENPEP	4q25	0.2467	0.9248	0.4139	1.53E-05
8	rs391798	SNTG1	8q11-q12	0.9659	0.483	0.8758	1.53E-05
4	rs13146994	ATP8A1	4p14-p12	0.3128	0.2118	0.4046	1.66E-05
20	rs6072394			0.2119	0.9787	0.3763	1.77E-05
3	rs9834626	RPL10AP6	3p14.2	0.6774	0.454	0.5157	1.78E-05
15	rs870335	SMAD3	15q22.33	0.9354	0.4891	0.411	2.28E-05
20	rs2326424	ADRA1D	20p13	0.04644	0.01311	0.2266	2.69E-05
14	rs17111534	LOC100288480	14q11.2	0.3991	0.9245	0.1688	2.78E-05
2	rs12692083	ISCA1P6	2q14.3	0.7659	0.2809	0.09403	2.83E-05
15	rs2415037	SMAD3	15q22.33	0.8592	0.5387	0.4015	3.00E-05
15	rs1532758	SMAD3	15q22.33	0.2269	0.973	0.07676	3.50E-05
8	rs1023979			0.2049	0.377	0.5902	3.65E-05
6	rs11970411	TNFA1P3	6q23	0.9345	0.8628	0.8696	3.66E-05
5	rs6595727			0.1771	0.02356	0.9237	3.68E-05
8	rs4412393			0.6117	0.06589	0.3106	3.75E-05

Table 11 (Continued)							
8	rs7832224	EFR3A	8q24.22	0.2561	0.8041	0.2023	4.21E-05

[*] CHR = Chromosome Number; SNP = Single Nucleotide Polymorphism; MDD = Major Depressive Disorder;

[a, b, c] Wald test asymptotic P-value for the association of neuroticism with MDD in the total, male, and female sample.

Genome-Wide Association Analysis of Age at Onset of Major Depressive Disorder

The top 20 significant ($p < 10^{-4}$) SNPs associated with age at onset in MDD cases are tabulated below in Table 12. The highlighted portions in the table show SNPs that are significant at the genome-wide level (i.e., p-value < $5x10^{-7}$). These are – rs734253 in the gene GPR143 ($p = 2.54x10^{-12}$), rs6641545 in the gene ASS1P4 ($p = 9.75x10^{-10}$), rs5983118 in the gene MXRA5 ($p = 5.67x10^{-8}$) and rs6654462 flanked by the genes MAGEC1 and MAGEC2 ($p = 3.99x10^{-7}$). In addition we have found three more highly significant ($p < 10^{-5}$) SNPs at the 10^{-6} level which are – rs32589 ($p = 4.72x10^{-6}$), rs9294523 ($p = 5.83x10^{-6}$), and rs10515638 ($p = 9.77x10^{-6}$). The complete list of associated ($p < 10^{-4}$) SNPs for age at onset association is shown in Table D-5 in Appendix D. In chapter 5, we elaborate more on the implicated genes for age at onset in terms of any known functionality, association with other known diseases in both animal and human models.

Table 12

Genome-Wide Association Study Results of Significant SNPs for Age at Onset of MDD*[*]

CHR	Reference SNP	Gene Name	Gene Location	P-MDD-Total[a]	P-MDD-Male[b]	P-MDD-Female[c]	P-Neuroticism[d]	P-AAO[e]
23	rs734253	**GPR143**	Xp22.3	0.3747	N/A	0.8116	0.2357	**2.54E-12**
23	rs6641545	**ASS1P4**	Xp22.33	0.7671	0.2733	0.2551	0.64	**9.75E-10**
23	rs5983118	**MXRA5**	Xp22.33	0.398	N/A	0.1451	0.5025	**5.67E-08**

Table 12 (Continued)

23	rs6654462	**Flanked by MAGEC1 and MAGEC2**	Xp26-Xp27	0.3369	0.7451	0.1007	0.1432	**3.99E-07**
5	rs32589	PPARGC1B	5q32	0.7074	0.0051	0.0168	0.5262	4.72E-06
6	rs9294523	RPL5P19	6q15	0.5052	0.0952	0.0459	0.6942	5.83E-06
5	rs10515638	PPARGC1B	5q32	0.8893	0.0133	0.1193	0.2302	9.77E-06
5	rs26125	PPARGC1B	5q32	0.494	0.0068	0.0066	0.5332	1.04E-05
23	rs5939057	ZCCHC16	Xq23	0.284	0.5247	0.1672	0.6689	1.34E-05
5	rs17110375	PPARGC1B	5q32	0.62	0.0716	0.0686	0.0779	1.42E-05
6	rs6570806	N/A	N/A	0.1601	0.0838	0.0129	0.8842	3.23E-05
8	rs16938568	RDH10	8q21.11	0.6015	0.1612	0.9382	0.1362	3.36E-05
12	rs11107320	LOC100132126	12q22	0.4294	0.2577	0.1283	0.1142	3.41E-05
11	rs3016415	GRIK4	11q22.3	0.6102	0.0176	0.0251	0.8042	3.50E-05
6	rs10947795	KCNK16	6p21.2-p21.1	0.5245	0.0112	0.4695	0.2657	3.71E-05
22	rs34074034	TTLL12	22q13.31	0.2529	0.0134	0.9774	0.4047	3.85E-05
5	rs11738269	CANX	5q35	0.8539	0.0293	0.1304	0.2636	3.95E-05
6	rs1876155	SIM1	6q16.3-q21	0.1498	0.0672	0.5046	0.9709	3.97E-05
15	rs11630901	RPAP1	15q15.1	0.1481	0.0037	0.9168	0.5768	4.07E-05
8	rs17211245	RDH10	8q21.11	0.3178	0.1653	0.6283	0.0366	4.70E-05

[*] CHR = Chromosome Number; SNP = Single Nucleotide Polymorphism; MDD = Major Depressive Disorder;

[a, b, c] Wald test asymptotic P-value for the association with MDD in the total, male, and female sample.

[d] Wald test asymptotic P-value for the association of neuroticism with MDD in the total sample.

[e] Wald test asymptotic P-value for the association of age at onset with MDD in the total sample.

CHAPTER 5

DISCUSSION

Some of the expected phenotypic dependencies of MDD are discussed in the context of our study findings and compared to the literature. The results of our genome-wide studies of MDD in the Netherlands-based sample are discussed and the newly discovered significant SNPs for MDD, neuroticism and age at onset of MDD will be considered. The study hypotheses presented in Chapter 1 are considered according to our study results. A summary conclusion is made and recommendations for future studies and directions for research into MDD using GWAS are given.

Risk Factors of Major Depressive Disorder

Findings from the Univariate Analyses

The population used for univariate analyses contained 1,150 men and 2,206 women, 1,738 cases and 1,618 controls. Of these, 530 men and 1,208 women had MDD affection. There were 1,054 current smokers and 2,279 nonsmokers in the sample with a statistically significant risk for smokers to be diagnosed with MDD and smoking as a risk factor for MDD has been well reported in the literature (Husky et al., 2008).

A significant majority (73.77%) of the sample is identified as current users of alcohol, but alcohol is not seen as a predictor for MDD from the multiple logistic regression analyses that were performed later (details in subsequent sections) on the sample. Most of the subjects in the study sample are either married or living with a partner while a minority reported a status of being single or living without a partner. The association of having familial risk for MDD in the study sample is significantly more (85.58%) as compared with those without any reported

familial risks for MDD (14.42%). The presence of a first degree family member affected with a depression disorder was identified using the Family Tree Assessment (Fyer & Weissman, 1999) and was a part of the CIDI-based interview that was conducted on the subjects of the sample during the data gathering process. The descriptive characteristics of cases and controls summarized earlier in Chapter 3 in Tables 1 and 2 generally agree with the results of our univariate analyses.

Findings from the Bivariate Analyses

Our study of regressing MDD as a binary trait using logistic regressions against selected predictors seems to be one of the few in the literature as no other study has chosen the seven predictors we have chosen, namely age, gender, marital status, smoking status, alcohol use, family risk, and NEO score. Most other studies have concentrated on age, gender, marital status as classic demographic variables along with such other socioeconomic variables as family income, educational level, and living styles. Most studies using the dbGaP data seem to be restricted to genotypic studies and not necessarily phenotypic in nature. Our study might be one of the first to look at the epidemiology of MDD using the dbGaP data.

Our results support the findings of prior studies that female gender, smoking, and having an affected first degree relative are all significant risk factors for being afflicted with MDD. The association of family risk with an OR of 3.112 confirms the strong association of MDD within families and that MDD is a genetically heritable disease. The association between gender and MDD is shown by the 1.7:1 frequency among women as compared with men and an OR of 1.43 (95% CI: 1.24 – 1.64). This compares well to the literature where Kessler et al. (2003) reported an OR of 1.4 (95% CI: 1.1 – 1.8) and Bromet et al. (2011) who recently reported a total OR of 1.8 (95% CI: 1.6 – 2.0) in high-income countries (e.g., Belgium, France, Germany, Japan,

Netherlands, etc.) and low-income countries (e.g., Brazil, India, Columbia, Mexico, etc.) around the world.

According to the study by Bromet et al. (2011), MDD is a significant public health concern across all regions of the world and is strongly linked to social conditions. Three predictors – age, gender, and marital status – are available as common study predictors for comparison of our results. Comparing their gender results with our European (in particular, the Netherlands population) sample the OR value of our sample (1.43) is much smaller than Bromet et al. study sample which is reported to be 2.3 (95% CI: 1.5 – 3.5). They categorized age into four groups – 18-34, 35-49, 50-64, and 65+ and they also categorized marital status into five groups - never married, separated, divorced, widowed, and currently married. But in our data set, age is a continuous variable and marital status is binary and thus we cannot compare our results directly. Our study must reconsider age as categorical and reformat our data to reflect their four age groups and the regression rerun before we can compare the age as a risk factor. Their ORs for marital status range from 2.6 to 1.9 with '65+' category being the reference at 1.0. Similarly our marital status OR is 3.05, but we need to categorize our marital status and the regression rerun before attempting any comparisons. Their ORs range from 2.7 to 1.8 with currently married category being the reference at 1.0.

MDD is also comorbid with smoking as shown by an OR of 2.88 (95% CI: 2.48 – 3.35) which is consistent with other studies (Husky et al., 2008). Prevalence of depression among cases who have a first degree relative being also diagnosed with some form of MDD is shown to be the highest predictor (OR: 3.45 with 95% CI: 2.42 – 4.92) in our study. The family risk for MDD shows up as a leading predictor for MDD in the study sample from the logistical regression analyses. Several studies have linked chronic MDD to family history of depression

(Camp et al., 2005; Gonda et al., 2006; Lazary et al., 2009). Lazary et al. (2009) use a different phenotype called affective temperament that is measured via special instrument called TEMPS-A (Kesebir et al., 2005). They concluded that a crucial part of inherited factors of depression is mediated by affective temperament.

In addition, both alcohol usage and being married (or living with a partner) showed a moderating effect with respective ORs of 0.498; 95% CI: (0.426-0.583) and 0.327; 95% CI: (0.275 – 0.389) for being diagnosed with MDD. These results may point to the emotional and/or calming (coping) effects of family bonds and alcohol. However, there are also many studies about alcohol usage and MDD (see Chapter 2 for details) that shows an opposite effect.

Findings from the Multiple Variable Analyses

The results in Table 5 show that only age as a risk factor is removed upon multiple logistic regression. The remaining risk factors marital status, smoking status, family risk, alcohol use and NEO score continue to stay significant at $p = 0.028$ or much below that. The corresponding OR values do not change much indicating minimal amounts of confounding among the risk factors.

As shown by Tables 6, 7, and 8 the analyses with interaction effects among the three major predictors, gender, smoking, and family risk show little or no statistical significance. However, the risk factor ‘smoking status’ has been classified as a binary variable and we do not have any knowledge (data) of whether a person’s smoking status was pre- or postdiagnosis of being afflicted with MDD. Thus the major risk factors for MDD from this study that could be ascertained are gender, neuroticism, and family risk.

In terms of looking at our results, the odds ratio of 1.28 and 95% CI of (1.078 – 1.602) gender is a significant risk factor with a modest chi-square value of 4.82 and a p-value of 0.028. Smoking status has an OR of 3.07 (95% CI: 2.45 – 3.87) even after adjustment for other

confounding variables and is a major risk factor of MDD. Looking back at the available literature on smoking and MDD (Chapter 2.2), female gender and prior smoking together increase the risk of MDD. From our results it seems that the smoking status of this particular female population could have been "heavy or daily" as opposed to "occasional" or "prior". This confirms other studies on smoking and MDD that were cited earlier in Chapter 2 (Cardenas et al., 2002; Gulec et al., 2003; Husky et al., 2008). Our results show an OR of 3.07 which is higher than the highest found in the Husky et al. study where the smoking females had an OR value ranging from 1.36 to 2.52 corresponding to "prior user status" to "current daily user" status.

Both marital status (being married or living with a partner) and alcohol use seem to be beneficial as indicated by their negative parameter values. However, from our earlier discussion in Chapter 2 alcohol use has a more complex relationship with MDD. As has been reported in the literature alcohol consumption when moderate could be beneficial to a person's health, but specific studies showing benefits for mental health are not found in the literature. Comparing the OR of marital status from the bivariate analysis to the multiple variable analyses, we see a jump in value from 0.327 to 0.432. While both values are well below 1.0, the confounding seems to be beneficial and further analyses are needed to determine the particular risk factor that confounds marital status as a risk factor for MDD. However, as South and Krueger (2008) point out, it is the "marriage quality", not just being married (which is all the information we have in our data) might not reveal any suggestive relationship between MDD and marital status. South and Krueger further point to environmental influences along with genetic influences that internalize stressors, which in turn affects the psychopathology. Suffice is it to say that simple odds ratios (albeit adjusted for confounding) may not be good indicators of possible risks that a marriage may bring in terms of influencing a person's (esp. a female's) mental disorder. In summary, the

genetic effects on the internalizing factor contribute to the development or worsening of MDD. Further discussion of this complex subject is beyond the scope of this report.

As to be expected, the family history of MDD has a huge bearing on MDD shown by the OR of 3.11 (95% CI: 2.17 – 4.46) which is slightly greater than what we have found for smoking. Again, this association is more complex to be analyzed and results like adjusted odds ratios might not reflect the underlying functional relationships. MDD is known to be heritable (up to 70% as some studies have shown) and both twin and family studies have contributed heavily and will continue to do so in the future as genetic studies unearth more specific genetic regions and SNPs where the predisposition to MDD may lie and how any pattern of inheritance could be mapped. Multiple variable analyses may not throw any more light on this risk factor than the simple bivariate analysis discussed earlier. However, our initial Hypothesis 1 (stated in Chapter 1) is true as has already been shown in the literature; i.e., being female, smoking, and family history are all major risk factors for MDD.

The effect of NEO Score has been positive but the relationship between MDD and neuroticism was further evaluated using GWAS methods by treating it as an endophenotype or a quantitative trait of MDD. The results from our study are further discussed in the next section.

Genetic Association Analyses

We conducted genome-wide studies of two MDD related phenotypes, namely, neuroticism and age at onset. Results from each of these studies are discussed separately below.

