INVESTIGATION OF SCHIZOPHRENIA RELATED GENES AND PATHWAYS THROUGH GENOME WIDE ASSOCIATION STUDIES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

HÜSEYİN ALPER DÖM

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
BIOINFORMATICS

JANUARY 2013

INVESTIGATION OF SCHIZOPHRENIA RELATED GENES AND PATHWAYS THROUGH GENOME WIDE ASSOCIATION STUDIES

Submitted by **Hüseyin Alper DÖM** in partial fulfillment of the requirements for the degree of **Master of Science, Bioinformatics Program**,
Middle East Technical University by,

Prof. Dr. Nazife Baykal _____

Director, Informatics Institute

Assist. Prof. Dr. Yeşim Aydın Son _____

Head of department, Medical Informatics, METU

Assist. Prof. Dr. Yeşim Aydın Son _____

Supervisor, Health Informatics, METU

Examining Committee Members

Prof. Dr. Özcan Uzun _____

GMMA, PSYCHIATRY

Assist. Prof. Dr. Yeşim Aydın Son _____

METU, MIN

Assist. Prof. Dr. Aybar Can Acar _____

METU, MIN

Assoc. Prof. Dr. Tolga Can _____

METU, CENG

Assoc. Prof. Dr. Vilda Purutçuoğlu Gazi _____

METU, STAT

Date: 30.01.2013

I hereby declare that all information in this document has been obtained and resented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Hüseyin Alper Döm
Signature :

ABSTRACT

INVESTIGATION OF SCHIZOPHRENIA RELATED GENES AND PATHWAYS THROUGH GENOME WIDE ASSOCIATION STUDIES

Hüseyin Alper Döm

M.Sc., Bioinformatics Program

Advisor: Assist. Prof. Dr. Yeşim Aydın Son

January 2013, 124 pages

Schizophrenia is a complex mental disorder that is commonly characterized as deterioration of intellectual process and emotional responses and affects 1% of any given population. SNPs are single nucleotide changes that take place in DNA sequences and establish the major percentage of genomic variations. In this study, our goal was to identify SNPs as genomic markers that are related with schizophrenia and investigate the genes and pathways that are identified through the analysis of SNPs. Genome wide association studies (GWAS) analyse the whole genome of case and control groups to identify genetic variations and search for related markers, like SNPs. GWASs are the most common method to investigate genetic causes of a complex disease such as

schizophrenia because regular linkage studies are not sufficient. Out of 909,622 SNPs analysis of the dbGAP Schizophrenia genotyping data identified 25,555 SNPs with a p-value $5x10^{-5}$. Next, combined p-value approach to identify associated genes and pathways and AHP based prioritization to select biologically relevant SNPs with high statistical association are used through METU-SNP software. 6,000 SNPs had an AHP score above 0.4, which mapped to 2,500 genes suggested to be associated with schizophrenia and related conditions. In addition to previously described neurological pathways, pathway and network analysis showed enrichment of two pathways.

Melanogenesis and vascular smooth muscle contraction pathways were found to be highly associated with schizophrenia. We have also shown that these pathways can be organized in one biological network, which might have a role in the molecular etiology of schizophrenia. Overall analysis results revealed two novel candidate genes SOS1 and GUCY1B3 that have a possible relation with schizophrenia.

Keywords: *METU-SNP, AHP, GWAS, pathway and gene discovery, schizophrenia.*

ÖZ

ŞİZOFRENİ İLE İLGİLİ GENLERİN VE YOLAKLARIN GENOM ÇAPINDA İLİŞKİLENDİRME ÇALIŞMALARI (GWAS) ARACILIĞIYLA İNCELENMESİ

Hüseyin Alper Döm

Yüksek Lisans, Bioenformatik Programı

Tez Danışmanı: Yrd. Doç. Dr. Yeşim Aydın Son

Ocak 2013, 124 sayfa

Şizofreni genellikle zihinsel süreçleri ve duygusal tepkileri etkileyen kompleks bir zihinsel bozukluk olarak karakterize edilir ve herhangi bir toplumun %1ini etkiler. SNPler, DNA sekansında meydana gelen tek bazlık nükleotid değişimleri olup genomik varyasyonların büyük çoğunluğunu oluştururlar. Bu çalışmada amacımız şizofreni ile ilişkili olan SNP'lerin genomik belirteçler olarak tanımlanması ve SNP'lerin analizleri sonucu tanımlanmış gen ve yolakların araştırılmasıydı. Genom çapında ilişkilendirme çalışmaları (GWAS) genetik varyasyonları tanımlamak ve SNP gibi bu varyasyonlarla ilişkili belirteçleri araştırmak için hasta ve kontrol gruplarının bütün genomlarının analiz edilmesidir. GWASlar sıradan genetik bağlantı çalışmaları yetersiz kaldığından şizofreni

gibi kompleks hastalıkların araştırılmasında en sık kullanılan yöntemlerdir. dbGAP şizofreni genotip veri setinin analizi 909.622 SNPden p-değeri $5x10^{-5}$'ten başlayan 25.555 SNP tanımladı. Sonrasında, kombine p-değeri yaklaşımıyla ilişkili gen ve yolakları tanımlamak ve de AHP'ye dayalı istatistiksel olarak yüksek anlamlı ve biyolojik olarak anlamlı SNPlerin önceliklendirilmesi çalışmaları için METU-SNP yazılımı kullanılmıştır. AHP puanı 0,4 ve üzerisi olan ve 2.500 gene atanan 6.000 SNPnin şizofreni ve şizofreniyle ilgili kondisyonlarla ilişkili olduğu önerilmiştir. Daha önce tanımlanan nörolojik yolaklara ilaveten, yolak ve ağ analizleri zenginleştirilmiş iki yeni yolak ortaya çıkardı.

Melanogenesis ve damarsal düz kas kasılma (vascular smooth muscle contraction) yolaklarının şizofreni ile yüksek ilgisinin olduğu bulundu. Ayrıca bu yolakların şizofreninin moleküler etiyolojisinde rol oynayabilecek tek bir biyolojik ağ oluşturduğunu gösterdik. Tüm analizlerin sonuçları neticesinde de şizofreni ile muhtemel ilgisi bulunan SOS1 ve GUCY1B3 genleri iki yeni aday gen olarak ortaya çıkmıştır.

Anahtar kelimeler: *METU-SNP, AHP, GWAS, yolak ve gen keşfi, şizofreni*

To My Family and Fiancée

ACKNOWLEDGEMENTS

First and foremost I would like to express my deepest gratitude and thanks to my supervisor Assist. Prof. Dr. Yeşim Aydın Son. Her contributions to both my academic career and my life are invaluable and irreplaceable. She has supported and guided me from the very beginning of the programme. She is always very nice, helpful and has positive attitude towards me. She has also supported me throughout my thesis with her patience and knowledge whilst allowing me space to work in my own way. One simply could not wish for a better or friendlier 'super'visor.

I am grateful to Prof. Dr. İnci Togan for all her support during my first year in the programme. Working as a project assistant in her group and under TURK-HAYGEN project provided an opportunity to me develop my molecular biology laboratory skills. That first year contributed me a lot academically and professionally, and also provided financial support which helped me to pursue my academic career. Her just and knowledge will always guide me.

I also thank Dr. Gürkan Üstünkar for developing METU-SNP software, which forms the backbone of this study.

I am thankful to Mehmet Ali Döke, Onur Baloğlu, Alper Mutlu, Cansaran Saygılı and Mustafa Demir for being few among the many. Your friendships mean a lot to me.

I extend my special thanks to Damla (Güldağ) Taş for her concern and endless friendship.

I am greatly thankful to Hande Acar for being the definition of sincere and dearest friend

and appreciate every single thing and every glimpse of time we share together.

I cannot express enough how I'm thankful and grateful to my brothers in life Tutku Koç, Süleyman Taşkın and Onur Büyükpolat. It's great to know you, be a part of your life, and feel the presence of you right here by my side. It is a privilege of me having you in my life.

I am very happy and excited for my soon to be family-in-law for their warm welcome to their lives. I am thankful for making me feel at home every time I'm with you and for your sincere concern, interest and care.

I am also grateful to my supporting family for being there every time I need. I thank my parents for devoting their life on me. I thank my brother for being there understanding me where no one knows.