Genome-Wide Association Analysis of Neuroticism

A quick search in the OMIM and Entrez Gene databases for the implicated genes did reveal some significant gene function or association with some mental disorders. The most significant

association of neuroticism is with the gene TMPRSS9 containing the SNP rs4806846 ($p = 7.79x10^{-6}$) at 19p13.3. This gene has been associated with the regulation of physiological and pathological phenomena on the cell surface and has been implicated in congenital or childhood-onset form of deafness (Hayama et al., 2007). Functionally, Polyserase-1 transcripted from the gene TMPRSS9 has alternative transcripts serase-1B and serase-2B (Okumura et al., 2006). These have been found to be novel type II transmembrane serine proteases which may play a role in regulating physiological and pathological phenomena on the cell surface. In other functional studies, TMPRSS9 has been seen to possess a unique structure with three tandem serine protease domains (serase-1, 2, and 3) (Cal et al., 2003).

The second best SNP rs220549 in the GRIN2B gene at 12p12 has been associated with Schizophrenia (Doi et al., 2009), Bipolar Disorder (Fallin et al., 2005), Alzheimer's disease (Stein et al., 2010) and Obsessive Compulsive Disorder (OCD) (Arnold et al., 2009). Functionally, GRIN2B is associated with Glutamate and N-Methyl D-Aspartic acid (NMDA) receptors and have been found to participate as an excitotoxin for an operant behavior in laboratory animals (Watkins et al., 2006). Both linkage and association studies of GRIN2B have shown strong associations with Schizophrenia and Bipolar Disorder (Doi et al., 2009; Fallin et al., 2005) in Ashkenazi Jewish population samples. A significant ($p = 0.04$) association of GRIN2B has been identified by Arnold et al. (2009) with a decreased glutamatergic concentration in the anterior cingulate cortex (ACC) that has been consistently implicated in OCD. In addition, the NMDA protein encoded by the gene GRIN2B and is involved in learning and memory along with excitotoxic cell death has age-dependent prevalence in the synapse and is already a therapeutic target in Alzheimer's disease (Stein et al., 2010).

The third best neuroticism associated gene SGCA has been associated with limb girdle muscular dystrophy, type 2D (LGMD2D) (Mendell et al., 2010) and embryonal rhabdomyosarcoma in laboratory mice (Fernandez et al., 2010). In terms of association, mutations in the SGCA gene may have caused LGMD2D in patients who were then subjected to successful early phase gene therapy using adeno-associated virus (AAV) gene transfer under control of a muscle specific promoter. Long-term, sustained gene expression of alpha-sarcoglycan may result in an effective treatment for LGMD2D (Mendell et al., 2010). Functionally, the dystrophin-associated glycoprotein (DAG) complex is a group of muscle proteins whose loss gives rise to muscular dystrophy. In laboratory mice lack of alpha-sarcoglycan (expressed by SGCA as a member of the DAG complex) results in a spontaneous development of muscle-derived embryonal rhabdomyosarcoma (Fernandez et al., 2010).

Among other associated SNPs in Table 11, we found the SNP rs8068962 ($p = 3.32x10^{-5}$) in the gene STX8 at 17p12 associated with the Crohn's disease (AceView, 2011). Two other disease associated SNPs were – rs2344734 ($p = 3.61x10^{-5}$) in the gene NEB associated with Nemaline Myopathy (Wallgren-Petterson et al., 2007), and rs6065392 ($p = 5.02x10^{-5}$) in the gene ZHX3 associated with Primary Glomerula disease (Liu, Clement, Kanwar, Avila-Casado, & Chugh, 2006). In fact, Wang et al. (2009) and Wilson et al. (2003) have suggested predictive association of neuroticism with dementia and Alzheimer's disease (AD) respectively. High neuroticism has been associated with greater risk of dementia, while an active and socially integrated lifestyle may modify the risk for dementia. Wang et al. (2009) conducted a prospective study of 506 older people for 6 years to conclude that a combination of high extraversion and low neuroticism is the personality trait required to maintain a low risk for dementia (Wang et al., 2009). Wilson et al. (2003) used NEO-FFI neuroticism score to assess the

disposition to experience psychological distress in a population of older Catholic nuns, priests, and brothers participating in a Religious Orders Study of aging and Alzheimer's disease. Their findings showed the tendency to experience psychological distress predicted development of incident AD and rate of cognitive decline (Wilson et al., 2003).

In addition, as shown by Table 12, several of these significant SNPs in the region showed up as being significant for ASPD in an Australian sample (OZALC study) thus linking neuroticism and antisocial behavior (Ducci et al., 2008; Schmidt et al., 2000; Wang et al., 2011). However, further studies and replications are required before we can establish firm genome-wide associations between neuroticism and anti-social behavior.

An analysis of the statistical power of our study using QUANTO software shows a power of 96% for the main effect using the neuroticism as a quantitative trait at a mean value of 39.3 ± 8.0, an allele frequency of 0.05, and a marginal R^2 of 0.005. The main effect regression coefficient was derived to be 1.8353. Looking back at our Hypothesis 2 (see Chapter 1), we stated that neuroticism as a quantitative trait to conduct genome-wide association analysis may increase the statistical power. A high statistical power of 96% that was achieved for this particular sample population of Dutch origin in the Netherlands makes our hypothesis true. However, many more replication studies are needed using other human populations before we can confirm our hypothesis.

To our knowledge, this is the first GWA analysis of neuroticism as a quantitative trait of MDD. Identification of a common genomic region between neuroticism in the NESDA population and ASPD in the OZALC population may signal possible association between neuroticism and ASPD that needs to be further investigated.

Gene x Gender Interaction Analysis of Neuroticism

According to literature the HMCN1 gene has been associated with gene x environment interaction in patients presenting atherosclerosis and two of the intervening environmental factors have been gender and smoking (McGeachie et al., 2009). Another implication of the gene HMCN1 is in age-related macular degeneration (AMD) that affects the renal functions (Thompson et al., 2007). The authors report that the association of HMCN1 is not phenotype-specific and it appeared to influence the longitudinal rate of change of AMD while also accounting for other risk factors like family history.

The gene SNTG1 encodes a protein specifically expressed in the brain and its transcript variants have not been fully described, nor has their full-length nature been determined as of this writing. The gene has also been implicated in personalized smoking cessation studies (Rose, Behm, Drgon, Johnson, & Uhl, 2010) in terms of affecting the dosage of nicotine as an environmental variable. The third most significant SNP rs2612437 ($p = 9.3x10^{-6}$) is in the region between genes MAP2K4 and MYOCD. The gene MAP2K4 encodes a dual specificity protein kinase which is an activator of receptors to stressful signals from the environment, while the gene MYOCD encodes a nuclear protein expressed in heart, aorta and other smooth muscle containing tissues.

Thus so far, no specific or direct role of the implicated gene x gender interaction of neuroticism genes have been connected to MDD, but they all seem to have some functional role in the gene x environment interaction effects. Further studies and more functional analyses are required to determine the specific roles of these genes. Detection of gene x gender interactions provides a preliminary basis for the second part of Hypothesis 2 (stated in Chapter 1). However, we again caution the need for replication studies using other samples before we could generalize

our results as a determinant for the observed preponderance of higher prevalence of MDD in females. In fact, the study of gene x gene and gene x environment interactions in GWAS is extremely important before declaring any compelling associations.

In a perspective written by The Psychiatric GWAS Consortium Steering Committee (2009), the authors recognize a new type of study called "integrated mega-analysis" that is different from the conventional meta-analysis. Mega-analysis uses individual-level genotype and phenotype data as opposed to summary data (such as Odds Ratios used in meta-analyses) and integrated mega-analyses of all GWAS data might lead us toward the etiology of psychiatric diseases such as MDD. According to the committee a conventional definition of MDD with the identification of one or more compelling associations with a SNP, haplotype, or copy number variant could become the "holy grail" of psychiatric genetics.

Genome-Wide Association Analysis of Age at Onset of Major Depressive Disorder

The SNP rs734253 in gene GPR143 was the best SNP with the strongest p-value at $p = 2.54 \times 10^{-12}$ along with three more genome-wide significant SNPs - rs6641545 in the gene ASS1P4, rs5983118 in the gene MXRA5, rs6654462 flanked by the genes MAGEC1 and MAGEC2. In addition, the gene PPARGC1B contained four SNPs (rs32589, $p = 4.72 \times 10^{-6}$; rs10515638, $p = 9.77 \times 10^{-6}$; rs26125, $p = 1.04 \times 10^{-5}$; rs17110375, $p = 1.42 \times 10^{-5}$) of high significance ($p < 10^{-4}$) along with the gene GRIK4 which contained one SNP rs3016415 also at high significance ($p = 3.5 \times 10^{-5}$).

While the top four SNPs have been found to be of GWAS significance, the associated genes are novel in that none of them have been implicated in any specific disease association earlier. Some passing observational references have been made in the gene databases, such as, the MAGEC1/2 genes are seen in human melanoma and the GPR143 has been implicated in Ocular

albinism. But, no direct association with MDD or any other mental disease has been made for these novel genes.

A quick search in the gene databases revealed that the gene PPARGC1B has been associated with obesity, metabolism, and body energy regulation (e.g., performance of athletes) and the gene GRIK4 which has been implicated in neuronal signaling (Joslyn, Ravindranathan, Brush, Schuckt, & White, 2010) and in an individual's performance on standardized tests (Cirulli et al., 2010). The gene GRIK4 has been implicated for strong remission of depression in depressive patients undergoing anti-depressant therapy (Horstman et al., 2009). Functionally, GRIK4 gene encodes a protein that belongs to the glutamate-gated ion channel family. Glutamate functions as major excitatory neurotransmitter in the central nervous system through activation of ligand-gated ion channels and G-protein coupled membrane receptors.

Many genetic linkage studies have been undertaken in the context of recurrent and early onset MDD and the early age at onset has been considered as below the age of 31. Separate adolescent studies have also been done and thus the current literature seems to have dealt with age at onset under two categories – adolescent MDD and adult MDD.

Many associations have been made in the literature between early or late age at onset of MDD and risky behaviors like alcohol abuse and cigarette smoking (e.g. Sund, Larsson, & Wichstrom, 2011) and a study of brain anatomy (volume of amygdala) of older patients with late-life depression (Burke et al., 2011). However, other than a study by Shi et al (2011) the current literature does not show specific GWAS of AAO in MDD. The Shi study excludes lifetime MDD and also focuses on people who were diagnosed with MDD before the age of 31 years.

Thus, the current study reported in this dissertation seems to be first of its kind (population-based study of age at onset of both adolescent and adult MDD) with a large 18-65 years MDD cohorts which has found four novel genes (GPR143, ASS1P4, MXRA5 and two SNPs between MAGEC1 and MAGEC2) with GWAS significance. It is hoped that future studies will be able to replicate these GWAS significant MDD implicating genes in other populations in order that AAO as a quantitative trait of MDD can become a basis for further elucidation of MDD and its genetic variants.

The above finding of novel genetic variants for MDD using age at onset as a quantitative trait means that our Hypothesis 3 (stated earlier in Chapter 1) is true. Again, we hasten to caution that replication in other populations and specific functional associations need to be ascertained before implicating these novel genes for MDD.

Strengths of the Study

The main strength of this study is that it is the first GWA study of the MDD related quantitative traits of neuroticism scores and age at onset of MDD. Next, we have looked at the gene x gender interactions of neuroticism associated with MDD cases and have unearthed some novel genes and associated SNPs by the analysis of age at onset as a quantitative trait. These new genes and their associated genomic regions can further be examined for functional significance and other tag SNPs that may have been reported by other researchers for various mental disorders and other complex diseases.

The gender differences we have found in the differing significant SNPs ($p < 10^{-4}$) for Neuroticism in men and women with MDD and the clear interactions between gene and gender for the mean NEO scores (see Figure 5 in Appendix C) show the strength of using quantitative

traits in GWAS. Elucidating the genetic basis of MDD by exploring other quantitative phenotypes of MDD might lead to a narrowing of the set of genes that might be implicated for MDD.

Discovery of novel genetic variants of MDD by the use of another quantitative trait (namely, age at onset of MDD) points to possible first steps in understanding the etiology of MDD in adolescents and young adults. Knowing the fact that young adults in the age group of 18-29 (Kessler et al., 2003) have one of the highest ORs (3.0) for MDD, understanding the genes implicated by the age at onset of MDD further is an important public health undertaking in order to craft early interventional strategies for that group of young adults. Furthermore, finding common and significant ($p < 0.05$) genetic variants in two different populations (as we did for neuroticism in the NESDA and ASPD in the Australian sample) might lead us to common biological functional underpinnings that might one day help us in identifying newer interventional strategies based on the specific functional proteins and signaling pathways.

While we are still years away from elucidating the etiology of MDD, adding new and significant biomarkers (SNPs and genetic regions) to the existing databases may pave the way for detecting and/or understanding of how genes can affect the development of MDD in vulnerable populations.

As we have seen from our limited results of MDD vulnerability, both gender and family risks must be studied further to look for possible public health intervention strategies to prevent early onsets of MDD, progression from temporal to lifelong suffrage of MDD, and suicidal deaths from MDD. Genetic and genome-wide association analyses are only the first step in terms of using the power of newer genetic and bioinformatics technologies in order to enhance the epidemiological methods available for combating MDD in our society.

Limitations of the Study

A number of limitations must be recognized in this study. This is a preliminary study looking at neuroticism as an endophenotype for MDD and as such needs to be replicated in several other samples. Other than showing some associations at suggestive level ($p < 5x10^{-4}$), not all implicated SNPs reach the GWA significance level ($p < 5x10^{-7}$). Our findings must be interpreted with caution because of limitations of sample size and power.

While the initial phenotypic analyses using logistic regression methods confirmed the major risk factors for MDD (female gender, marital status, etc.) further work needs to be done by categorization of certain risk factors as age (in terms of specific age groups), marital status (in terms of specific statuses as being single, widowed, same sex partner, etc.), and alcohol use (in terms of teetotaler, casual drinker, moderate consumer, heavy drinker, and abuser). The alcohol and/or tobacco use as a risk factor for any medical condition are in general viewed as being contentious and both risk factors are highly complex to analyze. Our study does not adequately address these two risk factors for MDD.

Genome-Wide studies have been criticized for not being able to find main effects of genes on complex diseases such as MDD. Thus, this study must be expanded to include investigations of specific vulnerable genetic regions (esp. within families) by using the latest New Generation Sequencing (NGS) technologies, albeit with added expense to procure new equipment and design new strategies to limit research for functional understanding of the known genes.

Conclusions and Further Recommendations

Our results showed female gender, family history of MDD, and smoking as major risk factors for MDD, findings that are consistent with past studies. We found several significant SNPs (at $p < 10^{-4}$) for the association of neuroticism in MDD and significant gene x gender interactions

implicating several new genes – TMPRSS9, GRIN2B, SGCA, HMCN1 and SNTG1. More interestingly, we found four novel genes at genome-wide significance ($p < 5x10^{-7}$) for age at onset association with MDD. These novel genes are – GPR143 ($p = 2.54x10^{-12}$), ASS1P4 ($p = 9.75x10^{-10}$), MXRA5 ($p = 5.67x10^{-8}$), and the genomic region between MAGEC1 and MAGEC2 ($p = 3.99x10^{-7}$).

Much work remains to be done in the elucidation of genetic variations and the etiology of MDD as a complex and chronic disease that affects millions of people around the world and poses as a major global burden of disease. The results of our study contribute to expanding the genetic basis of MDD. Replication studies in other population in order to elucidate the genetic variants of MDD are needed.

A few recommendations for further study are in order –

- Conduct replication studies of other non-European populations that are vulnerable to MDD using neuroticism and age at onset as quantitative traits.
- Discover other mental disorders that have same or similar genomic regions as what we have found and study the interrelations and gene functionality that might explain further biological mechanisms that make it possible for MDD to be inheritable in and across families at rates of 40%-70%.
- Use Haploview and other modern genetic software tools to find the tag SNPs from the significant SNPs found in our study.
- Procure funding for exploring the pertinent genomic regions significant for neuroticism and age at onset using Next Generation Sequencing of human chromosomes 1, 8, 12, 17, 19, and 23 (or X chromosome).

REFERENCES

Angst J., Angst F., & Stassen H.H. (1999). Suicide risk in patients with major depressive disorder. *J Clin Psychiatry 60*. 57-62.

Anderson N.L., & Anderson N.G. (1998). Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis 19(11).* 1853–1861.

Aragam N.R., Wang K.S., & Pan Y. (2011). Genome-Wide association analysis of gender differences in major depressive disorder in the Netherlands NESDA and NTR population-based samples. *Journal of Affective Disorders 133.* 516-521.

Arnold P.D., MacMaster F.P., Richter M.A., Hanna G.L., Sicard T., Burroughs E., Mirza Y..., & Rosenberg D.R. (2009). Glutamate receptor gene (GRIN2B) associated with reduced anterior cingulated glutamatergic concentration in pediatric obsessive-compulsive disorder. *Psychiatry Res 172*. 136-139.

APA, American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders,* (4th ed) (DSM-IV). Washington, DC.

Aube J., Fichman L., Saltaris C., & Koestner R. (2000). Gender differences in adolescent depressive symptomatology: Towards an integrated social-developmental model. *Journal of Social and Clinical Psychology 19*. 297-313.

Balding D.J. (2006). A tutorial on statistical methods for population association studies. *Nat Rev Genet 7*. 781-791.

Barrett J.C., Fry B., Maller J., & Daly M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics 2*. 263-265.

Barrett J.C., & Cardon L.R. (2006). Evaluating coverage of genome-wide association studies. *Nat Genet 38*. 659-662.

Bashiardes S., Veile R., Allen M., Wise C.A., Dobbs M., Morcuende J.A., ... Lovett M. (2004). SNTG1, the gene encoding Y1-syntrophin: a candidate gene for idiopathic scoliosis. *Hum Genet 115*. 81-89.

Baum A.E., Akula N., Cabanero M., Cardona I., Corona W., Klemens B., ... McMahon F.J. (2008). A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol. Psychiatry 13.* 197-207.

Belson W.A. (1981). *The design and understanding of survey questions*. *Aldershot: Gower*.

Benjamin J., Ebstein R.P., Lesch K.P. (1998). Genes for personality traits: Implications for psychopathology. *Int J Neuropsychopharmacol 1*. 153-168.

Bennett D. (2005). Growing pains for metabolomics. *The Scientist* 19(8). 25-28.

Bienvenu O.J., Brown C., Samuels J.F., Liang K.Y., Costa P.T., Eaton W.W., & Nestadt G. (2001). Normal personality traits and comorbidity among phobic, panic and major depressive disorders. *Psychiatry Res 102*. 73-85.

Bilder R.M., Sabb F.W., Cannon T.D., London E.D., Jentsch J.D., Stott-Parker D., ... Freimer N.B. (2009). Phenomics: The systematic study of phenotypes on a genome-wide scale. *Neuroscience 164*. 30-42.

Bilgi M.M., Ozalay O., Eker M.C., Kitis O., Ozan E., Cagdas-Eker O., ... Gonul A.S. (2010). Small frontal gray matter volume in first-episode depression patients. *Turkish Journal of Psychiatry 2010.* 1-9.

Birley A.J., Gillespie N.A., Heath A.C., Sullivan P.F., Boomsma D.I., et al. (2006). Heritability and nineteen-year stability of long and short EPQ-R Neuroticism scales. *Personality and Individual Differences 40.* 737-747.

Bochdanovits Z., Verhage M., Smit A.B., de Geus E.J., Posthuma D., Boomsma D.I., ... Heutink P. (2009). Joint reanalysis of 29 correlated SNPs supports the role of PCLO/Piccolo as a causal risk factor for major depressive disorder. *Mol Psychiatry 14.* 650-652.

Bogunovic D., O'Neill D.W., Belitskaya-Levy I., Vacic, Yu Y.L., Adams S., ... Bhardwaj N. (2009). Immune profile and mitotic index of metastatic melanoma lesions enhance clinical staging in predicting patient survival. *PNAS 106.* 20429-20434.

Boomsma D.I., Beem A.L., van den Berg M., Dolan C.V., Koopmans J.R., Vink J.M., ... Slagboom P.E. (2000). Netherlands twin family study of anxious depression (NETSAD). *Twin Research 3*. 323-334.

Boomsma D.I., Willemsen G., Sullivan P.F., Heutink P., Meijer P., Sondervan D., ... Penninx B.W.J. H. (2008). Genome-Wide association of major depression: description of samples for the GAIN Major Depressive Disorder study: NTR and NESDA biobank projects. *European Journal of Human Genetics 16*. 335-342.

Bouchard T.J., Jr., Loehlin J.C. (2001). Genes evolution and personality. *Behav Genet 31*. 243-273.

Bromet E., Andrade L.H., Hwang I., Sampson N.A., Alonso J., de Girolamo G., ... Kessler R.C. (2011). Cross-national epidemiology of DSM-IV major depressive episode. *MMC Medicine 9*. 90.

Brummett B.H., Boyle S.H., Kuhn C.M., Siegler I.C., Williams R.B. (2008). Associations among central nervous system serotonergic function and neuroticism are moderated by gender. *Biol Psychol 78.* 200-203.

Bulik C.M., Sullivan P.F., Tozzi F., Lichtenstein P., & Pedersen N.L. (2006). Prevalence, heritability, and prospective risk Factors for Anorexia Nervosa. *Arch Gen Psychiatry 63*. 305-313.

Burke J., McQuoid D.R., Payne M.E., Steffens D.C., Krishnan R.R., & Taylor W.D. (2011). Amygdala volume in late-life depression: relationship with age at onset. *Am J Geriatr Psychiatry 19.* 771-776.

Cal S., Quesada V., Garabaya C., & Lopez-Otin C. (2003). Polyserase-1, a human polyprotease with the ability to generate independent serine protease domains from a single translation product. *PNAS 100.* 9185-9190.

Calboli F.C.F., Tozzi F., Galwey N.W., Antoniades A., Mooser V., Preisig M., ... Balding D.J. (2010). A Genome-Wide association study of neuroticism in a population-based sample. *PLoS One 5(7).* 1-7.

Camp N.J., Lowry M.R., Richards R.L., Plenk A.M., Carter C., Hensel C.H., ... Cannon-Albright L.A. (2005). Genome-Wide linkage analyses of extended Utah pedigrees identify loci that influence recurrent, early-onset major depression and anxiety disorders. *American Journal of Human Genetics Part B (Neuropsychiatric Genetics) 135B.* 85-93.

Cannell C.F., Miller P.V., Oksenberg L. (1981). Research on interviewing techniques. In S Leinhardt (Ed.) *Sociological methodology*. San Francisco: Josey-Bass. 389-437.

Cardenas L., Tremblay L.K., Naranjo C.A., Herrman N., Zack M., Busto U.E. (2002). Brain reward system activity in major depression and comorbid nicotine dependence. *The Journal of Pharmacology and Experimental Therapeutics 302.* 1265-1271.

Centers for Disease Control and Prevention. (n.d.). IOM-2005. From, http://www.cdc.gov/genomics/about/reports/2005/letter.htm

Chanok S.J., Manolio T., Boehnke M., Boerwinkle E., Hunter D.J., Thomas G., ... Abecasis G. (2007). Replicating genotype-phenotype associations. *Nature 447.* 655-660.

Cheeseman I.M., Niessen S., Anderson S., Hyndman F., Yates III JR., Oegema K., & Desai A., (2004). A conserved protein network controls assembly of the outer kinetochore and its ability to sustain tension. *Genes & Development 18*. 2255-2268.

Cirulli E.T., Kasperaviciute D., Attix D.K., Need A.C., Ge D., Gibson G., & Goldstein D.B. (2010). Common genetic variation and performance on standardized cognitive tests. *European Journal of Human Genetics 18*. 815-820.

Clark H.H., & Schober M.F. (1992). Asking questions and influencing answers. In JM Tanur (Ed.) *Questions about questions: Inquiries into the cognitive bases of surveys.* New York: Russell Stage Foundation. 15-48.

Cohen J. (1992). A power primer. *Psychol Bull 112*. 155-159.

Cordell H.J., & Clayton DG. (2005). Genetic Epidemiology 3: Genetic association studies. *Lancet 366*. 1121-1131.

Costa P.T., & McCrae R.R. (1992). *Revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI).* Odessa, FL: Psychological Assessment Resources.

Costa P.T., Jr., & McCrae R.R. (1995). Domains and facets: hierarchical personality assessment using the revised NEO personality inventory. *J Pers Assess 64*. 21-50.

Costa P.T., & McCrae R.R. (2006). Age changes in personality and their origins: Comment on Roberts, Walton, and Viechtbauer (2006). *Psychological Bulletin 132(1).* 26-28.

Cyranowski J., Frank E., Young E., Shear K. (2000). Adolescent onset of the gender difference in lifetime rates of major depression: A theoretical model. *Archives of General Psychiatry 57*. 21-27.

Daly R.J., Sanderson G.M., Janes P.W., & Sutherland R.L., 1996. Cloning and characterization of GRB14, a novel member of the GRB7 gene family. *The Journal of Biological Chemistry 271.* 12502-12510.

Diabetes Genetics Initiative. (2007). Genome-Wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science 316.* 1331-1336.

Doi N., Hoshi Y., Itokawa M., Usui C., Yoshikawa T., & Tachikawa H. (2009). Persistence criteria for susceptibility genes for Schizophrenia: A discussion from an evolutionary viewpoint. *PLoS one 4.* e7799.

Ducci F., Enoch M.A., Hodgkinson C., Xu K., Catena M., Robin R.W., & Goldman D. (2008). Interaction between a functional MAOA locus and childhood sexual abuse predicts alcoholism and antisocial personality disorder in adult women. *Mol Psychiatry 13*. 334-347.

Eaves L.J., Heath A.C., Neale M.C., Hewitt J.K., & Martin N.G. (1998). Sex differences and non-additivity in the effects of genes on personality. *Twin Res 1*. 131-137.

Elston R.C., Igo R., & Cartier K.C. (2010). Statistical analysis for genetic epidemiology (S.A.G.E.): Short Course. *School of Medicine, Case Western Reserve University*. August 15-18, 2010. Cleveland, OH.

Essau C.A., Lewinsohn P.M., Seeley J.R., Sasagawa S. (2010). Gender differences in the developmental course of depression. *J Affect Disord 127.* 185-190.

Eysenck S.B.G., Eysenck H. (1975). *Manual of the Eysenck personality questionnaire (junior and adult).* London: Holder and Stoughton.

Fanous A., Gardner C.O., Prescott C.A., Cancro R., & Kendler K.S. (2002). Neuroticism, major depression and gender: a population-based twin study. *Psych Med 32*. 719-728.

Farrer L.A., Cupples L.A., Haines J.L. et al. (1997). Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer's disease. A meta-analysis. APOE and Alzheimer's Disease Meta-Analysis Consortium. *JAMA 278*. 1349-1356.

Floderus-Myrhed B., Pedersen N., & Rasmuson I. (1980). Assessment of heritability for personality based on a short-form of the Eysenck personality inventory: A study of 12,898 twin pairs. *Behav Genet 10*. 153-162.

Frayling et al. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science 316*. 889-894.

Garriock H.A., Kraft J.B., Shyn S.I., Peters E.J., Yokoyama J.S., Jenkins G.D., ... Hamilton S.P. (2010). A genomewide association study of citalopram response in major depressive disorder. *Biol Psychiatry 67*. 133-138.

Gibbs R.A. et al. (2003). The international HapMap project. *Nature 426*. 789-796.

Golster-Dubner T., Galili-Weistubb E., & Segman R.H. (2010). Genetics of unipolar major depressive disorder. *Isr J Psychiatry Relat Sci 47(1).* 72-82.

Gonda X., Rihmer Z., Zsombok T., Bagdy G., Akiskal K.K., & Akiskal H.S. (2006). The 5-HTTLPR polymorphism of the serotonin transporter gene is associated with affective temperaments as measured by TEMPS-A. *Journal of Affective Disorders 91*. 125-131.

Goodwin R.H., & Gotlib I.H., 2004. Gender differences in depression: The role of personality factors. *Psychiatry Research 126*. 135-142.

Goodyer I., Herbert J., Tamplin A., & Altham P.M. (2000). Recent life events, cortisol, dehydroepiandrosterone, and the onset of major depression in high-risk adolescents. *British Journal of Psychiatry 177*. 499-504.

Gotlib I.H., Lewinsohn P.M., & Seely J.R. (1995). Symptoms vs. a diagnosis of depression: Differences in psychosocial functioning. *Journal of Consulting and Clinical Psychology 63*. 90-100.

Grant B.F., Dawson D.A., Stinson F.S., Chou P.S., Kay W., & Pickering R. (2003). The alcohol use disorder and associated disabilities schedule (AUDADIS). Reliability of alcohol consumption, tobacco use, family history of depression, and psychiatric diagnostic modules in a general population. *Drug Alcohol Depend 71*. 7–16.

Grant J.D., Agrawal A., Bucholz K.K., Madden P.A. et al. (2009). Alcohol consumption indices of genetic risk for alcohol dependence. *Biological Psychiatry 66*. 195-800.

Guerrera R.J., Dickey C.C., Niznikiewicz M.A., Voglmaier M.M., Shenton M.E., & McCarley R.W. (2005). The five-factor model in schizotypal personality disorder. *Schizophr Res 80.* 243-251.

Guessous I., Gwinn M., Khoury, M.J. (2009). Genome-wide association studies in pharmacogenomics: Untapped potential for translation. *Genome Med 1(4).* 46.

Gulec M., Bakr B., Ozer M., Ucar M., Klc S., & Hasde M. (2005). Association between cigarette smoking and depressive symptoms among military medical students in Turkey. *Psychiatry Research 134*. 281-286.

Hakamata Y., Iwase M., Iwata H., Kobayashi T., Tamaki T., Nishio M., Matsuda H., ... Inada T. (2009). Gender difference in relationship between anxiety-related personality traits and cerebral brain glucose metabolism. *Psychiatry Research: Neuroimaging 173*. 206-211.

Hamilton S.P. (2011). A new lead from genetic studies in depressed siblings: Assessing studies of chromosome 3. *Am J Psychiatry 168*. 783-789.

Hankin B.L., & Abramson L. (2001). Development of gender differences in depression: An elaborated cognitive vulnerability-transactional stress theory. *Psychological Bulletin 127*. 773-796.

Hankin B.L., Fraley R.C., Lahey B.B., & Waldman ID. (2005). Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population based sample. *Journal of Abnormal Psychology 114*. 96-110.

Hayama M., Okumura Y., Takahashi E., Shimabukuro A., Tamura M., Takeda N., ... Kido H. (2007). Identification and analysis of the promoter region of the type II transmembrane serine protease polyserase-1 and its transcript variants. *Biol Chem 388*. 853-858.