I am deeply and ultimately grateful to my dear, Özge. You are unique in so many ways that dazzle me. You have been my sanctuary when I needed. You are the peace, tranquility, harmony and serenity in my life that I've been long lasting. You have brought joy and cheer that was missing. I appreciate for all these and much, much more.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

PREFACE

In this research we have investigated schizophrenia related genes and pathways through genome wide association studies. The genotyping data from 1,385 subjects as control and 1,576 cases are downloaded from dbGaP of NCBI. We have used several well-known bioinformatics tools that are publicly available as well as recently published METU-SNP software, which enables users to filter and reduce the dataset into a statistically significant and a biologically relevant subset. In order to identify novel genes and/or pathways related with schizophrenia based on the GWAS DAVID and Reactome, and GeneMANIA tools were used for pathway and network discovery respectively. We have found that melanogenesis and vascular smooth muscle contraction pathways were highly associated with schizophrenia. We have also shown that these pathways can be organized in one biological network, which might have a role in the molecular etiology of schizophrenia. Overall analysis results revealed two novel candidate genes SOS1 and GUCY1B3 that have a possible relation with schizophrenia. Our results were discussed in Chapter 3 in details and summarized in Chapter 4.

CHAPTER I

1 INTRODUCTION

1.1. Single Nucleotide Polymorphisms and Genome Wide Association Studies

1.1.1. Human Genome

Deoxyribonucleic acid (DNA) is the genetic material of all living organisms and many viruses. DNA comprises four repeating units. These are namely: Adenine (A), Guanine (G), Thymine (T) and Cytosine (C). First two nucledotides are called purines and latter ones are called pyrimidines and sum of purines equals sum of pyrimidines. In 1953, James Watson and Francis Crick suggested the most common form of DNA in which DNA has a double helix structure. In this model, two chains of DNA run antiparallel meaning one chain runs 5' to 3' direction whereas other chain runs 3' to 5' direction. The interactions that hold double helix structure together on a sugar phosphate backbone is provided via hydrogen bonding as a result of pairing of A with T (double bond) and G with C (triple bond) [1].

Genes are the functional units of heredity. They are sequences within genome and are transferred from parent to progeny. Early systematic approaches to understand genes

suggested that one gene codes for one protein and a mutation that is carried out on a gene causes the disruption of biological processes. However, later studies showed that a gene is not necessarily codes for a single protein, instead it may code for a precursors of a protein or a subunit of a protein. Thus, the "one gene: one protein" hypothesis modified into one gene: one polypeptide chain [2].

Genome is the whole set of genetic information of an organism and encoded in DNA or in RNA for many types of virus such as HIV. The genome consists of genes, which are functional units that are packed in chromosomes. Each species has a unique genome and genomes of individual members of the same species also differ from each other at some extent. Although genomes differ in size, there is no correlation between the complexity of organism and its' genome size [3].

In 1990 the Human Genome Project has been started by Aristides Patrinos, head of the Office of Biological and Environmental Research in the U.S. Department of Energy's Office of Science. Francis Collins directed the US National Institutes of Health (NIH) National Human Genome Research Institute efforts. The aim was to identify all the genes in human DNA; detect the sequence of human DNA; build a database and store the information; develop and improve tools for analysis; transfer the technology to the private sector. In June 2000 the first draft of the human genome project was ready. Later the complete working draft and its analysis are published in February 2001. Although HGP was originally planned to be successfully finalized by the end of 2005, international efforts and developing technology allowed it to be completed ahead of the schedule by April 2003 [4].

Human genome is a complex structure with ~3.2 billion of bases (haploid human genome). It relies on histones to preserve its compact feature and it only occupies a microscopic space inside the nucleus with its two meter length in form of chromatins [5]. The human genome project revealed that human genome contains roughly 20,000

protein-coding genes [6]. Today there are 25,910 unique sequences identified as human gene structures in the GRch37.p8 release of the human genome. When it is compared to rest of the genome, which comprises non-coding RNA molecules, regulatory DNA sequences, introns and other non-annotated, the coding region of the genome is accountable only for 2% of the whole haploid genome [7].

Although humans have the same genetic material in terms of biochemistry, two individual genetically varies from each other and rest of the population by at least 0.1% [5]. Variations among humans ranges from modifications in karyotype to single nucleotide changes [8]. Even monozygotic twins are genetically different due to mutations during development and gene copy number variations [9]. There are total of five major types of human genetic variations (Figure 1 [10]).

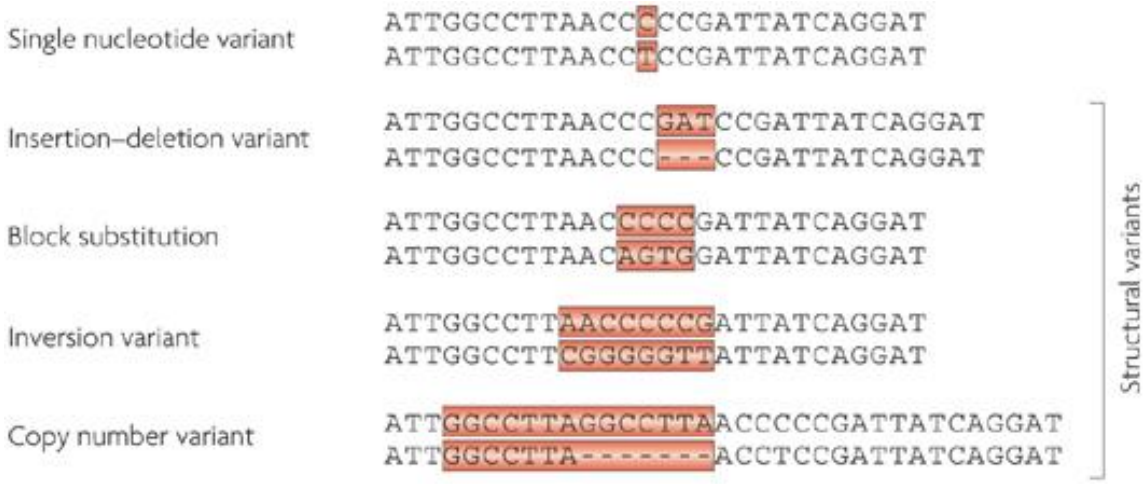

Figure 1. Classes of human genetic variants (adopted from [10])

The single nucleotide variants show alteration on only one nucleotide (A, T, G or C) of the genomic sequence. Second class of variations is insertion-deletion variants in which more than one base pair is missing in some sequences whereas the subsequence is present in others. In some cases number of missing bases can reach up to 80000 and more [10], [11]. The third class is observed when a string of neighboring bases varies

between genomes which are called block substitutions. In inversion variant cases it is shown that the order of a DNA sequence is reversed. The last class is called copy number variants as the number of repeated subsequences differs between individuals [10].

Genetic variation occurs both within and among populations. For instance a given gene can be found in different varieties in a population and such genes are called polymorphic genes. In other words, that gene has multiple alleles in a given population. As well as there are genes with multiple alleles in a population, there are also many fixed genes meaning that those genes have only single allele [12].

1.1.2. Single Nucleotide Polymorphisms

SNPs are single nucleotide changes that take place in DNA sequences. As mentioned earlier SNPs institute 90% of the genomic variations. Such variations cause modifications in traits of an individual ranging from individual's being prone to a disease to physical features of the individual [13]. For instance, occurrence of a single nucleotide change in the Apolipoprotein E (APOE) gene on chromosome 19 corresponds to an increased risk for Alzheimer's disease [14]. As it has been shown a SNP may influence risk of a common disease if it is occurred in a regulatory region of a gene. Yet, most SNPs are observed in non-coding regions of genome [15]. Depending on an SNP's location, it might have several uses considering medical and research applications. Coding SNPs that are responsible for a monogenic disorder can be used for prediction and diagnosis of the target disease and can be utilized for gene therapy approaches. Furthermore, an SNP that is responsible for alteration of the primary structure of a protein involved in drug metabolism is a perfect candidate for pharmacogenomic studies [16]. However, targeting just one SNP is not sufficient for most diseases. For instance, it has been shown that in osteoporosis focusing on a group of SNPs reveals the relation

between their role and emergence of such complex disease [17].

After the human genome project had completed in 2003 it was estimated that there were 10 million SNPs in human genome which was corresponding to one SNP per 300 bases on average [18]. However, as of 26 June 2012, in dbSNP build 137 [19] there are 53,558,214 human SNPs submitted to the database and 38,097,002 of them are validated to date [20]. The biallelic SNPs, which only have two variants, contributes their rapid increase since they can be genotyped relatively easy via automated, high-throughput genotyping methods [21]. The high numbers of SNPs identified also emphasize their importance. Moreover, they are perfect candidates as genomic markers due to their low recurrent mutation rate parallel to the evolutionary properties of human genome [21]. Today, there is a big effort going on to identify disease associated SNP profiles and genome wide studies are enabling researchers not only to detect genetic variation more thoroughly and with different perspective, but also it helps clinical applications such as disease detection, preventative and curative medicine gain momentum and develop further [22].