Hek K., Mulder C.L., Luijendijk H.J., van Duijn C.M., Hofman A., Uitterlinden A.G., & Tiemeier H. (2010). The PCLO gene and depressive disorders: replication in a population-based study. *Human Molecular Genetics 19.* 731-734.

Herbert A., Garry N.P., McQueen M.B., et al. (2006). A common genetic variant is associated with adult and childhood obesity. *Science 312*. 279-283.

Hettema J.M., Prescott C.A., & Kendler K.S. (2004). Genetic and environmental sources of covariation between generalized anxiety disorder and neuroticism. *Am J Psychiatry 161*. 1581-1587.

Hettema J.M., Neale M.C., Myers J.M., Prescott C.A., & Kendler K.S. (2006). A population-based twin study of the relationship between neuroticism and internalizing disorders. *Am J Psychiatry 163.* 857-864.

Hirschfield R.M., Klerman G.L., Clayton P.J., Keller M.B., *et al.* (1983). Assessing personality: effects of the depressive state on trait measurement. *Am J Psychiatry 140*. 695-699.

Hirschhorn M., & Daly M.J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics 6*. 95-108.

Hogeweg P. (2011) Searls D.B. (Ed.) The roots of bioinformatics in theoretical biology. *PLoS Computational Biology 7 (3).* e1002021.

Holliday R. (1990). Mechanisms for the control of gene activity during development. *Biol. Rev. Cambr. Philos. Soc. 65*. 431-471.

Holmans P., Zubenko G.S., Crowe R.R., DePaulo Jr. J.R., Scheftner W.A., Weissman M.M., ... Levinson D.F. (2004). Genomewide significant linkage to recurrent, early-onset major depressive disorder on chromosome 15q. *Am J Hum Genet 74*. 1154-1167.

Horstmann S., Lucae S., Meke A., Hennings J.H., Ising M., Roeske D., ... Binder EB. (2010). Polymorphisms in GRIK4, HTR2A, and FKBP5 show interactive effects in predicting remission to antidepressant treatment. *Neuropsychopharmacology 35*. 727-740.

Huezo-Diaz P., Tandon K., & Aitchison K.J. (2005). The genetics of depression and related traits. *Curr Psychiatry Rep 7*. 117-124.

Husky M.M., Mazure C.M., Paliwal P., & McKee S.A. (2008). Gender differences in the comorbidity of smoking behavior and major depression. *Drug Alcohol Depend 93.* 176-179.

Hyde J.S., Mezulis A.H., & Abramson L.Y. (2008). The ABCs of depression: Integrating affective, biological, and cognitive models to explain the emergence of the gender difference in depression. *Psychological Review 15.* 291-313.

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature 431*. 931-945.

Jabbi M., Korf J., Ormel J., Kema I.P., & den Boer J.A. (2008). Investigating the molecular basis of Major Depressive Disorder etiology: A functional convergent genetic approach. *Ann N Y Acad Sci 1148*. 42-56.

Jang K.L., Livesley W.J., & Vernon P.A. (1996). Heritability of the big five personality dimensions and their facets: A twin study. *J Pers 64*. 577-591.

Jardine R., Martin N.G., & Henderson A.S. (1984). Genetic covariation between neuroticism and the symptoms of anxiety and depression. *Genet Epidemiol 1*. 89-107.

Jorm A.F. (1987). Sex differences in neuroticism: A quantitative synthesis of published research. *Aust NZ J Psychiatry 21*. 501-506.

Joslyn G., Ravindranathan A., Brush G., Schuckt M., & White R.L. (2010). Human variation in alcohol response is influenced by variation in neuronal signaling genes. *Alcohol Clin Exp Res 34*. 800-812.

Jostins L. (2009). Basics: Sequencing DNA, part 2. Genetic Interference blog. From, http://www.genetic-inference.co.uk/blog/2009/08/basics-sequencing-dna-part-2/

Kao C.F., Fang Y.S., Zhao Z., & Kuo P.H. (2011). Prioritization and evaluation of depression candidate genes by combining multidimensional data resources. *PLoS ONE 6(4).* 1-9.

Kathiresan et al. (2008). Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nature Genet. 40*. 189-197.

Kendler K.S., Neale M.C., Kessler R.C., Heath A.C., & Eaves L.J. (1992). Major depression and generalized anxiety disorder: Same genes (partly) different environments? *Arch Gen Psychiatry 49*. 716-722.

Kendler K.S., Neale M.C., Kessler R.C., Heath A.C., & Eaves L.J. (1993). http://www.ncbi.nlm.nih.gov/pubmed/15939839 A longitudinal twin study of personality and major depression in women. *Arch Gen Psychiatry 50*. 853-862.

Kendler KS, & Prescott C.A. (1999). http://www.ncbi.nlm.nih.gov/pubmed/9892254 A population-based twin study of lifetime major depression in men and women. *Arch Gen Psychiatry 56.* 39-44.

Kendler K.S., Gardner C., & Prescott C. (2002). Toward a comprehensive developmental model for major depression in women. *American Journal of Psychiatry 159*. 1133-1145.

Kesebir S., Vahip S., Akdeniz F., Yuncu Z., Alkan M., & Akiskal H. (2005). Affective temperaments as measured by TEMPS-A in patients with bipolar I disorder and their first-degree relatives: A controlled study. *Journal of Affective Disorders 85*. 127-133.

Kessler R.C., Wittchen H.-U., Abelson J.M., McGonagle K., Schwartz N., Kendler K.S., ... Zhao S. (1998). Methodological studies of the composite international diagnostic interview (CIDI) in the US national comorbidity survey. *Intl J of Methods in Psychiatric Research 7*. 33-55.

Kessler R.C., & Ustun T.B. (2000). The World Mental Health (WMH) survey initiative version of the World Health Organization (WHO) Composite international diagnostic interview (CIDI). *International Journal of Methods in Psychiatric Research 13(2).* 93-121.

Kessler R.C., Berglund P., Demler O., Jin R., Koretz D., Merikangas K.R., ... Wang P.S. (2003). The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA 289(23).* 3095-3105.

Kessler R.C., Chiu W.T., Demler O., Merikangas K.R., & Walters E.E. (2005). http://www.ncbi.nlm.nih.gov/pubmed/15939839 Prevalence, severity, and comorbidity

of 12-month DSM-IV disorders in the national comorbidity survey replication. *Arch Gen Psychiatry 62(6).* 617-627.

Klein D.N., Lewinsohn P.M., Seeley J.R.,& Rohde P. (2001). A family study of major depressive disorder in a community sample of adolescents. *Arch Gen Psychiatry 58.* 13-20.

Kovacs M., & Lopez-Duran N. (2010). Prodromal symptoms and atypical affectivity as predictors of major depression in juveniles: Implications for prevention. *J Child Psychol Psychiatry 51.* 472-496.

Krishnan V., & Nestler E.J. (2008). The molecular neurobiology of depression. *Nature 455.* 894-902.

Lake R..I, Eaves L.J., Maes H.H., Heath A.C., & Martin N.G. (2000). Further evidence against the environmental transmission of individual differences in neuroticism from a collaborative study of 45850 twins and relatives on two continents. *Behav Genet 30.* 223-233.

Lange C., DeMeo D., Silverman E.K., Weiss S.T., & Laird N.M. (2004). PBAT: Tools for family-based association studies. *Am J Hum Genet 74.* 367-369.

Lavretsky H., Kurbanyan K., Ballmaier M., Mintz J., Toga A., & Kumar A. (2004). Sex differences in brain structure in geriatric depression. *Am J Geriatr Psychiatry 12.* 653-657.

Lazary J., Gonda X., Benko A., Gacser M., & Bagdy G. (2009). Association of depressive phenotype with affective family history is mediated by affective temperaments. *Psychiatry Research 168.* 145-152.

Lesch K.P., Bengel D., Heils A., Sabol S., Greenberg B., Petri S., ... Murphy D. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* 274. 1527-1530.

Lesch K.P. (2004). Gene-environment interaction and the genetics of depression. *J Psychiatry Neurosci 29.* 174-184.

Levinson D.F. (2006). The genetics of depression: A review. *Biol Psychiatry* 60(2). 84-92.

Lewinsohn P.M., Solomon A., Seeley J.R., & Zeiss A. (2000). Clinical implications of "subthreshold" depressive symptoms. *Journal of Abnormal Psychology 109*. 345-351.

Leyfer O.T., Woodruff-Borden J., Klein-Tasman B.P., Fricke J.S., & Mervis C.B. (2006). Prevalence of psychiatric disorders in 4 to 16-year-olds with Williams syndrome. *Am J Med Genet B Neuropsychiatr Genet. 141B.* 615-622.

Liu G., Clement L.C., Kanwar Y.S., Avila-Casado C., & Chugh S.S. (2006). ZHX proteins regulate podocyte gene expression during the development of nephrotic syndrome. *The Journal of Biological Chemistry 281*. 39681-39692.

Liu X., Xu Y., Jiang S., Cui D., Qian Y., & Jiang K. (2009). Family-based association study between brain-derived neurotrophic factor gene and major depressive disorder of Chinese descent. *Psychiatry Res 169.* 169-172.

Loftus E.F., & Palmer J.C. (1974). Reconstructions of automobile destructions: an example of the integration between language and memory. *Journal of Verbal Language and Verbal Behavior 13*. 585-589.

Luby J.L. (2009). Early childhood depression. *Am J Psychiatry 166.* 974-979.

Manchia M., Squassina A., Congiu D., Chillotti C., Ardau R., Severino G., & Del Zompo, M., (2009). Interacting genes in lithium prophylaxis: Preliminary results of an exploratory

analysis on the role of DGKH and NR1D1 gene polymorphisms in 199 Sardinian bipolar patients. *Neuroscience Letters 467*. 67-71.

Marcus S.M., Young E.A., Kerber K.B., Kornstein S., Farabaugh A.H., Mitchell J., ... Rush, A.J. (2005). Gender differences in depression: Findings from the STAR*D study. *J Affect Disord 87(2-3).* 141-150.

Martin N., Goodwin G., Fairburn C., Wilson R., Allison D., et al. (2000). A population-based study of personality in 34,000 sib-pairs. *Twin Res 3*. 310-315.

Mathers C.D., & Loncar D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS medicine 3*. e442.

McCarthy M.I., Abecasis G.R., Cardon L.R., Goldstein D.B., Little J., Ioannidis J.P., & Hirschhorn J.N. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat. Rev. Genet 9.* 356-369.

McGeachie M., Ramoni R.L., Mychaleckyj J.C., Furie K.L., Dreyfuss J.M., Liu Y., ... Ramoni M.F. (2009). Integrative predictive model of coronary artery calcification in atherosclerosis. *Circulation 120.* 2448-2454.

McGuffin P., Katz R., Watkins S., & Rutherford J. (1996). A hospital based twin register of the heritability of DSM IV unipolar depression. *Arch Gen Psychiatry 53.* 129-136.

McGuffin P., Knight J., Breen G., Brewster S., Boyd P.R., Craddock N., ... Farmer A.E. (2005). Whole genome linkage scan of recurrent depressive disorder from the depression network study. *Human Molecular Genetics 14.* 3337-3345.

McMahon F.J., Akula N., Schulze T.G., Muglia P., Tozzi F., Detera-Wadleigh S.D., ..., Rietschel M. (2010). Meta-analysis of genome-wide association data identifies a risk locus for major mood disorders on 3p21.1. *Nat Genet 42.* 128-131.

Merikangas K.R., Zhang H., Avenevoli S., Acharyya S., Neuenschwander M., & Angst J. (2003). Longitudinal trajectories of depression and anxiety in a prospective community study: The Zurich Cohort Study. *Arch Gen Psychiatry 60*. 993-1000.

Miguel-Hidalgo J.J., Waltzer R., Whittom A.A., Austin M.C., Rajkowska G., & Stockmeier C.A. (2010). Glial and glutamatergic markers in depression, alcoholism, and their comorbidity. *J Affect Disord 127(1-3).* 230-240.

Millon T. (2004). *Masters of the mind: Exploring the story of mental illness from ancient times to the new millennium*. *John Wiley & Sons*. ISBN 9780471469858.

Morton, N.E. (1982). *Outline of genetic epidemiology. Karger*. ISBN 380552269X.

Mun-Keat Looi. (2009). Feature: Genomics – the next generation. From, http://www.wellcome.ac.uk/News/2009/Features/WTX056032.htm.

Myung W., Lim S.W., Kim J., Lee Y., Song J., Chang K.W., & Kim D.K. (2010). Serotonin transporter gene polymorphisms and chronic illness of depression. *J Korean Med Sci 25*. 1824-1827.

Neale B.M., & Purcell S. (2008). The positives, protocols, and perils of genome-wide association. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 147B.* 1288-1294.

Neale B.M., Ferreira M.A.R., Medland S.E., Posthuma D. Eds. (2009). *Statistical Genetics: Gene mapping through linkage and association. Taylor & Francis Group*. London. ISBN 9780415410403.

Nelson E.C., Joyce P.R., Sullivan P.F., Bulik C.M. et al. (2004). Genetic epidemiology of alcohol-induced blackouts. *Archives of General Psychiatry 61*. 257-263.

Nolen-Hoeksema S., & Girgus J. (1994). The emergence of gender differences in depression during adolescence. *Psychological Bulletin 115*. 424-443.

Nurnberger Jr. J.I., Foroud T., Flury L., Meyer E.T., Hu K., Crowe R., ... Reich W. (2001). Evidence for a locus on Chromosome 1 that influences vulnerability to alcoholism and affective disorder. *Am J Psychiatry 158*. 718-724.

Oksenberg L., Cannell C.F., & Kanton G. (1991). New strategies for pretesting survey questions. *Journal of Official Statistics 7*. 349-365.

Ollila H.M., Soronen P., Silander K., Palo O.M., Kieseppa T., Kaunisto M.A., ... Paunio T. (2009). Findings from bipolar disorder genome-wide association studies replicate in a Finnish bipolar family-cohort. Mol. Psychiatry *14.* 351-353.

Parker G., & Hadzi-Pavlovic D. (2004). Is the female preponderance in major depression secondary to a gender difference in specific anxiety disorders? *Psychological Medicine 34.* 461-470.

Peng ZW, Chen XG, & Wei Z, 2007. Cryptochrome1 may be a candidate gene of schizophrenia. *Medical Hypotheses 69.* 849-851.

Penninx B.W.J.H., Beekman A.T.F., Smit J.H., Zitman F.G., Nolen W.A., Spinhoven P. ..., van Dyck R. (2008). The Netherlands study of depression and anxiety (NESDA): Rationale, objectives and methods. *Int J Methods Psychiatr Res 17.* 121-140.

Persson M.L., Wasserman D.G., Jonsson E., Bergman H., Terenius L., Gyllander A., ... Geijer T. (2000). Search for the influence of the tyrosine hydroxylase (TCAT)n repeat polymorphism on personality traits. *Psychiatry Research 95.* 1-8.

Pervin L.A., & John O.P. (1990). *Handbook of personality: Theory and research.* New York: Guilford Press.

Plink. (n.d.). Whole genome data analysis toolset. From, http://pngu.mgh.harvard.edu/~purcell/plink/.

Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A., & Reich D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet 38.* 904-909.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., ... Sham P.C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet 81*. 559-75.

Rao V., Bertrand M., Rosenberg P., Makley M., Schretlen D.J., Brandt J., & Mielke M. (2010). Predictors of new-onset depression after mild traumatic brain injury. *J Neuropsychiatry Clin Neurosci 22(1).* 100-104.

Regier D.A. (2009). The conceptual development of DSM-V. *Am J Psychiatry 166.* 645.

Rietschel M., Beckmann L., Strohmaier J., Georgi A., Karpushova A., Schirmbeck F., ... Schulze T.G. (2008). G72 and its association with major depression and neuroticism in large population-based groups from Germany. *Am J Psychiatry 165.* 753-762.

Risch N., & Merikangas K. (1996). The future of genetic studies of complex human diseases. *Science. 273.* 1516-1517.

Rose J.E., Behm F.M., Drgon T., Johnson C., & Uhl G.R. (2010). Personalized smoking cessation: Interactions between nicotine dose, dependence and quit-success genotype score. *Mol Med 16.* 247-253.

Sanna et al. (2008). Common variants in the GDF5-UQCC region are associated with variation in human height. *Nature Genet. 40.* 198-203.

Schmidt L.G., Sander T., Kuhn S., Smolka M., Rommelspacher H., Samochowiec J, & Lesch K.P. (2000). Different allele distribution of a regulatory MAOA gene promoter polymorphism in antisocial and anxious-depressive alcoholics. *J Neural Transm 9.* 103-112.

Serretti A., Cusin C., Cristina S., Lorenzi C., Lilli R., Lattuada E., ... Nappi G. (2003). Multicentre Italian family-based association study on tyrosine hydroxylase, catechol-O-methyl transferase and Wolfram syndrome 1 polymorphisms in mood disorders. *Psychiatr Genet 13.* 121-126.

Shi J., Potash J.B., Knowles J.A., Weissman M.M., Coryell W., Scheftner W.A., ... Levinson D.F. (2010). Genome-wide association study of recurrent early-onset major depressive disorder. *Mol Psychiatry 16.* 193-201.

Shifman S., Bhomra A., Smiley S., Wray N.R. et al. (2008). A whole genome association study of neuroticism using DNA pooling. *Mol Psychiatry 13.* 302-312.

Shorey R.C., Brasfield H., Febres J., & Stuart G.L. (2011). Gender differences in depression and anxiety among victims of intimate partner violence: Moderating effect of shame proneness. *J Interpers Violence 26.* 2681-2697.

Shyn S.I., Shi J., Kraft J.B., Potash J.B., Knowles J.A., Weissman M.M., ... Hamilton S.P. (2009). Novel loci for major depression identified by genome-wide association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies. *Mol Psychiatry 16.* 202-215.

Shyn S.I., Hamilton S.P. (2010). The genetics of major depression: Moving beyond the monoamine hypothesis. *Psychiatr Clin North Am 33.* 125-140.

Silverstein B., (1999). Gender difference in the prevalence of clinical depression: The role played by depression associated with somatic symptoms. *Am J Psychiatry 156.* 480-482.

Singh A.L., D'Onofrio B.M., Slutske W.S., Turkheimer E., Emery R.E., Harden K.P., ... Martin N.G. (2011). Parental depression and offspring psychopathology: A children of twins study. *Psychol Med 41.* 1385-1395.

Skol A.D., Scott L.J., Abecasis G.R., & Boehnke M. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet 38.* 209-213.

South S.C., & Krueger R.F. (2008). Marital quality moderates genetic and environmental influences on the internalizing spectrum. *J Abnorm Psychol 117.* 826-837.

Stein J.L., Hua X., Morra J.H., Lee S., Hibar D.P., Ho A.J., ... Alzheimer's Disease Neuroimaging Initiative. (2010). Genome-Wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegradation in Alzheimer's disease. *Neuroimage 51.* 542-554.

Sullivan P.F., Neale M.C., & Kendler K.S. (2000). Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry 157.* 1552-1562.

Sullivan P.F., Eaves L.J., Kendler K.S., & Neale M.C. (2001). Genetic case-control association studies in neuropsychiatry. *Arch Gen Psychiat 58.* 1015-1024.

Sullivan P.F., de Geus E.J.C., Willemsen G., James M.R., Smit J.H., Zandbelt T., ... Penninx B.W. (2009). Genomewide association for major depressive disorder: A possible role for the presynaptic protein Piccolo. *Mol Psychiatry 14.* 359-375.

Sund A.M., Larsson B., & Wichstrom L. (2011). Prevalence and characteristics of depressive disorders in early adolescents in central Norway. *Child Adolesc Psychiatry Ment Health.* 2011 Aug 31; *5(1).* 28. [Epub ahead of print]

Szewczyk B., Albert P.R., Burns A.M., Czesak M., Overholser J.C., Jurjus G.J., ... Austin M.C. (2009). Gender-specific decrease in NUDR and 5-HT1A receptor proteins in the prefrontal cortex of subjects with major depressive disorder. *Int J Neuropsychopharmacol 12.* 155-168.

Terraciano A., Sanna S., Uda M., Deiana B., et al. (2010). Genome-wide association scan for five major dimensions of personality. *Mol Psychiatry 15.* 647-656.

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

Thompson C.L., Klein B.E.K., Klein R., Xu Z., Capriotti J., Joshi T., ... Iyengar S.K. (2007). Complement factor H and hemicentin-1 in age-related macular degeneration and renal phenotypes. *Human Molecular Genetics 16.* 2135-2148.

Tochigi M., Otowa T., Hibino H., Kato C., Otani T., Umekage T., ... Sasaki T. (2006). Combined analysis of association between personality traits and three functional polymorphisms in the tyrosine hydroxylase, monoamine oxidase A and catechol-O-methyltransferase genes. *Neuroscience Research 54.* 180-185.

Tonna S., Dandapani S.V., Uscinski A., Appel G.B., Schlondorff J.S., Zhang K., ... Pollack M.R. (2008). Functional genetic variation in aminopeptidase A (ENPEP): Lack of clear association with focal and segmental glomerulosclerosis (FSGS). *Gene 410(1).* 44-52.

Traks T., Koido K., Eller T., Maron E., Kingo K., Vasar V., Koks S. (2008). Polymorphisms in the interleukin-10 gene cluster are possibly involved in the increased risk for major depressive disorder. *BMC Medical Genetics 9.* 111.

Tsuang M.T., Bar J.L., Stone W.S., & Faraone S.V. (2004). Gene-environment interactions in mental disorders. *World Psychiatry 3(2).* 73-83.

Uddin M., Koenen K.C., Aiello A.E., Wildman D.E., de los Santos R., & Galea S. (2010). Epigenetic and inflammatory marker profiles associated with depression in a community-based epidemiologic sample. *Psychological Medicine 41.* 997-1007.

U.S.Department of health and Human Services. (1999). Mental health: A report of the Surgeon General – executive summary. Rockville, MD: U.S.Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health. Superintendent of Documents, Pittsburgh, PA.

U.S.Public Health Service. (1999). The surgeon general's call to action to prevent suicide. Washington, DC: Author.

van den Oord E.J., Kuo P.H., Hartmann A.M., Webb B.T., Moller H. J., Hettema J.M., ... , Rujescu D. (2009). Genomewide association analysis followed by a replication study implicates a novel candidate gene for neuroticism. *Arch Gen Psychiatry 65.* 1062-1071.

Vink J.M., Smit A.B., de Geus E.J., Sullivan P., Willemsen G., Hottenga J.J., ... Boomsma D.I. (2009). Genome-wide association study of smoking initiation and current smoking. *Am J Hum Genet 84.* 367-379.

Wade T.D., Bulik C.M., Neale M., & Kendler K.S. (2000). Anorexia nervosa and major depression: Shared genetic and environmental risk factors. *Am J Psychiatry 157.* 469-471.

Wang K.S., Liu X.F., Zhang Q.Y., Aragam N.R., Pan Y. (2011). Genome-Wide association analysis of age at onset in schizophrenia in an European-American sample. *Am J Med Genet B Neuropsychiatr Genet 156B(6).* 671-680.

Wang K.S., Liu X.F., Aragam N.R., Jian X., Mullersman J.E., Liu Y., & Pan Y. (2011). Family-based association analysis of alcohol dependence in the COGA sample and replication in the Australian twin-family study. *J Neural Transm*. *118 (9).*1293-1299. Epub 2011 Mar 29.

Wang K.S., Liu X..F, & Aragam N.R. (2010) A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophr Res. 142 (1-3).* 192-199.

Wang H.X., Karp A., Herlitz A., Crowe M., Kareholt .I, et al. (2009). Personality and lifestyle in relation to dementia incidence. *Neurology 72*. 253-259.

Weedon et al. (2007). A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nature Genet. 39.* 1245-1250.

Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature 447.* 661-678.

Whittle S., Allen N.B., Lubman D.I., & Yucel M. (2006). The neurobiological basis of temperament: towards a better understanding of psychopathology. *Neuroscience and Biobehavioral Reviews 30*. 511-525.

Wilson R.S., Evans D.A., Bienias J.L., de Leon C.F.M., Schneider J.A., et al. (2003). Proneness to psychological distress is associated with risk of Alzheimer's disease. *Neurology 61*. 1479-1485.

Wyatt K., White H.E., Wang L., Bateman O.A., Slingsby C., Orlova E.V., & Wistow G. (2006). Lengsin is a survivor of an ancient family of class I glutamine synthetases in eukaryotes that has undergone evolutionary reengineering for a role in vertebrate eye lens. *Structure 14(12).* 1823-1834.

Zubenko G.S., Hughes H.B., Maher B., Stiffler J., Zubenko W., & Marazita M. (2002). Genetic linkage of region containing the CREB1 gene to depressive disorders in women from families with recurrent early-onset, major depression. *American Journal of Medical Genetics 114*. 980-987

Zubenko G.S., Maher B., Hughes III H.B., Zubenko W.N., Stiffler J.S., Kaplan B.B., & Marazita M.L. (2003). Genome-wide linkage that influence the development of depressive disorders in families with recurrent, early-onset major depression. *American Journal of Medical Genetics Part B 123B.* 1-18.

APPENDICES

APPENDIX A

A1. Lifetime prevalence of Major Depressive Disorder and Age at Onset distributions

The literature indicates that until 2005 little was known about the lifetime prevalence or age at onset of a number of DSM-IV disorders including MDD. Kessler et al. (2005) conducted a study of 9282 English-speaking, nationally representative Americans aged 18 and older (from the NCS-R sample) in order to estimate the distribution of lifetime prevalence and age at onset of DSM-IV disorders. Their findings reported that about half of Americans will meet a DSM-IV disorder criterion sometime in their life, with the first onset usually in childhood or adolescence. Thus any intervention to prevent or treating early the symptoms of MDD must be focused on the youth according to Kessler et al. (2005).

The results from Kessler et al. (2005) epidemiologic study indicate that the lifetime prevalence estimates are as follows –

Table A1-1

Lifetime Prevalence Estimates

Age Group	Prevalence % (SE)
Total sample (18 – 60+)	16.6 (0.5)
Age 18 - 29	15.4 (0.7)
Age 30 - 44	19.8 (0.9)
Age 45 - 59	18.8 (1.1)
Age over 60	10.6 (0.8)

Ages at selected percentiles on the standardized age at onset distributions of MDD with projected lifetime risk at age 75 are as follows –

Table A1-2

Projected Lifetime Risk at Age 75

Projected Lifetime Risk at age 75; 23.2% (0.6 SE)	**Actual Age**
Percentile 5	12
Percentile 10	14
Percentile 25	19
Percentile 50	32
Percentile 75	44
Percentile 90	56
Percentile 95	64
Percentile 99	73

The authors of the above study, however, point to four biases that should be noted -

1. People with a history of mental illness are less likely to participate in the study.
2. Well-known bias of underreporting embarrassing behaviors.
3. Estimation of lifetime risk was based on the assumption of constant conditional risk of first onset in a given year of life among people of different ages at interview.
4. Recall bias of age at onset.

Given the enormous public health consequences of MDD, the above observations should lead us to thinking about interventions early in childhood and adolescent years and prevention of MDD and its comorbidities (Kessler et al., 2005).

A2. An Overview of WMH-CIDI

The World mental Health (WMH) Composite International Diagnostic Interview (CIDI) is a collection of a screening module and has 40 sections with 22 sections that focus on diagnoses, four sections on functioning, 2 sections on treatment, 4 sections on risk factors, 7 sections on socio-demographic correlates, and 2 sections on methodological factors. A computer-based version of the interview is available with direct data entry software using a keyboard. Diagnoses based on ICD-10 and DSM-IV criteria are generated from the WMH-CIDI software. In addition, CDROM-based training materials are available to teach the interviewers how to administer the interview and teach the supervisors how to monitor the quality of data collected. A brief description of the details (Kessler & Ustun, 2000) is presented below.

The diagnostic survey methodological issues consists of question and task comprehensions, motivation, ability to answer diagnostic questions accurately, section to review one's lifetime, details of Part I and Part II diagnostics, and the expanded section to include such lifetime factors as the socio-economic status (SES), family, and childhood experiences. The aim of the interview is to obtain *valid* information about the prevalence, correlates, and unmet needs for treatment of mental disorders in the general population as well as to determine the treatment adequacy and the societal burden of mental disorders.

Question comprehension

Answering structured interview questions that are ambiguous is easily misconstrued by more than 70% of the respondents as was shown by Belson (1981) and others (Oksenberg et al., 1991). The problem in surveys is that the flow of questions is predetermined by the researcher and the normal rules of conversation (e.g., give-and-take) do not apply and this is compounded when the

topic of discussion is the emotional experiences of the respondent. Four discriminating features among the high and low levels of misunderstandings are complexity, vagueness, multiple interpretations of an odd experience, and contextual. The solution arrived to avoid the problems in question understanding was to breakdown the complex questions into simpler subquestions.

Task comprehension

Second type of misunderstood response in CIDI came from people who did not understand what they were being tasked to do when a response needed a considerable thought. Cannell et al. (1981) has shown that this problem could be solved by instructing respondents to answer the question completely and accurately. The stem question preceding questions that need clear comprehension of the task at hand must be timed properly and also placed appropriately within the survey along with a preface – "The next question may be difficult, so please take your time before answering".

Motivation

Getting valid responses needs motivation on the part of the respondent and thus motivational instructions and contingent reinforcement during the interview (e.g., "You answered that very quickly. Was there anything else, even something unimportant that you want to add?") are needed to enhance the validity of the responses. CIDI uses this type of contingent reinforcement a part of their training materials developed for the WMH surveys. Another motivational aspect of getting the respondent demur explicitly or tacitly work hard to give an honest answer is a part of the common understanding between the interviewer and the respondent (e.g., Clark et al., 1992; Loftus & Palmer, 1974).

Ability to answer accurately

Episodic and semantic memories play a role in influencing the respondent's accuracy in answering questions. For example, the respondent may have a semantic memory of what panic attacks are like, but after having many such attacks in one's lifetime, one may not be able to recall one attack accurately. Research has shown that people are more likely to recall episodic memories more accurately than the semantic memories (say, from panic attacks) that are frequent, typical, and regular. Thus, asking questions without knowing the limits of one's memory should be based on episodic memory for accuracy, whereas, clarification must precede any questions about semantic memories. Thus, the designers of WMH-CIDI used pilots where respondents were debriefed with an explicit eye towards pinpointing questions that were difficult to answer.