1.1.3. Genome Wide Association Studies

Genome wide association studies (GWAS) or whole genome association studies compares the difference between case and control groups with a high coverage that is representative of the whole genome in order to identify genetic variations by searching for related markers, namely SNPs. Furthermore, in a given GWAS finding the genetic variations is not the sole purpose. Main goal is to associate the variation profiles with a disease or a phenotype [23]. After a successful attempt to associate a genetic variation with a disease, further studies comprise development of new and better methods to fight, prevent or cure the pathological conditions [24].

Genetic linkage studies were popular prior to introduction of GWAS and they had high success rates in identification of single gene disorders [25], [26]. Through genetic linkage studies Huntington's disease and Alzheimer's disease were shown to be inherited and the genes in interest were successfully isolated [27], [28]. Yet, for complex diseases including schizophrenia, bipolar disorder, diabetes, etc. conducting linkage studies did not reveal any specific genetic factors for detection. Risch examined whether the genetic study of complex disorders reached its limits in 1996. He suggested that future of the genetics of complex diseases should be studied through association analysis since results from linkage studies between complex diseases and various loci were not replicable, and supported his hypothesis for association analysis through statistical calculations [29]. In this context, the human genome project also contributed to progress of GWAS through biobanks founded afterwards and the International HapMap Project [18].

GWAS utilizes high throughput techniques and SNPs as markers distributed throughout genome to correlate allele frequencies with trait variation in samples collected from a population. In GWAS it is hypothesized that a causal variant is located on a haplotype. Accordingly, an association should be observed between a marker allele in linkage disequilibrium (LD) between the causal variant and the trait of interest. GWAS involve only genomic markers throughout the analysis and they are considered as unbiased because previous knowledge of the trait etiology is not used as opposed to genetic studies [30].

In 2005, Klein *et al.* had carried out first successful GWAS on age related macular degeneration (AMD). They analyzed roughly 116204 SNPs in 96 cases and 50 controls. They found out that complement factor H gene (CFH) on chromosome 1 is strongly associated with AMD (nominal P-value $<10^{-7}$) [31]. Afterwards, GWA studies continued to successfully associate traits with diseases such as IL23R gene on chromosome 1p31 with Crohn’s Disease [32], lymphotoxin-alpha gene (LTA) located on chromosome 6p21 with myocardial infarction [33], TNFSF15 (the gene encoding tumor necrosis

factor superfamily, member 15) on chromosome 9q32 and inflammatory bowel disease [34], and IDE-KIF11-HHEX, EXT2-ALX4 and type 2 diabetes [35]. Yet, the most comprehensive study was achieved by Wellcome TrustCase Control Consortium in 2007. They concluded GWAS results for seven common diseases, including bipolar disorder (BD), coronary artery disease (CAD), Crohn's disease (CD), hypertension (HT), rheumatoid arthritis (RA), type I diabetes (T1D), and type II diabetes (T2D). They examined ~2,000 individuals for each of aforementioned diseases and a shared set of ~3,000 controls with 500,568 SNPs. Their findings included 24 independent association signals: 1 in BD, 1 in CAD, 9 in CD, 3 in RA, 7 in T1D and 3 in T2D with a relatively high confidence of at $p < 5x10^{-7}$ [36].

One of the main goals of the human genome project was emergence of biobanks, where biological information is stored in data warehouses. Although it is not directly related with HGP, enormous amount of data resulting from GWA studies also lead to foundation of GWAS databases. Currently there are 5 major databases: GWAS Central formerly known as HGVbaseG2P [37], GWASdb [38], NHGRI GWAS Catalog [39], Genetic Association Database [40] and PharmGKB [41]. All of these databases allow access to SNP information and its associated phenotypes. NCBI also harbors a genotype – phenotype interaction database, where the many GWAS experimental data with additional clinical information on case and controls is stored and shared with users in a controlled manner [42].

1.2. Schizophrenia

1.2.1. Epidemiology

Almost 100 years have passed since Kraepelin and Bleuler suggested the modern concept of schizophrenia into literature [43]. The detailed studies that were carried later on around 1980s concluded correspondingly as Kraepelin and Bleuler' studies [44], [45]. Schizophrenia is defined as a complex psychiatric disorder caused by genetic and environmental factors as well as their interactions [46]. Generally, schizophrenia emerges in early adult life, emerging of schizophrenia symptoms usually coincides to 20 to 35 years of age [47]. Regardless of a gender, chance of an individual's presenting schizophrenic spectrum is approximately %1 [46]. Yet, incidence rate varies across cultures. For instance people that reside or were born in an urban setting are under a greater risk for developing schizophrenia [48]. Furthermore, immigrants are considered as a high risk group, especially if they have different skin color than residents [49]. Additionally, Bresnahan *et al.* showed that African Americans are more prone to develop schizophrenia than European Americans by three times [50]. Mortality is another issue for schizophrenia patients. Suicides (5.6% of schizophrenia patients commits suicide [51]) and premature deaths [52] are the main reason for increased mortality.

1.2.2. Symptoms and Subtypes of Schizophrenia

Symptoms of schizophrenia are categorized into three as: positive (delusions, hallucinations), negative (reduced emotions, speech, interest) and disorganized symptoms (disrupted syntax and behavior), as well as cognitive symptoms in many cases [46].

There are four classical subtypes of schizophrenia; Kahlbaum's catatonia, Hecker's hebephrenia, Kraepelin's dementia paranoids and Diem's simple dementia, and schizophrenia today is considered as a merger of these disorders [53]. Kraepelin [54] states that these four disorders should be combined because it is impossible to differ them strictly. Although distinctiveness was apparent at initial stages any disorder may transform into another one as Bleuler [55] explained this transformation: 'a case which begins as a hebephrenic may be a paranoid several years later'.

Paranoid schizophrenic patients suffer from delusions and hallucinations. These delusions and hallucinations may be continuous or may last for long periods of time. Gradually, individual becomes isolated and withdrawn due to affecting delusions and hallucinations. Finally, hospitalized patient reaches a more or less stable end state. In spite of the fact that current delusions are no longer as strong as early delusions, at this stage patient is indifferent, apathetic and sustains life without any ambition or goal[53].

Second subtype of schizophrenia is composed of hebephrenic schizophrenia patients. Thought disorder stands out for this group as well as the symptoms affective flattening and incongruity. Although first symptoms are negligible such as absentmindedness, dreaminess or forgetfulness, acute symptoms are observed during the course of this type of schizophrenia like staring into space and talking to oneself. Eventually, patients start dealing with delusions. Hallucinations, on the other hand, are almost absent [53].

Catatonic schizophrenia is the third group of classical subtypes. Multiple motor, volitional and behavioral disorders are observed in this subtype. Severe symptoms as florid delusions, hallucinations and formal thought disorder follow of quietness, withdrawal, absentmindedness, etc. Catatonic schizophrenia differs from other subtypes via its catatonic nature that is originated from stupor or excitement [53].

The final group of classical subtypes of schizophrenia is the simple schizophrenia. This

form is apart from Kraepelin's original concept of schizophrenia. Simple schizophrenia patients do not develop florid symptoms. Non-specific prodrome is observed; patient becomes unsuccessful, cold, unsympathetic and alienated. The final stage of the disorder differs from patient to patient [53].

1.2.3. Risk Factors

Although symptoms, progress and subcategories of schizophrenia are well documented, exact cause of the disorder still remains unknown. As mentioned before schizophrenia is a complex psychiatric disorder and environmental and genetic factors play a role in development of schizophrenia [46].