Lifetime review section

One of the most important aspects of the CIDI is the life review cycle that is administered at the beginning of the interview. This will both motivate and facilitate active memory search in answering the diagnostic stem questions. This modification to the earlier CIDI led to a significant increase in the proportion of respondents who endorsed the diagnostic stem questions (Kessler et al., 1998). While some upward biasing questions were raised from some early international surveys (e.g., as much as 50% increase in lifetime criteria for ICD or DSM mental disorders), clinical calibration proved otherwise leading to add a clinical significance criterion to many disorders in DSM-IV. However, some critics have argued for including some subthreshold cases (e.g., Merikangas et al., 2003) in the definitions for mental disorders. Research-wise this

might be important because full exploration of the continua rather than the currently established threshold guidelines yields greater power in studies of genetic and environmental risk factors (Benjamin et al., 1998). Further discussion on subthreshold disorders, symptom persistence and severity, internal and external impairments, and the ranking of physical and mental disorders are available in Kessler & Ustun, 2000.

Part I and Part II diagnoses

The WMH-CIDI is an elaborate instrument that takes up as many as 2 full hours to complete the interview. In practice this would entail complications in order that the interview is done in two parts – Part I and Part II. Part I includes all core diagnostic assessments and those who report having no lifetime history of disorder are terminated at this midpoint of the interview. Part II is completed by those diagnosed as being with lifetime disorder issues. Further elaboration of subsampling of noncases to be included in Part II and other case-control and statistical power issues are more fully discussed in Kessler and Ustun, 2000.

Other factors

The main reason behind developing WMH-CIDI was to expand a previous version of the instrument and include assessments of risk factors, consequences, and treatment. Accordingly, socio-demographics, treatment and pharmacoepidemiology, nonspecific psychological distress, family burden, and childhood experiences are all included in specific subsections of the interview. Socio-demographics include such predictors (or, risk factors) as age, gender, race, educational attainment, marital status, and employment status. The idea here is to include the dynamic information about achieved statuses in the interview schedule. Treatment and

pharmacoepidemiology sections assess both internally and externally the patterns and predictors of delays in initial treatment after first onset of disorder, lifetime hospitalization, and the type of professionals the respondents have sought and seen. Nonspecific psychological distress information gives the frequency and a scaled score from a screen for serious mental illness (SMI) 30 days before the interview and in the worst month of the past year. This information would be very useful as an inexpensive mental health needs surveillance tool for public health professionals. While the assessment of the respondent gives rise to excellent understanding of how mental disorders affect the people with the disorders, it ignores the patient's family. Thus, the enormous family burden needs to be measured and a special section has been introduced to assess the same in WMH-CIDI. After inquiring about a network of first degree relatives, the respondent is asked about 12 serious health problems (cancer, heart problems, memory problem, mental disorders like schizophrenia, manic depression, etc.) of those first order relatives. The data collected here also helps in laying the ground work for any family-based genetic studies at a later date. Finally, such childhood experiences as having a single parent or nonbiological parents, trauma caused by authoritarian, overprotective or neglectful parents, any sexual abuse, etc. are collected by a special section at the end of the CIDI interview. Separate assessments of parental depression, panic disorder, general anxiety disorder, antisocial behaviors, and substance-abuse disorders are made here.

The results from the WMH-CIDI interview are summarized in terms of quantitative scores that are stored in a database and a final qualitative diagnosis of MDD is reached by a computer program.

APPENDIX B

Brief Summary of the Genomics and Genome-Wide Association Analyses Literature

Considering that the human genome was fully sequenced (i.e., over 90% of the functional genome) in 2004 (International Human Genome Sequencing Consortium, 2004), and the first GWAS of seven common diseases was published in 2007 (Wellcome Trust Case Control Consortium, 2007), the literature of GWAS of MDD is rather limited. However, the pathophysiology of MDD at genetic and molecular levels has been studied extensively in the past 3 to 5 years and many genes (and/or gene regions) have been implicated to be strongly associated with MDD and its related phenotypes.

In addition to the complete sequencing of the functional human genome, the International HapMap project, which was a collaborative project started in 2002 and completed in 2009, has given us the ability to search the entire human genome for genetic variants. The HapMap project was a collaborative effort between researchers in Canada, China, Japan, Nigeria, UK, and the US and used samples from four groups as follows:

- 30 adults-and-both-parents trios (90 people) from Ibadan, Nigeria (YRI)
- 30 trios (90 people) US residents of northern and western European ancestry (CEU)
- 44 unrelated individuals from Tokyo, Japan (JPT), and
- 45 unrelated Han Chinese individuals from Beijing, China (CHB)

The genotyping of these diverse human populations has resulted in a SNP database (dbSNP) of the complete human genome that has provided over 27 million SNPs for use in genetic research and Genome-Wide studies.

A brief summary of definitions, post-genomic era accomplishments, consensus, challenges and uncertainties, positives and perils, and phenomics – the systematic study of phenotypes on a genome-wide scale are discussed briefly below.

Definitions

In recent years (since 1980s), many new areas of genetics research have emerged, and thanks to the completion of sequencing the known functional human genome in 2004 there has been an even newer emergence of other 'omics era. Some of the more important areas of research and the terminology attributable to this new research are –

- Public Health Genomics – Centers for Disease Control and Prevention (CDC) defines Public Health genomics as an emerging field of study that assesses the impact of genes and their interaction with behavior, diet and the environment on the population's health (Centers for Disease Control and Prevention, n.d.).
- Genetic and Molecular Epidemiology – Morton defines this as "a science which deals with the etiology, distribution, and control of disease in groups of relatives and with inherited causes of disease in populations" (Morton, 1982). It is closely allied to both molecular epidemiology and statistical genetics, but these overlapping fields each have distinct emphases, societies, and journals.
- Gene x Gene and Gene x Environment interactions – When two or more genes interact to produce a single phenotype, gene x gene interaction is said to have taken place. Genes may also interact with an environmental factor such as gender and alcohol or tobacco usage to produce phenotypes that might be different than that

would have resulted from the influence of the gene alone or with other genes (gene x gene interaction). In mental disorders many normal physiological conditions such as blood pressure and cognitive abilities such as intelligence probably result from the combined action of multiple genes, each producing a small effect together with a variety of environmental factors (Tsuang et al., 2004).

- Genotype-phenotype databases – The genetic information is described in terms of the sequence of nucleotides (A, C, T, G) comprising the DNA and the most common form of genetic variation is in terms of a single nucleotide polymorphism (SNP). In concert with the rapid availability of several genotyping technologies, the systematic interrogation of the whole genome has become possible (Chanock et al., 2007). A compilation of these varying genetic sequences of a specimen from an organism (including the human) in an orderly way into a database represents a genotype database. The establishment of an association between a genotype and a phenotype constitutes a map, which when compiled into a database of associations would constitute a phenotype database. Some of the well-known examples of genotype and phenotype databases are – dbGaP (Database of Genotypes and Phenotypes), a publicly available database for well-known diseases from NIH, GWAS Central, containing the most comprehensive collection of summary level p-value GWAS data. Also, dbSNP, the Single Nucleotide Polymorphism Database is a free public archive for genetic variation within and across different species developed and hosted by the

National Center for Biotechnology Information (NCBI) in collaboration with the National Human Genome Research Institute (NHGRI).

- Proteomics – A new branch of genomic sciences involving the large-scale study of proteins, particularly their structures and functions (Anderson & Anderson, 1998).
- Systems Biology and Metabolic Modeling - Kell (2006) opines that – "systems biology involves an iterative interplay between more or less high-throughput and high-content 'wet' experiments, technology development, theory and computational modeling, and that it is the involvement of computational modeling, in particular, in the process that sets systems biology apart from the more traditional and more reductionist molecular biology". Simply put, metabolic modeling consists of studying metabolism at a molecular level using sophisticated computer-based algorithms to determine the changes in the concentrations of enzymes (and the transcripts that encode them) that have substantial effects on the concentrations of metabolic intermediates.
- Bioinformatics and Biobanks - In simple terms, Bioinformatics is the application of computer science and statistics to the field of molecular biology. Since the 1980's the field of bioinformatics has grown to include the areas of genetics, genomics, and DNA sequencing, while including creation and advancement of databases, algorithms, computational and statistical techniques, and theory to solve problems arising from the management and analyses of biological data (Hogeweg, 2011).

Biobanks are the natural evolution of bioinformatic databases to store and document access to biological samples and donor information (Macleod et al., 2009).

- Epigenetics - Epigenetics is the study of heritable changes in the phenotype or gene expression by something (usually environmental) other than changes in the DNA sequence (Holliday, 1990).
- Next Generation Sequencing - The genome sequencing efforts started back in 1991 and they entailed a laborious radiation-based method. According to Mun-Keat Looi (2009), it used to take half a million British pounds (£'s) and 3 years to generate one bacterial genome sequence. Now one can do 100 bacterial genomes with a single run and newer technologies can do things that were inconceivable 2 or 3 years ago. Without going into all the details (Jostins, 2009; Mun-Keat Looi, 2009), the newer sequencing techniques use repeated cleaving (i.e., cutting up the DNA into fragments), reading the base (by recognizing a specific colored dye terminator), and moving to the next base in sequence without moving the piece of the segmented DNA itself, but regenerating the whole segment in place. By having the DNA cut up into several hundreds or thousands of contiguous pieces (called beads) in an array and using automated techniques, machines are now able to read one gigabase of DNA sequence in a couple of days at a cost of $0.02 per 1000 bases.

Post-genomic era accomplishment

Lesch (2004) summarizes the post-genomic era research into the genetics of depression as follows:

- A predominant role in individual differences in depression-related traits (i.e., phenotypes) is confirmed to be played by the variation in genetic expression among individuals.
- Functional Neuroimaging has substantiated the gene-phenotype correlations.
- Concept of gene networks controlling brain development has been recognized.
- Gene-environment interactions in humans and animal models give rise to both complexity of traits and establishment of specific phenotypes.

Hirschhorn and Daly (2005) acknowledge that genetic association studies offer a potentially powerful approach for mapping causal genes with modest effects but also warn that only a small number of genes can be studied at one time and that the contributions made by any single gene are small. They also note the post-genomic era accomplishments in terms of deposition of millions of SNPs into public data bases, rapid improvements in genotyping technology, and the completion of the International HapMap Project (Gibbs et al., 2003).

Consensus, Challenges, and Uncertainties

The first wave of GWAS has improved our understanding of the genetic basis of many complex traits such as diabetes, prostate and breast cancers, inflammatory bowel disease, and many loci implicating these diseases have been found (McCarthy et al., 2008). In addition, the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI) have published an updated list of GWAS as a catalog which is available on their web sites. Several common variants influencing such continuous traits as lipids, height, and fat mass have also been found (Diabetes Genetic Initiative, 2007; Frayling et al., 2007; Katheriresan et al., 2008; Sanna

et al., 2008; Weedon et al., 2007). However, many challenges remain and compelling signals highlighting previously unsuspected biology and known variants of implicated genes explain only a fraction of the observed familial aggregation, thus limiting potential application for early interventions. Much work needs to be done to obtain a complete inventory of variants at each locus and determine the molecular mechanisms that are in play in order to determine the disease risk. In addition, the joint effects of the gene and the environment need to be modeled and further analyses are needed to determine the complete range of susceptibility variants to explain the clustering in families and the multifactorial nature of the complex traits (McCarthy et al., 2008).

Positives, Protocols, and Perils

Neale and Purcell (2008) discuss GWAS in a more practical way of detecting phenotypic associations of modest effect contrasting it from the previous linkage and candidate gene approaches. Accordingly, they look at GWAS as being a study of hundreds of thousands of single nucleotide polymorphisms (SNPs) in thousands of individuals looking to comprehensively survey common genetic variation. More importantly, they outline some key analytic considerations, study design, quality control, and data cleaning along with analysis and replication of results in other samples.

Presently, two major initiatives, one in UK (Wellcome Trust Case Control Consortium – WTCCC), and one in the US (Genetic Association Information Network – GAIN) are generating GWAS data. The UK study is comprised of 2,000 case sample cohorts for each of the following diseases – Tuberculosis, coronary heart disease, types 1 and 2 diabetes, Crohn's disease, rheumatoid arthritis, bipolar disorder, heart disease, and hypertension. The US study, on the

other hand, has generated approximately 600K markers for schizophrenia, bipolar disorder, diabetic nephropathy, MDD, ADHD, and psoriasis. The data used in this dissertation also come from the GAIN initiative in the US that can be found at http://www.fnih.org/gain2/home_new.shtml.

Numerous resources (SNP chips, online resources, collaboration, and consortia) have been instrumental in making GWAS popular in recent years. Commercially available SNP chips are being offered by Illumina, Affymetrix, and Perlegen to provide genotyping services for individuals and/or families. A number of internet-based resources provide information for accessing and understanding the results of a GWAS – for example, the National Center for Biotechnology Information (NCBI - http://www.ncbi.nlm.nih.gov) hosts dbGaP, the database of Genotypes and Phenotypes. Other resources include linking information to professional and peer-reviewed articles in PubMed (a searchable index of publications), GenBank (a genetic sequence database), and Entrez (a search engine for nucleotide, protein, structure, taxonomy, genome, expression, and chemical databases). In addition, a number of shared controls sets (for a case-control study) are available from WTCCC and GAIN (pending an application process).

Neale and Purcell (2008) opine that a successful GWAS is through increasing sample size that enables detection of variants of small effect. Pooling of case samples from across the UK and US as well as drawing from the experience of analysts, geneticists, and clinicians on major collections provide additional benefits. Some examples for the collaborative work include major type 2 diabetes projects DGI, FUSION, and Novartis (Saxena et al., 2007), the International

Multi-center ADHD Genetics (IMAGE) project (Bookes et al., 2006; Kuntsi et al., 2006), and the NTR and NESDA biobank project (Boomsma et al., 2008).

Some of the perils that come with GWAS are selection and use of computer software packages, study designs, and data quality control. Neale and Purcell (2008) describe a number of issues arising out of software selection, design of a required study, and cleaning up of data. Software packages like PLINK (Purcell et al., 2007), principal components analysis (Price et al., 2006), Haploview (Barrett et al., 2005), and PBAT (Lange et al., 2004) are most commonly used in GWAS as can be gleaned from the literature. Both case-control and association study designs may be employed for a GWAS; however, case-control studies have been more popular and careful selection of matched controls with a particular focus on the ancestry has been recommended. Finally, the quality of the data used for analysis determines the accuracy with which any results from a GWAS may be used to draw conclusions. Neale and Purcell (2007) point to the following important issues –

- Prior probability of a SNP showing true significance is rather low ($5x10^{-7}$ is the accepted GWAS significance as set up by the Wellcome Trust Consortium).
- A good indicator of genotype probe performance is the call rate across the sample. Call rates in excess of 85% are the expected norms used in GWAS.
- Genotype reproducibility is another key measure assessed through international sample duplication.
- Potential batch effects arise since samples (DNA sources) are not done with the same product at the same time.