1.2.3.1.Environmental Risk Factors

There are numerous researches that show the effect of environment. For instance, a review that was conducted on three different groups with natal complications showed a concordance with schizophrenia. The groups were selected according to the time that complications had occurred (during pregnancy, at the time of delivery and abnormal fetal growth and development). And it was reported that the magnitude of risk of developing schizophrenia was doubled in individuals that survive natal complications compared to individuals without such complications [56]. The second trimester of pregnancy is also prominent for fetus's neurodevelopment. Any complication that may occur during this period doubles the risk of offspring for developing schizophrenia [57]. Environmental factors are not limited to complications that are occurred during pregnancy and at the time of delivery. Also, environmental factors that individuals face after birth have an impact for an individual to develop schizophrenia [43]. If an urban environment is considered, a high correlation can be easily observed between urban

inhabitance and birth. Yet, inordinate representation of schizophrenia is seen in those habitats [58]. Since there is an apparent correlation of birth and urban inhabitance, cause of increased numbers of schizophrenia incidents remain unknown. Boydell & Murray tried to find cause, and examine whether the reason of correlation derived from prenatal or perinatal factors, or whether urbanicity causes psychological stress and social isolation that may risk at a later point in development [49]. A recent study with a population size of 300,000 adolescents suggests that if an individual is genetically prone to schizophrenia, stress of city life may lead to development of schizophrenia [59]. It is not only urban stress that elevates the risk of developing schizophrenia, but also use of cannabis during adolescence increase the risk by two to three times [60]. Moreover, it is not just cannabis use but also alcohol abuse and drug abuse were reported for up to 60% of chronic schizophrenic patients. 24% of patients' found to be abusing alcohol and 14% of patients' found to be abusing drugs (twice the rates in general population) prior to first admission. Although alcohol abuse had generally emerged following the first symptom of schizophrenia, the numbers vary among drug abusers. 27.5% started abusing drugs prior to first symptom, 37.9% started drugs after the emergence of first symptom and 34.6% showed simultaneous emergence of drug abuse and first symptom [61]. There are also other studies on co-morbidity of substance abuse and schizophrenia [62], [63].

1.2.3.2.Genetic Risk Factors

Schizophrenia shows high heritability rates [64]. These figures rely on the studies that were conducted and being conducted for over eighty years [43], and a large collection of data that is acquired from family, twins and adoptee studies. The results of these studies confirm the importance of genetics and its association with schizophrenia and schizophrenia spectrum disorders [65]. According to twin studies there is a concordance

of 50% between monozygotic twins and 17% between dizygotic twins [66]. One of the most important studies that show potent effect of genes belongs to Gottesman. Gottesman showed that having a second-degree relative with schizophrenia increases the risk of individual to 3-4% from 1% baseline; having first degree schizophrenic relative increases the risk up to 13% (Figure 2) [67].

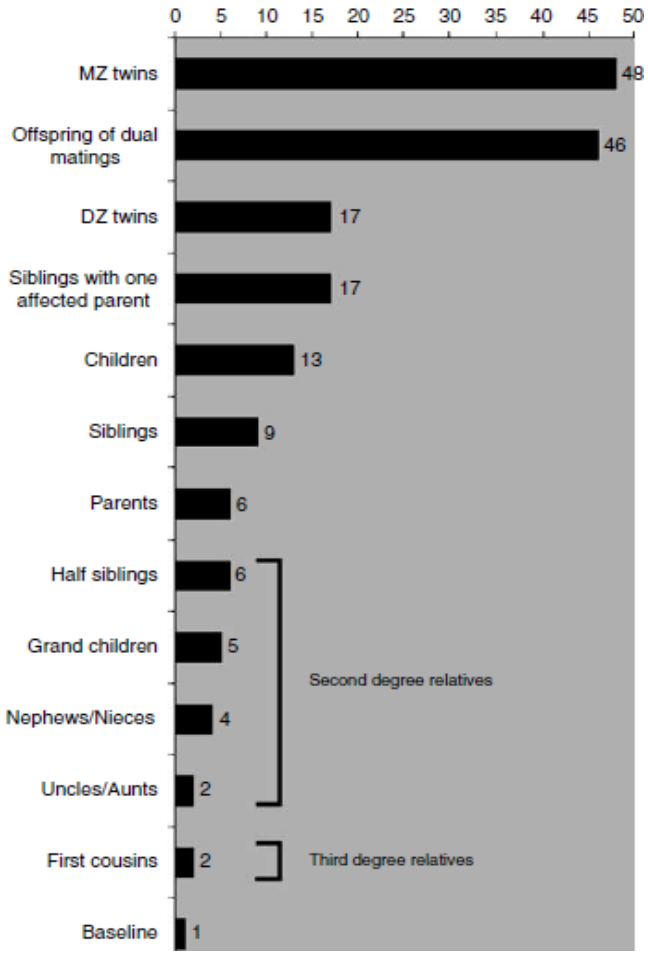

Figure 2. Risk rates of developing schizophrenia (adopted from [67])

Since perfect concordance is absent between monozygotic twins it would be accurate to state that schizophrenia is genetically mediated. Furthermore, some of the recent studies had also concluded overall risk for developing schizophrenia is more likely as a result of interaction between environment and genetics [68].

1.2.4. Linkage Studies of Schizophrenia

As common genetic studies were turned out that they are not efficient and reliable for the discovery of relevant genes and pathways of the disorder, scientists focus on genome wide association studies for schizophrenia and other complex diseases.

In a review prepared by Riley and Kendler in 2006 they listed eight linkage regions: 22q12–q13, 8p22–p21, 6p24–p22, 13q14–q32, and 6q21–q22. Yet, they mentioned the interpretation of results is controversial, since specification of replication for linkage to a complex trait remains unclear. Although the interpretation of results stands controversial, they reported results of the studies with promising candidate genes [65]. COMT [69–71], PRODH2 [72] and ZDHHC8 [73] were listed for chromosome 22q; NRG1 [74–77] and PPP3CC [78], [79] for 8p22-p21; DTNBP1 [80–89] for 6p24-p22; G72 and DAAO [90–96] for 13q14-q32; TRAR4 [97] for 6q21-q22; DISC1 [98–101] for 1q42; RGS4 [102–104] for 1q21-q22 as candidate genes. In addition, follow up studies showed that DTNBP1 interacts with AKT1 gene and further focus on this particular gene revealed association with schizophrenia [105–107], although there were some studies that showed no association [108].

Another review that was done by Tiwari *et al*. in 2010 reports 11 additional genes associated with schizophrenia [109] summarized in table 1.

Table 1. Schizophrenia related genes according to linkage studies**.**

Gene	Locus	Function	Reference
DRD2	11q23	Dopamine signaling	[110], [111]
ERBB4	2q33.3-q34	Receptor tyrosine kinase NRG1	[74], [112–114]
GARBB2	5q34	GABA signaling	[110], [115]
GRIN2B	12p12	Glutamate signaling	[110], [116]
HTR2A	13q14-21	Serotonin signaling	[110], [117], [118]
IL1B	2q14	Immune	[110], [119], [120]
NOTCH4	6p21.3	Neurodevelopment	[121–123]
NRXN1	2p16.3	Synapses	[124]
PDE4B	1p31	Synapses	[125–128]
PRODH	22q11	Glutamate synthesis	[110], [129], [130]
RELN	7q22	Synapses	[131–133]

1.2.5. Genome Wide Association of Schizophrenia

Linkage studies were the major tool before the GWAS was available. However, there was a serious debate on these studies that most of the chromosomes show linkage signals on thousand of genes. Moreover, accurately identifying a significant region by means of a linkage study is questionable due to heterogeneity and small effect sizes which are expected for schizophrenia-associated polymorphisms [109].

Linkage studies in large scales has an advantage over GWA studies by means of identifying regions with both multiple rare and common variants in susceptibility genes [134–136] as focusing on the affected members of a family within a population reduces heterogeneity. In linkage studies an affected chromosomal region is associated with a phenotype which results in thousands and millions of base pairs, thus tens or more genes. That's why linkage studies are considered to have low resolution. However, in a GWA study if the allele actually playing a role in emergence of the phenotype in interest or related with a causal allele, it is going to be reported since it will have a higher frequency. Moreover, when there is no gene, allele or region preceding the study GWAS

is the perfect tool to identify totally novel susceptibility factors [137], [138].

One of the first GWA studies was carried out in 2006 showed suggestive evidence of linkage at 8p23.3-p21.2 and 11p13.1-q14.1. Suarez *et al.* used a combined schizophrenia sample of 409 pedigrees (of 263 is European ancestry and of 146 is African American ancestry). Families in the sample were not also required to have a proband with schizophrenia but also a single or more siblings of the proband with schizophrenia or schizophrenic disorder. As a result of nonparametric multipoint linkage analysis of all families association of previously mentioned chromosome regions were found [139].

There is another GWA study from 2006 reporting a candidate gene, plexin A2 (PLXNA2) on chromosome 1q32. There were 320 patients of European descent and 325 controls in the study. The study utilized over 25,000 SNPs and found 62 markers. Yet, the most consistent one was determined as rs752016. The finding was successfully replicated in an independent sample of European Americans, in smaller subset of Latin and Asian Americans [140].

Lencz *et al.* reported identification of a novel locus associated with schizophrenia. The locus is located near colony stimulating factor, receptor 2 alpha (CSF2RA) gene in the pseudoautosomal region [141]. Further sequencing studies targeted the exonic sequences and upstream region of CSFR2A and its immediate neighbor, the interleukin 3 receptor alpha (IL3RA) revealed another association of one more polymorphism, rs6603272, in intron 5 of the IL3RA gene in independent samples of Han Chinese patients [142], [143].

According to GWA studies, certain copy number variations (CNVs) are shown to be related with high risk of schizophrenia and other psychiatric disorders [144–146]. In a GWAS of CNVs that was conducted in two phases by Stefansson, H. *et al*. in 2008. In phase I they found out nominal association between three deletions at 1q21.1, 15q11.2 and 15q13.3 and schizophrenia. They identified 66 *de novo* CNVs after the analysis of

9,878 transmissions from parents to offspring. Their sample was population based consisted of 1,433 schizophrenia cases and 33,250 controls. Their other sample for phase II covered 3,285 cases and 7,951 controls. As the results of the combined sample showed all three deletions significantly associate with schizophrenia and related psychoses [144]. In addition to previously mentioned study, The International Schizophrenia Consortium showed another association for large deletions on 22q11 but not on 15q13.3 or 1q21.1 in 2008. They included a sample consisted of 6,572 individuals whom 3,391 were patients. They used a total of 890 CNVs which occurs only once in either a case or a control. There were two independent signals pointing thioredoxin reductase 2 (TXNRD2), a gene that partially overlaps with catechol-O-methyltransferase (COMT) and the location within the genes phosphatidylinositol 4-kinase type 3 alpha (PI4KA) and serpin peptidase, clade D (heparin cofactor) member 1 (SERPIND1) [146]. In addition The International Schizophrenia Consortium reported ZNF804A, MYO18B and NOTCH4 as candidate genes in 2009 [147], which replicated the result of study of O'Donovan *et al.* that was reporting genome wide association of ZNF804A and schizophrenia in 2008 [148]. One year later, in 2009, a susceptibility gene fibroblast growth factor receptor 2 (FGFR2) on chromosome 10q25-q26 was identified by O'Donovan *et al.* as a result of a meta-analyses of 5,142 cases and 6,561 controls [149].

Another search for sequence variants was carried out by Stefansson, H. *et al.* in 2009. Their sample composed of 2,663 schizophrenia cases and 13,498 controls from eight European locations (England, Finland (Helsinki), Finland (Kuusamo), Germany (Bonn), Germany (Munich), Iceland, Italy and Scotland). There were 314,868 SNPs that were covered by the study. A significant association was exhibited for the following regions: the major histocompatibility complex (MHC) region on chromosome 6p21.3-22.1, HIST1H2BJ, PRSS16 (2 markers), PGBD1 and NOTCH4; a marker located upstream of the neurogranin gene (NRGN) on 11q24.2 and a marker in intron four of transcription

factor 4 (TCF4) on 18q21.2. Stefansson, H. *et al.* emphasized that the association of MHC region was consistent with an immune component to schizophrenia risk and the association of NRGN and TCF4 was implication of disruption of pathways involved in brain development, memory and cognition [121]. Moreover, additional findings were published in 2009 by Shi, J. *et al.* pointing a region on chromosome 6p22.1. The region of interest region includes a histone gene cluster (HIST1H2BJ, HIST1H2AG, HIST1H2BK, HIST1H4I and HIST1H2AH) and several immunity-related genes. The dataset in the study was a collection of 8,008 cases and 19,077 controls. As Shi *et al.* emphasized these studies (GWAS of CNVs) identified a region or regions of associations of common SNPs with schizophrenia, further studies were required to identify the sequence variation [46].

PLAA, ACSM1, ANK3 were found to be associated with schizophrenia as a result of GWAS by Athanasiu *et al.* in 2010. The study was performed in a Norwegian discovery sample of 201 cases and 305 controls with a focused replication analysis in a European sample comprising 2663 cases and 13,780 control subjects [150].

The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium identified five new schizophrenia loci in a two-stage study in 2011. The analysis covers 51,695 individuals. MIR137, CACNA1C, ANK3 and ITIH3-ITIH4 region were the reported findings found to be associated with schizophrenia [151].

1.2.6. Other Molecular Studies of Schizophrenia

Besides GWAS of CNVs there are also pathway analysis of GWAS data for the identification of genetic causes of schizophrenia [152–154]. Jia, P *et al.* were carried out such a study in 2010 to detect genes' combined effects of genes on mediating schizophrenia. As their results suggested glutamate metabolism pathway, TGF-beta

signaling pathway, and TNFR1 pathway which are respectively related to metabolism of glutamate, the process of apoptosis, inflammation, and immune system were found to be associated with schizophrenia. Gene Set Enrichment Analysis (GSEA) and hypergeometric test were incorporated along with Fisher's combined method to detect combined effect of genes. Their study utilized GAIN GWAS dataset for schizophrenia of unrelated European ancestry samples (1158 schizophrenia cases and 1378 controls). Yet, another study by O'Dushlaine *et al.* in 2010 found five pathways which were disparate from the results of Jia *et al.* O'Dushlaine *et al.* integrated SNP to ratio test (SRT) on the ISC GWAS and used GAIN GWAS dataset as validation dataset. They found glycan structures biosynthesis 1, cell cycle, SNARE, cell adhesion molecules (CAMs), and tight junction pathways associated with schizophrenia [155]. Thereupon, Jia *et al.* prepared a report in 2011 discussing benefits and limitations of pathway based analysis applied to GWAS datasets over a comparative study of theirs [152] and O'Dushlaine *et al.*'s. They concluded that there are several factors affecting results of GWAS data analysis through pathway based methods such as such as pathway annotations, pathway size, gene length, SNP density in genes, definition of gene boundaries, and assignment of a p-value to a gene which has multiple SNPs [153]. In this study, we also aimed to identify novel pathways and networks based on combined p-value analysis of associated SNPs for genes and pathways.

Another group that was using improved gene set enrichment analysis (i-GSEA), on 3 independent GWASs of schizophrenia (GAIN European Ancestry and African American Ancestry, and Duke Study European Ancestry) to identify pathways associated with schizophrenia and reported an association between the pathway substrate specific channel activity and schizophrenia. Their merged dataset included 3,446 cases and 3,209 controls, and over hundreds thousands of genotyped SNPs [154].

In a recent study that is unrelated to GWAS, Brzustowicz and Bassett presented miRNA-mediated risk for schizophrenia in 22q11.2 deletion syndrome (22q11.2DS)

[156]. This deletion syndrome confers a 25- to 30-fold increased risk for schizophrenia over the general population [157]. Although association between 22q11DS and schizophrenia were shown many years ago [158], Brzustowicz and Bassett presented a theory which mechanistically explains the link between 22q11.2DS, miRNAs, and schizophrenia risk [156].

1.2.7. Online Mendelian Inheritance In Man (OMIM)

OMIM is a public database that catalogs human genes and genetic disorders. The information OMIM contain is full-text, referenced overviews with all known Mendelian disorders and over 12,000 genes and updates daily. OMIM focuses on the relationship between phenotype and genotype [159]. Table 2 presents OMIM retrieved information that shows location of the associated gene, gene/locus name and MIM number linked to that gene as of January 2013.

Table 2. Schizophrenia related genes from OMIM

Location	Gene / Locus	MIM number	Location	Gene / Locus	MIM number
1p36.2	SCZD12	608543	11p14.1	GPR48	606666
1p36.22	MTHFR	607093	11q14-q21	SCZD2	603342
1q32.1	CHI3L1	601525	12q24.11	DAO	124050
1q42.2	DISC1	605210	13q14.2	HTR2A	182135
1q42.2	DISC2	606271	13q32	SCZD7	603176
3p25.2	SYN2	600755	13q33.2	DAOA	607408
3q13.31	DRD3	126451	14q32.33	AKT1	164730
5q23-q35	SCZD1	181510	18p	SCZD8	603206
6p23	SCZD3	600511	22q11.21	COMT	116790
6p22.3	DTNBP1	607145	22q11.21	RTN4R	605566
6q13-q26	SCZD5	603175	22q12.3	APOL4	607254
8p21	SCZD6	603013	22q12.3	APOL2	607252
10q22.3	SCZD11	608078			

1.2.8. Molecular Summary of Schizophrenia

Table 3 summarizes the current molecular findings on schizophrenia. The regions marked with * were found to be associated with schizophrenia by Stefansson, H. *et al*. in 2008. However, The International Schizophrenia Consortium could not find corresponding association in 2008.