- Minor allele frequency (MAF) thresholds are recommended (usually 1%) as many studies do not have the power to detect significant association for very rare variation.
- Testing for deviations from the Hardy-Weinberg equilibrium (HWE) or, more aptly, Hardy-Weinberg proportions (Elston et al., 2010) may provide further information about the validity of the genotypes from a SNP. A stringent threshold of 10^{-6} for deviation from HWE is recommended to assure data quality.
- Nonpaternity is a potential problem for family-based data when trio (i.e., patient and his or her parents) and sibship (i.e., patient and his or her siblings) designs are employed.
- Population stratification is a confounding factor in case-control, population based quantitative analysis. A correction for the inflated association may be applied using the principal component analysis (Price et al., 2006).
- The distribution of association of test statistics is a useful indicator for sources of biases.

Further data cleaning considerations are available from the literature (Chanock et al., 2007; Wellcome Trust Case Control Consortium, 2007). GWAS gives one the most extensive looks at the genome for uncovering variations of disease predisposal, while not necessarily pointing the new biochemical pathways leading to a complex and multifactorial disease. Given the difficulty of mapping the genetic variants of neuropsychiatric disease such as MDD, any GWAS for MDD and related disorders must be carried out with even greater care.

Phenomics and Epigenomics

Phenomics is another emerging transdiscipline where phenotypes are systematically studied on a genome-wide scale (Bilder et al., 2009). In the context of public health, phenomics research might be useful to relate neural systems functioning to human behavior. One of the advantages of studying phenotypes from a genome-wide perspective might be to generate phenotypes, not from a conventional laboratory or a web-based assessment of behavior, but rather, by iteratively refining phenotype assays based on prior genotype-phenotype associations. Vertically integrated research teams, novel analytic strategies, and the informatics infrastructure may help phenomics research in managing the sheer complexity involved in this sequel to genomics. Indeed, one of the nine interdisciplinary research consortia supported by NIH Roadmap Initiative that started in 2007 is funding the phenomics research at UCLA to investigate two cognitive phenotypes (Bilder et al., 2009).

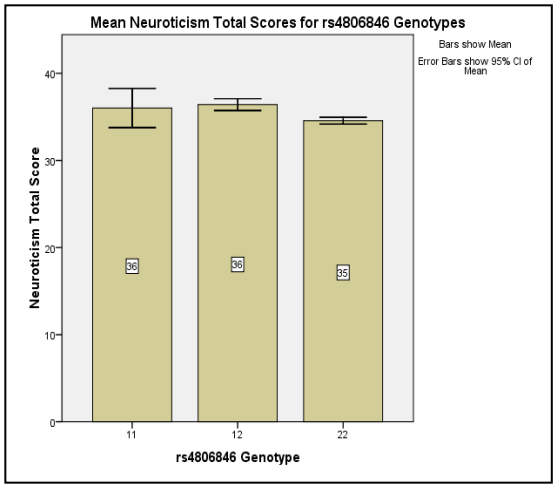
In addition to phenomics, yet another area of new research is epigenomics, the science of how genes might be modified by the environment while not changing the DNA sequence but still affecting the gene expression by affecting the chromatin structure, packaging, methylation, and other molecular modifications (Mehler, 2008). This is very important in understanding neuro-degenerative diseases such as schizophrenia, bipolar disorder, and other brain developmental (or neural regenerative) and/or evolutionary mechanisms. Further insights into these new research areas are the subject of NIH roadmap for the future aiming to develop epigenome-wide analysis as a staple of future biological characterization. Further details are available at the NIH roadmap web site - http://nihroadmap.nih.gov/epigenomics/initiatives.asp. Some researchers are even

calling these initiatives as the beginnings of the Human Phenome Project promising to realize the vision of personalized medicine and rational neuropsychiatric diagnosis and treatment (Bilder et al., 2009).

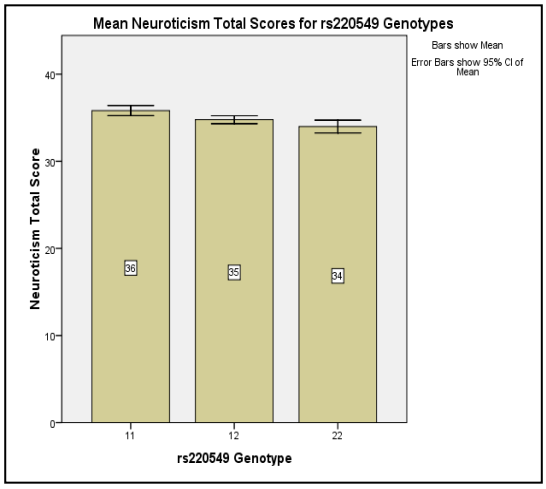
APPENDIX C

Additional Descriptive Histograms and Plots of Selected MDD Predictors

Figure 4 and Figure 5 below show the univariate and Gene x Gender analyses of neuroticism scores in our sample population.

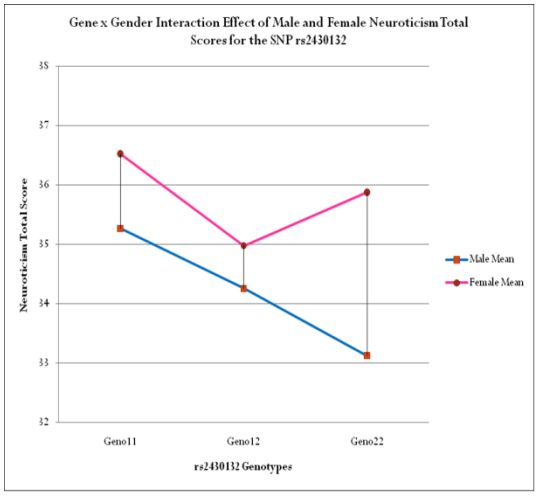


Mean Neuroticism Total Scores for the genotypes of the SNP – rs4806846.

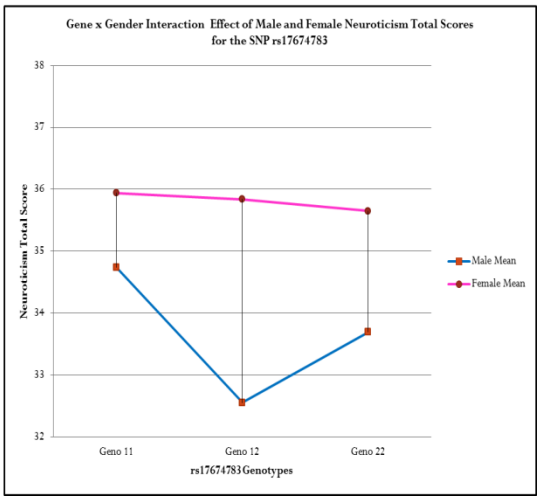


Mean Neuroticism Total Scores for the genotypes of the SNP – rs220549.

***Figure 4*. Distribution of Mean Neuroticism Scores**



Gene x Gender Interaction of Neuroticism Total Score for the SNP rs2430132 in HMCN1.



Gene x Gender Interaction of Neuroticism Total Score for the SNP rs17674783 in SNTG1.

***Figure 5*. Gene x Gender Interaction in Neuroticism Scores**

APPENDIX D

Supplementary Tables from GWAS Studies of Neuroticism and Age at Onset

Table D-1

Genome-Wide Association Study Results of 34 Significant SNPs for Neuroticism in MDD*

CHR	Reference SNP	Position[a]	Gene Name	Gene Location	P-value Total[b]	P-value Male[c]	P-value Female[d]
19	rs4806846	2361484	TMPRSS9	19p13.3	7.79E-06	0.001109	0.002405
12	rs220549	13828587	GRIN2B	12p12	1.05E-05	0.05337	0.0002437
17	rs1046329	45603905	SGCA	17q21	1.37E-05	0.1835	2.88E-05
2	rs6757820	147996163	ACVR2A	2q22.3	1.51E-05	0.003489	0.002758
6	rs4510687	96219582	(MANEA & FUT9)	6q16.1	2.49E-05	0.1229	0.0001593
3	rs985280	64083893	PSMD6	3p14.1	2.61E-05	0.07875	0.0009192
6	rs4342432	96195874			2.96E-05	0.1475	0.0001645
16	rs9807002	26184920			3.05E-05	0.1409	0.000289
14	rs9323020	21228477	Near EDDM3B	14q11.2	3.29E-05	0.06499	0.0001355
17	rs8068962	9254866	STX8	17p12	3.32E-05	0.05502	0.001416
2	rs2344734	152400075	NEB	2q22	3.61E-05	0.1872	0.0005467
2	rs12620464	101396078			4.09E-05	0.04817	0.0006929
8	rs1608361	123732642	LOC100131552	8q24.13	4.13E-05	0.08719	0.004728
2	rs10930046	162846229			4.68E-05	0.06334	0.001034
5	rs6894463	3266546			4.81E-05	0.001539	0.003872
17	rs2301685	32452234			4.94E-05	0.1001	0.002558
20	rs6065392	39870080	ZHX3	20q12	5.02E-05	0.06333	0.001298
11	rs10835855	32165526			5.15E-05	0.06333	0.001298
8	rs16915079	51685259	Near CYCSP22	8q11.22	5.54E-05	0.03473	0.006221

Table D-1 (Continued)

3	rs9834457	69604258	Near LOC642487	3p14.1	6.10E-05	0.02482	0.000309
9	rs4837349	131024550	GOLGA2	9q34.11	6.42E-05	0.04986	0.0004275
6	rs12198780	7501543	RPS26P29	6p24.3	6.53E-05	0.05282	0.002184
8	rs4871135	121924955			7.33E-05	0.1549	0.0005861
9	rs4740720	3196279			7.63E-05	0.4673	0.0003664
9	rs4836644	131024527	GOLGA2	9q34.11	8.00E-05	0.08905	0.0002971
5	rs298024	58977433			8.02E-05	0.05118	0.006706
17	rs17651665	32419258	TLK2P1	17q12	8.43E-05	0.122	0.003107
1	rs6702388	163236984			8.49E-05	0.01127	0.0007732
23	rs6525630	144651865	LOC347422	Xq27.3	8.67E-05	0.03364	0.1668
13	rs1927860	52964897	THSD1	13q14.3	9.23E-05	0.2213	0.001113
18	rs8096705	39769843			9.30E-05	0.01499	0.02828
8	rs7013587	52954726	ST18	8q11.23	9.53E-05	0.0183	0.07478
8	rs7818832	52856493			9.70E-05	0.0235	0.07048
23	rs5930284	127728822			9.97E-05	0.007143	0.1926

* CHR = Chromosome Number; SNP = Single Nucleotide Polymorphism

[a] Physical position is based on NCBI genome build 36.3.

[b] Wald test asymptotic P-value for the association of neuroticism with MDD in the total sample.

[c] Wald test asymptotic P-value for the association of neuroticism with MDD in the male sample.

[d] Wald test asymptotic P-value for the association of neuroticism with MDD in the female sample.

Table D-2

Genome-Wide Association Study Results of 46 Significant SNPs for Gene x Gender Interaction of Neuroticism*

CHR	Reference SNP	Gene Name	Gene Location	P-assoc Total[a]	P-assoc Male[b]	P-assoc Female[c]	P_GxE
1	rs2430132	HMCN1	1q25.3-q31.1	0.351	0.4283	0.6643	5.37E-06
8	rs17674783	SNTG1	8q11-q12	0.8544	0.5047	0.999	6.92E-06

Table D-2 (Continued)

17	rs2612437	(MAP2K4 & MYOCD)		0.9471	0.526	0.5177	9.30E-06
12	rs11614157	(PRP1R12A & C12orf64)		0.0612	0.3835	0.2374	1.48E-05
4	rs244052	ENPEP	4q25	0.2467	0.9248	0.4139	1.53E-05
8	rs391798	SNTG1	8q11-q12	0.9659	0.483	0.8758	1.53E-05
4	rs13146994	ATP8A1	4p14-p12	0.3128	0.2118	0.4046	1.66E-05
20	rs6072394			0.2119	0.9787	0.3763	1.77E-05
3	rs9834626	RPL10AP6	3p14.2	0.6774	0.454	0.5157	1.78E-05
15	rs870335	SMAD3	15q22.33	0.9354	0.4891	0.411	2.28E-05
20	rs2326424	ADRA1D	20p13	0.04644	0.01311	0.2266	2.69E-05
14	rs17111534	LOC100288480	14q11.2	0.3991	0.9245	0.1688	2.78E-05
2	rs12692083	ISCA1P6	2q14.3	0.7659	0.2809	0.09403	2.83E-05
15	rs2415037	SMAD3	15q22.33	0.8592	0.5387	0.4015	3.00E-05
15	rs1532758	SMAD3	15q22.33	0.2269	0.973	0.07676	3.50E-05
8	rs1023979			0.2049	0.377	0.5902	3.65E-05
6	rs11970411	TNFA1P3	6q23	0.9345	0.8628	0.8696	3.66E-05
5	rs6595727			0.1771	0.02356	0.9237	3.68E-05
8	rs4412393			0.6117	0.06589	0.3106	3.75E-05
8	rs7832224	EFR3A	8q24.22	0.2561	0.8041	0.2023	4.21E-05
8	rs203929			0.4532	0.9125	0.7004	4.25E-05
8	rs12541061			0.1925	0.3405	0.5777	4.28E-05
8	rs12478804	B3GALT1	2q24.3	0.05783	0.8438	0.3738	4.73E-05
2	rs12480036			0.2312	0.9315	0.4074	4.73E-05
8	rs203953			0.7064	0.7489	0.9391	4.79E-05
8	rs12543181			0.2308	0.2388	0.1487	5.11E-05
13	rs3905070	EFNB2	13q33	0.06431	0.2778	0.1601	5.44E-05
15	rs2946543	SYNGR2P1	15q13.2	0.3641	0.731	0.592	5.54E-05
1	rs11163755			0.6195	0.1876	0.3995	5.72E-05
8	rs2942805			0.786	0.7985	0.436	5.73E-05
18	rs10502673	GALNT1	18q12.1	0.05579	0.007379	0.1413	5.74E-05

Table D-2 (Continued)

15	rs11539519			0.7383	0.0272	0.1347	5.76E-05
5	rs6555484	ADCY2	5p15.3	0.1239	0.8236	0.01417	5.92E-05
3	rs2717294	ALG3	3q27.1	0.04133	0.4183	0.03417	6.02E-05
2	rs10194273			0.912	0.5944	0.8303	6.34E-05
8	rs4623452			0.8869	0.06411	0.1713	6.92E-05
17	rs7213039	TTLL6	17q21.32	0.2436	0.2449	0.7986	7.76E-05
4	rs3829753	DUX4C	4q35.2	0.6033	0.4975	0.7407	7.94E-05
8	rs1821122			0.169	0.2281	0.3058	8.02E-05
13	rs1325396			0.4129	0.5374	0.07007	8.28E-05
19	rs8112887			0.3088	0.7539	0.5337	8.31E-05
17	rs7209419	TTLL6	17q21.32	0.2892	0.2922	0.8286	8.37E-05
8	rs3104997			0.06863	0.06077	0.3434	8.54E-05
4	rs4532257	TN1P3	4q27	0.4698	0.4746	0.5356	8.68E-05
11	rs2457249	MIR708	11q14.1	0.3407	0.5289	0.1679	8.82E-05
3	rs1685442			0.8079	0.6051	0.2961	9.03E-05

[*] CHR = Chromosome Number; SNP = Single Nucleotide Polymorphism; MDD = Major Depressive Disorder;

[a, b, c] Wald test asymptotic P-value for the association of neuroticism with MDD in the total, male, and female sample.

Table D-3

***Genome-Wide Association Study Results of All Significant SNPs for ASPD in the Region of Implicated Genes for Neuroticism*[*]**

CHR	Reference SNP	Gene Name	P-Neoneu Total[a]	P-Neoneu Male[b]	P-Neoneu Female[c]	P-MDD-total[d]	P-MDD-male[e]	P-MDD-female[f]	OZ-ASPD p-value[g]
19	rs4806846	TMPRSS9	7.79E-06	0.0011	0.0024	0.002	0.061	0.0185	N/A
19	rs4453628	TMPRSS9	N/A	N/A	N/A	N/A	N/A	N/A	0.0538
12	rs220549	GRIN2B	1.05E-05	0.0534	0.0002	0.055	0.971	0.0145	0.2521
12	rs1158541	GRIN2B	N/A	N/A	N/A	N/A	N/A	N/A	0.0577

Table D-3 (Continued)

12	rs2192973	GRIN2B	N/A	N/A	N/A	N/A	N/A	N/A	0.0624
12	rs11055608	GRIN2B	0.9412	0.8302	0.5511	0.759	0.348	0.2884	0.0671
12	rs2216127	GRIN2B	0.0824	0.4308	0.0798	0.039	0.105	0.1781	0.0484
12	rs7974275	GRIN2B	0.865	0.9641	0.4977	0.461	0.757	0.2365	0.0194
14	rs9323020	EDDM3B	3.29E-05	0.0650	0.0001	0.008	0.234	0.0173	0.2237
14	rs10146821	OR10G2	0.0871	0.0701	0.7192	0.568	0.319	0.8758	0.0417
14	rs10483255	RANBP20P	N/A	N/A	N/A	N/A	N/A	N/A	0.0269
14	rs2874103	OR4E2	0.1635	0.8785	0.0596	0.223	0.329	0.3568	0.0802
14	rs970382	OR4E2	0.1826	0.8555	0.0638	0.225	0.305	0.3741	0.0901
14	rs6571990	EDDM3A	N/A	N/A	N/A	N/A	N/A	N/A	0.0028
14	rs8013476	LOC100287557	N/A	N/A	N/A	N/A	N/A	N/A	0.0735
14	rs718433	RNASE1	0.7877	0.6430	0.5419	0.841	0.732	0.5755	0.0573
14	rs2141971	RNASE1	0.7169	0.2254	0.3192	0.397	0.778	0.1487	0.0045
17	rs8068962	STX8	3.32E-05	0.0550	0.0014	0.013	0.297	0.0156	N/A
17	rs7216692	STX8	N/A	N/A	N/A	N/A	N/A	N/A	0.0707
17	rs6503203	STX8	N/A	N/A	N/A	N/A	N/A	N/A	0.0453
17	rs9904459	STX8	N/A	N/A	N/A	N/A	N/A	N/A	0.0661
17	rs7209802	STX8	N/A	N/A	N/A	N/A	N/A	N/A	0.0295
17	rs4791838	STX8	N/A	N/A	N/A	N/A	N/A	N/A	0.0014
17	rs3891720	STX8	N/A	N/A	N/A	N/A	N/A	N/A	0.0224
2	rs2344734	NEB	3.61E-05	0.1872	0.0005	0.076	0.789	0.0279	N/A
8	rs16915079	CYCSP22	5.54E-05	0.0347	0.0062	0.066	0.306	0.1175	N/A
8	rs12375352	SNTG1	N/A	N/A	N/A	N/A	N/A	N/A	0.0781

[*] CHR = Chromosome Number; SNP = Single Nucleotide Polymorphism; MDD = Major Depressive Disorder; ASPD = Anti-Social Personality Disorder.

[a, b, c] Wald test asymptotic P-value for the association of neuroticism with MDD in the total, male, and female sample.

[d, e, f] Wald test asymptotic P-value for the association with MDD in the total, male, and female sample.

[g] Wald test asymptotic P-value for the association with ASPD in the OZALC sample.

Table D-4

Genome-Wide Association Study Results of All Significant SNPs for Age at Onset of MDD*

CHR	Reference SNP	Gene Name	Gene Location	P-MDD-Total[a]	P-MDD-Male[b]	P-MDD-Female[c]	P-Neoneu[d]	P-AAO[e]
23	rs734253	GPR143	Xp22.3	0.3747	N/A	0.8116	0.2357	2.54E-12
23	rs6641545	ASS1P4	Xp22.33	0.7671	0.2733	0.2551	0.64	9.75E-10
23	rs5983118	MXRA5	Xp22.33	0.398	N/A	0.1451	0.5025	5.67E-08
23	rs17267483	HDX	Xq22.1	0.4358	0.01471	0.7501	0.1034	1.59E-07
23	rs5971527	CXorf29	Xp21.2	0.8627	0.01623	0.09991	0.6139	1.70E-07
23	rs6654462	N/A	N/A	0.3369	0.7451	0.1007	0.1432	3.99E-07
23	rs5928558	ENSG00000233571	N/A	0.04561	0.05098	0.1785	0.02767	5.45E-07
23	rs6629704	PTCHD1	Xp22.11	0.603	0.3504	0.3814	0.02766	1.01E-06
23	rs2490768	ZCCHC16	Xq23	0.03009	0.961	0.0249	0.9888	4.17E-06
5	rs32589	PPARGC1B	5q32	0.7074	0.00509	0.01677	0.5262	4.72E-06
23	rs5918490	SRPX	Xp21.1	0.00262	0.9286	0.00181	0.3342	4.97E-06
23	rs5935125	ARHGAP6	Xp22.3	0.07788	0.00667	0.5328	0.01839	5.38E-06
23	rs17343481	ENSG00000238327	N/A	0.7615	0.3565	0.5175	0.1023	5.40E-06
6	rs9294523	RPL5P19	6q15	0.5052	0.09523	0.04591	0.6942	5.83E-06
23	rs12393485	N/A	N/A	0.439	0.01915	0.7968	0.4857	5.92E-06
23	rs1468422	NDUFA1	Xq24	0.848	0.2351	0.5242	0.00049	7.46E-06
5	rs10515638	PPARGC1B	5q32	0.8893	0.01328	0.1193	0.2302	9.77E-06
5	rs26125	PPARGC1B	5q32	0.494	0.00681	0.00659	0.5332	1.04E-05

Table D-4 (Continued)

23	rs3861732	LOC100128601	Xq22.1	0.6901	0.1144	0.1834	0.9338	1.06E-05
23	rs17219420	HDX	Xq21.1	0.6002	0.04163	0.6945	0.153	1.25E-05
23	rs7883591	ZCCHC16	Xq23	0.2569	0.4755	0.1334	0.7061	1.32E-05
23	rs5939057	ZCCHC16	Xq23	0.284	0.5247	0.1672	0.6689	1.34E-05
5	rs17110375	PPARGC1B	5q32	0.62	0.07167	0.06859	0.07791	1.42E-05
23	rs16980772	SCML2	Xp22	0.5505	0.04284	0.1247	0.6471	1.65E-05
23	rs12846646	SAT1	Xp22.11	0.6008	0.2697	0.8674	0.1614	1.69E-05
23	rs17245924	N/A	N/A	0.05341	0.02078	0.2747	0.000149	1.73E-05
23	rs4562492	ENSG00000238327	N/A	0.873	0.4318	0.9855	0.1903	2.27E-05
23	rs11092550	NRK	Xq22.3	0.2293	0.9968	0.2005	0.8261	2.37E-05
23	rs5909462	SCML2	Xp22	0.5655	0.01616	0.07903	0.4144	2.53E-05
23	rs17301374	CCT4P2	Xq11.2	0.04026	0.8737	0.02666	0.7771	2.68E-05
23	rs5909169	SCML2	Xp22	0.4384	0.07897	0.1091	0.735	2.78E-05
23	rs1500725	IL1RAPL1	Xp22.1-p21.3	0.8097	0.08988	0.3676	0.9017	3.10E-05
23	rs4378112	SHC1P1	Xq11.1	0.3433	0.376	0.1647	0.7435	3.23E-05
6	rs6570806	N/A	N/A	0.1601	0.08379	0.01287	0.8842	3.23E-05
23	rs5970992	DDX53-	Xp22.11	0.5179	0.015	0.05175	0.7221	3.27E-05
8	rs16938568	RDH10	8q21.11	0.6015	0.1612	0.9382	0.1362	3.36E-05
12	rs11107320	LOC10013212 6	12q22	0.4294	0.2577	0.1283	0.1142	3.41E-05
11	rs3016415	GRIK4	11q22.3	0.6102	0.01755	0.02508	0.8042	3.50E-05
6	rs10947795	KCNK16	6p21.2-p21.1	0.5245	0.01124	0.4695	0.2657	3.71E-05
22	rs34074034	TTLL12	22q13.31	0.2529	0.01343	0.9774	0.4047	3.85E-05

Table D-4 (Continued)

23	rs35401330	BEND2	Xp22.13	0.3709	0.05472	0.06403	0.5005	3.89E-05
5	rs11738269	CANX	5q35	0.8539	0.02931	0.1304	0.2636	3.95E-05
6	rs1876155	SIM1	6q16.3-q21	0.1498	0.06722	0.5046	0.9709	3.97E-05
23	rs590779	MIR514-3	Xq27.3	0.7882	0.1753	0.8695	0.745	3.99E-05
15	rs11630901	RPAP1	15q15.1	0.1481	0.00367	0.9168	0.5768	4.07E-05
23	rs4279777	FRMPD4	Xp22.2	0.699	0.02068	0.4697	0.28	4.37E-05
23	rs989059	EDA	Xq12-q13.1	0.9774	0.1451	0.4568	0.1792	4.45E-05
8	rs17211245	RDH10	8q21.11	0.3178	0.1653	0.6283	0.03658	4.70E-05
11	rs4573685	GRIK4	11q22.3	0.8897	0.02289	0.09412	0.2065	5.02E-05
23	rs4824251	MIR514-3	Xq27.3	0.4736	0.2621	0.1843	0.53	5.27E-05
23	rs5907414	N/A	N/A	0.3992	0.3598	0.5648	0.05666	5.40E-05
23	rs5927283	LOC392439	Xp21.1	0.08112	0.00244	0.6162	0.01826	5.64E-05
10	rs11018112	N/A	N/A	0.8165	0.1413	0.311	0.4234	5.64E-05
23	rs5987957	PLS3	Xq23	0.7859	0.8008	0.6477	0.317	5.68E-05
23	rs7472960	ZCCHC16	Xq23	0.2121	0.03187	0.6583	0.4743	5.77E-05
23	rs2266835	MAMLD1	Xq28	0.1681	0.1089	0.3787	0.2941	5.79E-05
9	rs870186			0.6415	0.2125	0.2563	0.781	5.85E-05
10	rs10824745	LOC283050	10q22.3	0.7293	0.04759	0.08453	0.7109	6.25E-05
23	rs1076616	N/A	N/A	0.1157	0.2694	0.02248	0.3392	6.30E-05
23	rs6635622	N/A	N/A	0.7288	0.1708	0.3339	0.06482	6.61E-05
6	rs885463	LOC100127900	6q13	0.2756	0.02632	0.00426	0.3531	6.73E-05
1	rs10888527	ZNF687	1q21.3	0.3616	0.01012	0.00621	0.2164	7.28E-05
23	rs7883119			0.4177	0.01898	0.8102	0.6838	7.53E-05

Table D-4 (Continued)

11	rs4617622	LOC120364	11q23.1	0.8802	0.0819	0.3866	0.07239	7.56E-05
23	rs12687753	RRM2P3	Xp11.4	0.3099	0.2069	0.08466	0.2026	7.88E-05
23	rs5933909	ARHAGAP6	Xp22.3	0.05047	0.006547	0.3721	0.003498	7.89E-05
20	rs397883	TBC1D20	20p11.21	0.3593	0.01551	0.5936	0.5771	8.07E-05
23	rs1476012	ALAS2	Xp11.21	0.36	0.3883	0.1781	0.7389	8.22E-05
23	rs222108	KLHL4	Xq21.3	0.5008	0.06171	0.8959	0.8097	8.37E-05
16	rs8048267	ZFHX3	16q22.3	0.2389	0.03201	0.002845	0.5397	8.37E-05
8	rs7837426	RDH10	8q21.11	0.4974	0.2174	0.8439	0.02799	8.49E-05
2	rs13394360	N/A	N/A	0.4401	0.3643	0.1471	0.8868	8.90E-05
5	rs17617710	CANX	5q35	0.9256	0.02371	0.2289	0.4248	9.31E-05
23	rs5972323	DMD	Xq34	0.2193	0.05967	0.5964	0.01431	9.32E-05
23	rs5984014	N/A	N/A	0.3235	0.1835	0.09834	0.8646	9.81E-05
5	rs6895902	CANX	5q35	0.7987	0.0295	0.3488	0.338	9.91E-05

[*] CHR = Chromosome Number; SNP = Single Nucleotide Polymorphism; MDD = Major Depressive Disorder;

[a, b, c] Wald test asymptotic P-value for the association with MDD in the total, male, and female sample.

[d] Wald test asymptotic P-value for the association of neuroticism with MDD in the total sample.

[e] Wald test asymptotic P-value for the association of age at onset with MDD in the total sample.

VITA

NAGESH ARAGAM

Personal Data: Date of Birth: May 29, 1952

Place of Birth: Bangalore, India

Marital Status: Married

Education:

- Dr.P.H. (Doctor of Public Health), East Tennessee State University, Tennessee, USA, 2011.
- M.Sc.C.S. (Master of Computer Science), University of New Brunswick, Fredericton, New Brunswick, Canada, 1979.
- M. Tech. (Master of Technology), Indian Institute of Technology, Madras, India, 1975.
- B.E. (Bachelor of Engineering), Bangalore University, Bangalore, India, 1973.

Professional Experience:

- Postdoctoral Trainee, Dept. of Infectious Diseases, School of Medicine, University of North Carolina, Chapel Hill, North Carolina, USA. May 2011 – Present.
- Business Systems Analyst, Mountain States Health Alliance, Johnson City, Tennessee, USA, Aug 2008 – Sep 2009.
- Research Software Consultant, Oak Ridge National Laboratories, Oak Ridge, Tennessee, USA, Dec 2004 – June 2005.

- Research Specialist, IRIS Laboratory, University of Tennessee, Knoxville, Tennessee, USA, Apr 2003 – Dec 2004.
- Independent Consultant, TMS Consulting, Johnson City, Tennessee, USA, Jan 2001 – April 2004.
- Graphics Product Manager & Senior Systems Architect, Tech Source Inc., Altamonte Springs, Florida, USA, Mar 1999 – Jan 2002.
- Senior Consultant, Motorola Communications Enterprise Division, Richmond, British Columbia, Canada, Jul 1998 – Jan 1999.
- Senior Consultant, Hughes Aircraft of Canada, Richmond, British Columbia, Canada, Mar 1993 – Nov 1997.
- Various senior technical positions at Sun Microsystems, IBM, Coopers & Lybrand, The Foxboro Co., and Systems & Applied Sciences Corp., USA, Mar 1981 – July 1992.

Honors and Awards:

- Recipient of several Graduate Research and Teaching Assistantships;
- University Ranker all through my undergraduate studies in India;
- Professional recognition through bonuses, work citations, and an idea entry as a precursor to a patent application.
- Outstanding Doctoral Student Award, Dept. of Biostatistics and Epidemiology, East Tennessee State University, 2011.

- Several Journal Publications, Conference Posters, and Presentations.