Table 3. Summary of current knowledge on molecular etiology of schizophrenia

Source	**Gene / Term**
Linkage Studies	COMT, PRODH2, ZDHHC8, NRG1, PPP3CC, DTNBP1, G72, DAAO, TRAR4, DISC1, RGS4, AKT1, DRD2, ERBB4, GARBB2, GRIN2B, HTR2A, IL1B, NOTCH4, NRXN1, PDE4B, PRODH, RELN
GWAS	PLXNA2, CSF2RA, IL3RA, TXNRD2, COMT, PI4KA, SERPIND1, ZNF804A, MYO18B, NOTCH4, FGFR2, HIST1H2BJ, PRSS16, PGBD1, NOTCH4, NRGN, TCF4, HIST1H2BJ, HIST1H2AG, HIST1H2BK, HIST1H4I, HIST1H2AH, PLAA, ACSM1, ANK3, MIR137, CACNA1C, ANK3, ITIH3-ITIH4region, 8p23.3-p21.2, 11p13.1-q14.1, 1q21.1, 15q11.2*, 15q13.3*, 22q11, 6p21.3-22.1
Other	• glutamate metabolism pathway • TGF-beta signaling pathway • TNFR1 pathway • glycan structures biosynthesis 1 • cell cycle • SNARE • cell adhesion molecules (CAMs) • tight junction pathways • substrate specific channel activity pathway • 22q11.2 deletion syndrome
OMIM	SCZD12, MTHFR, CHI3L1, DISC1, DISC2, SYN2, DRD3, SCZD1, SCZD3, DTNBP1, SCZD5, SCZD6, SCZD11, GPR48, SCZD2, DAO, HTR2A, SCZD7, DAOA, AKT1, SCZD8, COMT, RTN4R, APOL4, APOL2
Common Genes	NOTCH4, HTR2A, COMT, HIST1H2BJ, ANK3

CHAPTER II

2 MATERIALS AND METHODS

2.1. Data

2.1.1. dbGAP (database of Genotypes and Phenotype)

The dbGaP public repository for individual-level phenotype, exposure, genotype and sequence data and the associations between them has been instituted by the National Center for Biotechnology Information (NCBI). Each studies and subsets of information from those studies in the database have its' unique identifier which allow published studies to discuss or cite the primary data in a specific way. There are two access types for usage of dbGaP. Public access allows researchers to study documents linked to summary data on specific phenotype variables, statistical overviews of the genetic information, position of published associations on the genome, and authorized access is required to obtain individual-level data [42].

2.1.2. Cases and Controls

Cases were selected according to criteria for schizophrenia (SCZ) or schizoaffective disorder (SA) per the Diagnostic and Statistical Manual of Mental Disorders version IV [160]. More information on inclusion criteria can be found online at [161]. Individuals in control group were selected unless they endorsed a history of these disorders. The control subjects have been recruited by Knowledge Networks, a survey and marketing research company in U.S. Each control filled out a questionnaire comprising 69 screening questions (45 for disease related traits and 24 for personality traits). The questionnaire and details can be accessed from [162]. In total there were 3,029 individuals in the dataset comprising 1,385 subjects as control and 1,576 subjects as case [163].

NonGAIN GWAS dataset includes 908,477 SNPs and two different sets of data apart from controls and cases. These are General Research Use (GRU) comprising 2,677 subjects (1,385 controls and 1,292 cases) and Schizophrenia and Related Conditions (SARC) comprising 352 cases. The dataset divided into two groups because of the type of data access consent that participants gave. The people in GRU group agreed their data to be accessible for general research use purposes and the SARC group agreed their data to be accessible in research which are solely related to schizophrenia and related conditions. The SARC group was excluded from this study, only data from GRU group was used for analysis.

The Affymetrix Genome-Wide Human SNP Array 6.0, or Affy 6.0 as commonly known, covers 1.8 million genetic markers as more than 906,600 single nucleotide polymorphisms (SNPs) and more than 946,000 probes for copy number variation (CNV). The platform also offers analysis tools that can truly link copy number and association high-resolution reference map and a copy number polymorphism (CNP) calling

algorithm developed by the Broad Institute [164].

In this study we handled a dataset, which is a part of Molecular Genetics of Schizophrenia (MGS) GWA study, downloaded from dbGaP [42] with study accession number of phs000167.v1.p1. The MGS study incorporates 3,972 cases (2,686 European ancestry (EA) and 1,286 African American (AA)) and 3,629 controls (2,656 EA and 973 AA) (analyzed sample remaining after quality control exclusions). Half of the EA sample and almost the entire AA sample were genotyped under the preservation of the Genetic Association Information Network (GAIN) with the Affymetrix 6.0 platform. Remaining samples were also genotyped with the same platform and referred as nonGAIN sample.

2.2. Analysis

2.2.1. Pre-processing of Schizophrenia Data

After the dataset was downloaded, it was required to be adjusted for determination of significance values of each SNP. Although most of the datasets do not require any tuning for next step, schizophrenia set was missing the phenotype entries for all of the participants. Upon correction of missing values through ID matching of the subject with the phenotype value of the subject from related study and phenotype files, dataset became available for further steps of the analysis.

2.2.2. Assignment of Statistical Association

PLINK [165] v1.07 was used for calculation of significance values. It is designed as a free, open-source, command-line tool to handle a bunch of beneficial analysis such as whole genome association analysis and a range of basic, large-scale analyses in a computationally efficient manner. The sole focus of the toolset is on analysis of genotype/phenotype data. Thus any data which will be analyzed through PLINK, should be undergone proper formatting [166].

Prior to statistical association PLINK applies initial quality control. This procedure is performed thorugh base filtering and preprocessing with default features (Minor Allele Frequency = 0.05, SNP Missingness Rate = 0.1, Individual Missingness Rate = 0.1, Hardy Weinberg Equilibrium = 0.001). The software was executed for binary dataset (*.bed, *.bim and *.fam files) with the following parameters:
plink --bfile "file names" --assoc --adjust --out "output file"

Upon completion of PLINK analysis, it generates *.qassoc.assoc as an output file in addition to basic *.assoc file. Former file includes a pre-sorted list of association results with following columns: Chromosome; SNP identifier; Unadjusted, asymptotic significance value; Genomic control adjusted significance value; Bonferroni adjusted significance value; Holm step-down adjusted significance value; Sidak single-step adjusted significance value; Sidak step-down adjusted significance value; Benjamini & Hochberg (1995) step-up FDR control; Benjamini & Yekutieli (2001) step-up FDR control. Among these statistical results unadjusted p-value was used in latter analysis because rest of the statical tests have an assumption that states each test subject is indepent of each other.

2.2.3. Data Preparation for METU-SNP

SNPs in the result files of PLINK are designated with Affy Ids. However, for combined p-value analysis and SNP prioritization through METU-SNP, SNPs must be given indicated with rs IDs (a reference SNP ID number assigned to each unique submitted SNP records by dbSNP). Accordingly, all of the affy IDs were converted into rs IDs. Yet, further adjustment on the result files was required with respect to format for the next analysis. Each column is restricted within a spesific character limit. First column is required 4, second is required 12, and the rest is required 11 including white spaces. After output files were reformatted according to those limitations, results were appropriate to be piped for SNP prioritization.

2.2.4. METU-SNP Analysis

METU-SNP is an integrated software application which is specifically designed for use in SNP based genome wide association studies. The application is publicly accessible at http://metu.edu.tr/~yesim/metu-snp.htm. There are also video tutorials and help files available on the website. METU-SNP is written in Java as a desktop application to prioritize biomarkers and discover genes and pathways related to diseases via analysis of the GWAS case-control data. METU-SNP is a very beneficial tool for GWA studies since: (1) it allows the user to prioritize hundreds thousands of SNPs that are statistically found to be associated with the phenotype, (2) it helps identifying of the smallest set of SNPs (informative SNPs) that can be integrated as a biomarker panel of the phenotype for follow-up analysis. For the identification of an informative SNP subset it is required to sort the SNPs according to well-defined criteria. By doing so, biologically more relevant SNPs are going to be clustered as they preserve their statistical significances. In order to perform such a task, METU-SNP employs a scoring mechanism for each SNP

that would reflect SNP's biological and statistical relevance by integrating AHP and Gene Set Enrichment Analysis frameworks into evaluation of SNPs. Furthermore, the software is equipped with machine learning based feature selection schemes which enables users to decrease the number of SNPs in the output. It's database was built with data gathered from major public databases such as dbSNP, Entrez Gene [167], KEGG [168] and Gene Ontology [169].

2.2.5. Associating Genes and Pathways

Fisher's method [170], also known as Fisher's combined probability test, is the classic method to combine p-values from independent tests. Fisher's method combines p-values into one test statistic (χ^2) using the following formula [171]:

$$(\chi)^2 = -2\sum_{n=1}^{k} \log_e (p_n)$$

P_k is the p-value for the n^{th} hypothesis test. As the formula suggests the p-values tend to be large, when the test statistic χ^2 will be small. Thus, the null hypotheses are failed to reject for every test [171].

Fisher's method is being used in METU-SNP analysis in order to find statistically significant genes through associated SNPs. Whereas, Hypergeometric test (Fisher's exact test [172]) is being used to determine statistically significant pathways. In the following formula total number of given genes is N, total number of genes which are significantly associated with the disease is S and *m* is the number of genes in the pathway. That's said p-value of observing k-significant genes in the pathway is estimated by [173]:

$$p = \sum_{n=0}^{k} \frac{\binom{S}{n}\binom{N-S}{m-i}}{\binom{N}{m}}$$

METU-SNP has two phases to carry out these calculations. Using Fisher's combination and Fisher's exact tests, the software calculates combined p-values first for the discovery of significant genes and then the pathways, respectively. Parameters were selected as following: SNP threshold p-value as $5x10^{-4}$, gene and pathway threshold p-values as 0.05.

2.3. SNP Prioritization with Analytic Hierarchy Process (AHP)

Analytic Hierarchy Process (AHP) is used in decision-making based on multiple different criteria. It was developed by T. L. Saaty in 1970s [174].

Inputs that can be used in an AHP study varies from actual measurements such as weight, height, price to subjective opinions such as preference, satisfactory level, feelings etc. Since humans are not consistent, some levels of inconsistencies are allowed in AHP. Principal Eigen vectors and principal Eigen values are utilized to determine the ratio scales from former and consistency index from latter [175]. Furthermore, AHP is a special tool with its' being flexible that can be integrated with various techniques such as Linear Programming, Quality Function Deployment, Fuzzy Logic and so on. AHP empowers users to achieve their goals in a better way by integrating such different techniques [176].

Analytic Hierarchy Process provides a numeric scale for the measurement of quantitative and qualitative performances. This scale ranges from 1/9 for 'least valued than', to 1 for 'equal', and to 9 for 'absolutely more important than' covering the entire

spectrum of the comparison [176].

In our study we have used METU-SNP's AHP based prioritization approach [177] to determine the biologically more relevant SNPs among statistically significant SNPs associated with schizophrenia. AHP score was determined as 0.4 for prioritization step. SNPs with a lower p-value than $5x10^{-5}$ are selected for the AHP prioritization, and the cut-off AHP score of 0.4 is used for further selection of SNP subset after prioritization.

2.4. Analysis of Associated SNPs and Gene Lists

These analyses include the utilization of different publicly available tools such as RegulomeDB [178] to identify DNA features and regulatory elements in non-coding regions of the human genome and SNPnexus [179] to summarize association of variations in the dataset with phenotypes.

RegulomeDB is a public database annotates SNPs with known and predicted regulatory elements such as regions of DNAase hypersensitivity, binding sites of transcription factors, and promoter regions that have been biochemically characterized to regulation transcription. The database utilizes datasets from GEO, the ENCODE project, and published literature [178]. It accepts query data in following formats: dbSNP IDs, 0-based coordinates, BED files, VCF files, and GFF3 files (hg19). RegulomeDB only requires exclusion of SNPs located on mitochondrial chromosomes.

SNPnexus is a knowledgebase for functional annotation of novel and publicly available genetic variants. It was designed to simplify and assist in the selection of functionally relevant Single Nucleotide Polymorphisms (SNP) for large-scale genotyping studies of multifactorial disorders. SNPnexus allows queries using dbSNP identifiers or

chromosomal regions for annotating known variants. It is synchronized with UCSC human genome annotation database via regular updates. Furthermore, SNPnexus allows users to compute the following data: genomic mapping and additional annotations, gene/protein consequences, effect on protein function, hapmap population data, regulatory elements conservation, phenotype & disease association and structural variations. In this study SNPnexus is used for summarizing association of variations in the dataset with phenotypes. SNPnexus retrieves phenotype and disease association data from: GAD, The Genetic Association Database [40]; COSMIC, The Catalogue Of Somatic Mutations In Cancer [180] and GWAS Catalogue, The Catalogue of Published Genome-Wide Association Studies[39], [179].

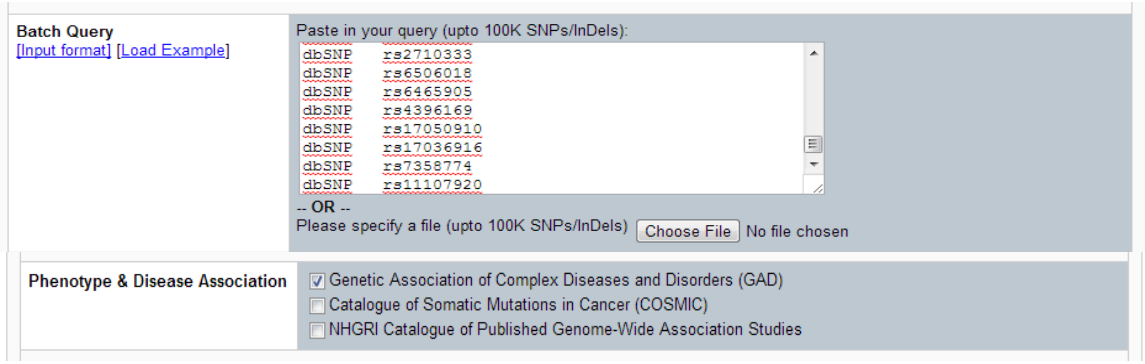

Figure 3. SNPnexus parameters for functional annotation

2.5. Biological Pathway Analysis

DAVID [181] and Reactome [182], which were publicly available bioinformatics tools, were employed for pathway analysis. This analysis helps us to understand and find out the biological context of the data by visualizing the overrepresented pathways in the dataset.

Database for Annotation, Visualization, and Integrated Discovery (DAVID), provides a

set of bioinformatics tools that present combination of functionally descriptive data and graphical displays. DAVID enhances the utility of resources which are focused on annotations and curations of gene-specific functional data such as Kyoto Encyclopedia of Genes and Genomes (KEGG) [168], BioCarta [183], PANTHER pathway [184] and much more by enabling batch query of genes [185].

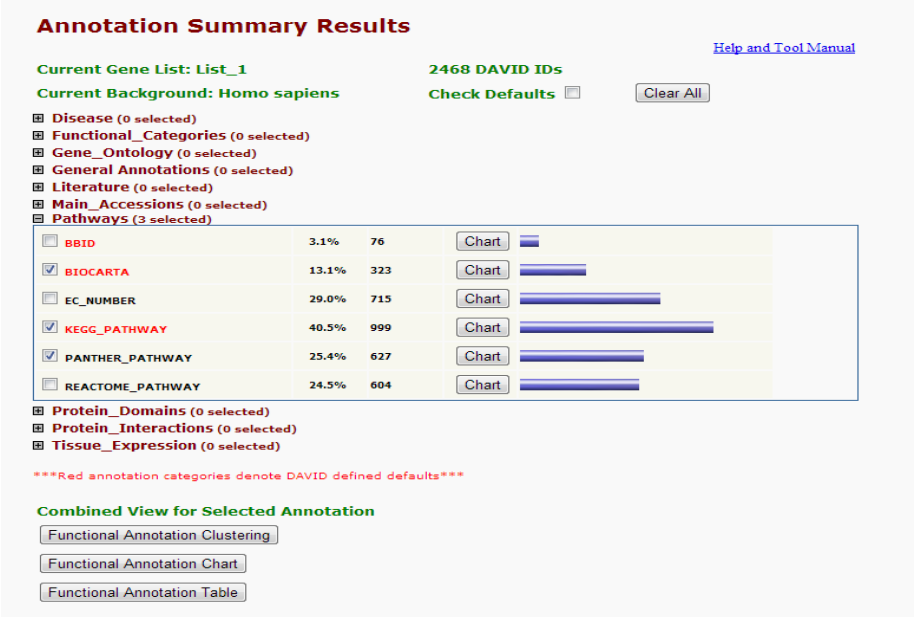

Figure 4. DAVID parameters for pathway analysis

After gene lists were uploaded as Entrez gene IDs to DAVID, functional annotation chart tool was used. All the default options were cleared and BIOCARTA, KEGG_PATHWAY and PANTHER_PATHWAY options were selected. There wasn't any other tuning than those that had been done for the analysis.

Reactome is an open source curated bioinformatics database which covers human pathways and reactions. It enables users to carry out ID mapping, pathway assignment and overrepresentation analysis via Pathway and Expression Analysis tools [182].

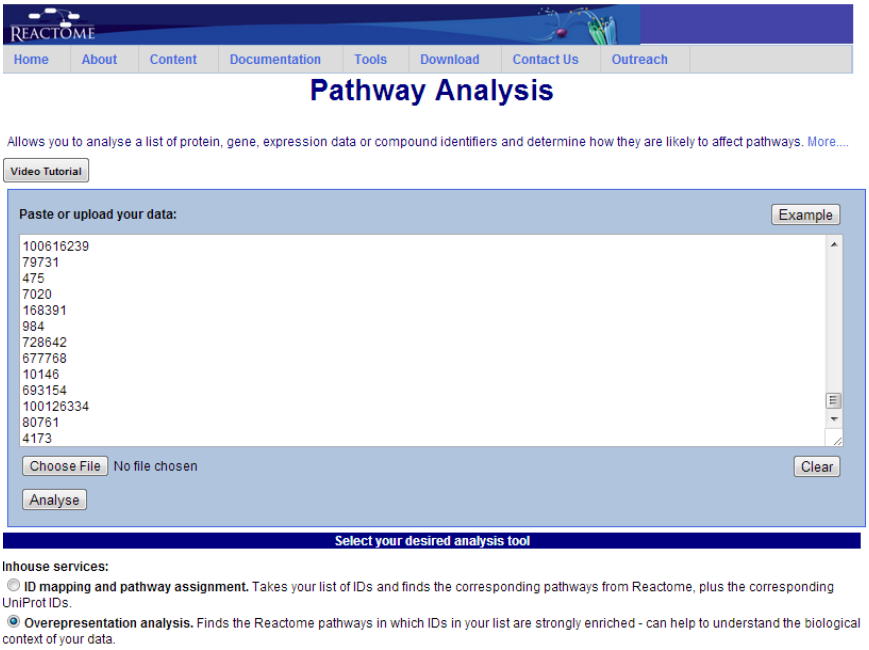

Figure 5. Reactome parameters for pathway analysis

To represent our data we uploaded lists of gene IDs (Entrez) to pathway analysis tool of Reactome and carry out overrepresentation analysis. This analysis helps users to understand the biological context of their data by finding enriched Reactome pathways with user list of genes.

2.6. Network Analysis

GeneMANIA is a publicly available gene network tool for predicting the function of query genes and gene sets. It finds other genes that are in a relationship with genes in interest by hosting a very large set of functional association data comprising protein and genetic interactions, pathways, co-expression, co-localization and protein domain similarity (currently, by Jan'13, indexing 1,464 association networks containing 292,680,904 interactions mapped to 149,747 genes from 7 organisms) [186].

GeneMANIA is allowed to integrate 10 related genes to network for gene discovery and 'Predicted' option is unchecked for network analysis. Other than these adjustments default options were used.

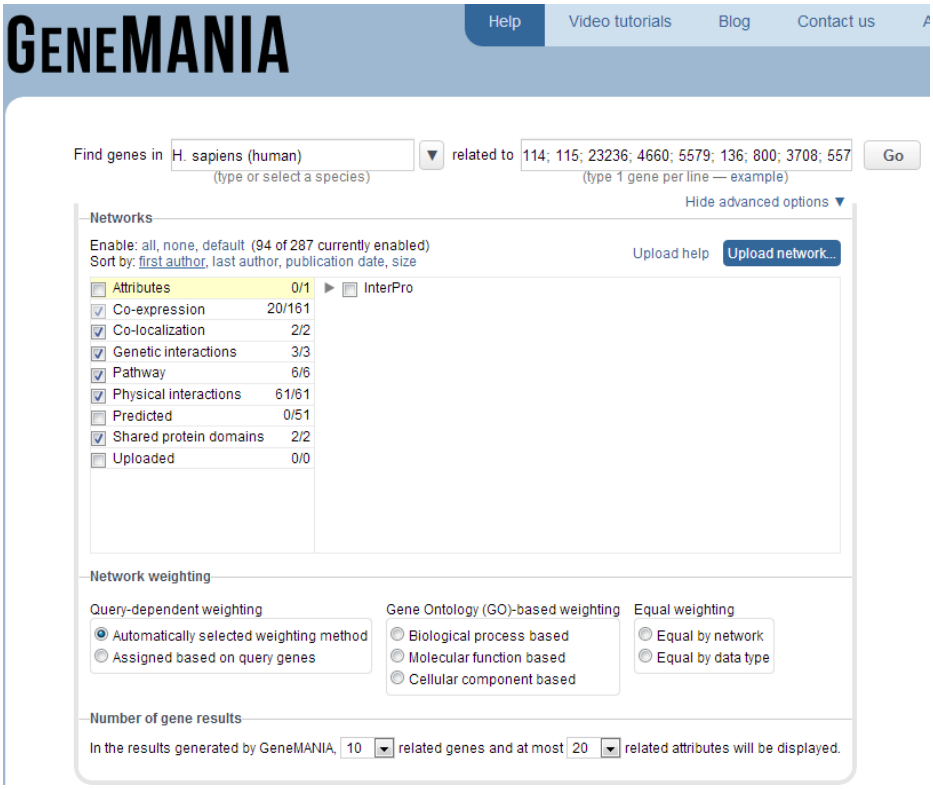

Figure 6. GeneMANIA parameters for network analysis

CHAPTER III

3 RESULTS AND DISCUSSIONS

3.1. PLINK Results and Analysis of Associated SNPs and Genes

P-values of 909,622 SNPs from 2,677 subjects were calculated through PLINK (Table 4). There were 25,555 SNPs with a p-value smaller than $5x10^{-5}$, and 4,375 of them were mapping to 5,365 Entrez gene IDs and 5,324 Ensembl [187] gene names according to Ensembl Biomart [188].

PLINK was solely used for calculation of p-values for SNPs. There is no consideration of biological significance of the SNPs or the genes that they are mapped in this analysis. None of the genes previously mentioned (1.4. Molecular epidemiology) were ranked among top 10 statistically significant SNPs (Table 4). Most of the SNPs were not mapped to any gene (7 out of 10), and three SNPs mapped to two genes. First one is PCDH11X from SNP rs2563850 and rs2562967. Second gene is PCDH11Y from SNPs rs2557227, rs2563850 and rs2562967. Both genes were associated with the GO terms calcium ion binding and homophilic cell adhesion. Although there are hits for the search of "PCDH11Y and schizophrenia", findings of the studies introduce definite proof of lack of association between PCDH11Y variants and schizophrenia as well as between PCDH11X and schizophrenia [189], [190]. And this is an example of a strong (10^{-19}) but an unvalid statistical result.

Table 4. GO Annotations of the genes which top 10 SNPs map

Chr.	p-value	SNP ID	Gene	GO Annotation (Function)
Ambiguous	4.16E-20	rs2574608	No gene model available	
Ambiguous	8.30E-20	rs2352301	No gene model available	
Ambiguous	8.78E-20	rs2771913	No gene model available	
X or Y	1.08E-19	rs2563850	PCDH11Y, PCDH11X	Calcium ion binding
Ambiguous	1.36E-19	rs2123561	No gene model available	
X or Y	1.43E-19	rs2562967	PCDH11Y, PCDH11X	Calcium ion binding
Y	1.43E-19	rs2558989	No gene model available	
X	1.57E-19	rs425231	No gene model available	
Y	1.67E-19	rs2557227	PCDH11Y	Calcium ion binding
X	2.01E-19	rs2755895	No gene model available	

Table 4 also includes location information of the given SNPs. As seen from the table locations are mostly ambiguous for given SNPs and the rest (4 SNPs) is either on Y chromosome or X chromosome. Although SNP locations were ambiguous according to ENSEMBL Biomart, it is highly probable that they reside on pseudoautosomal region as it is suggested in Manhattan plot (Figure 3) (Top 100 SNPs can be found at appendix A with associated p-values and full list is provided in the electronic supplements).

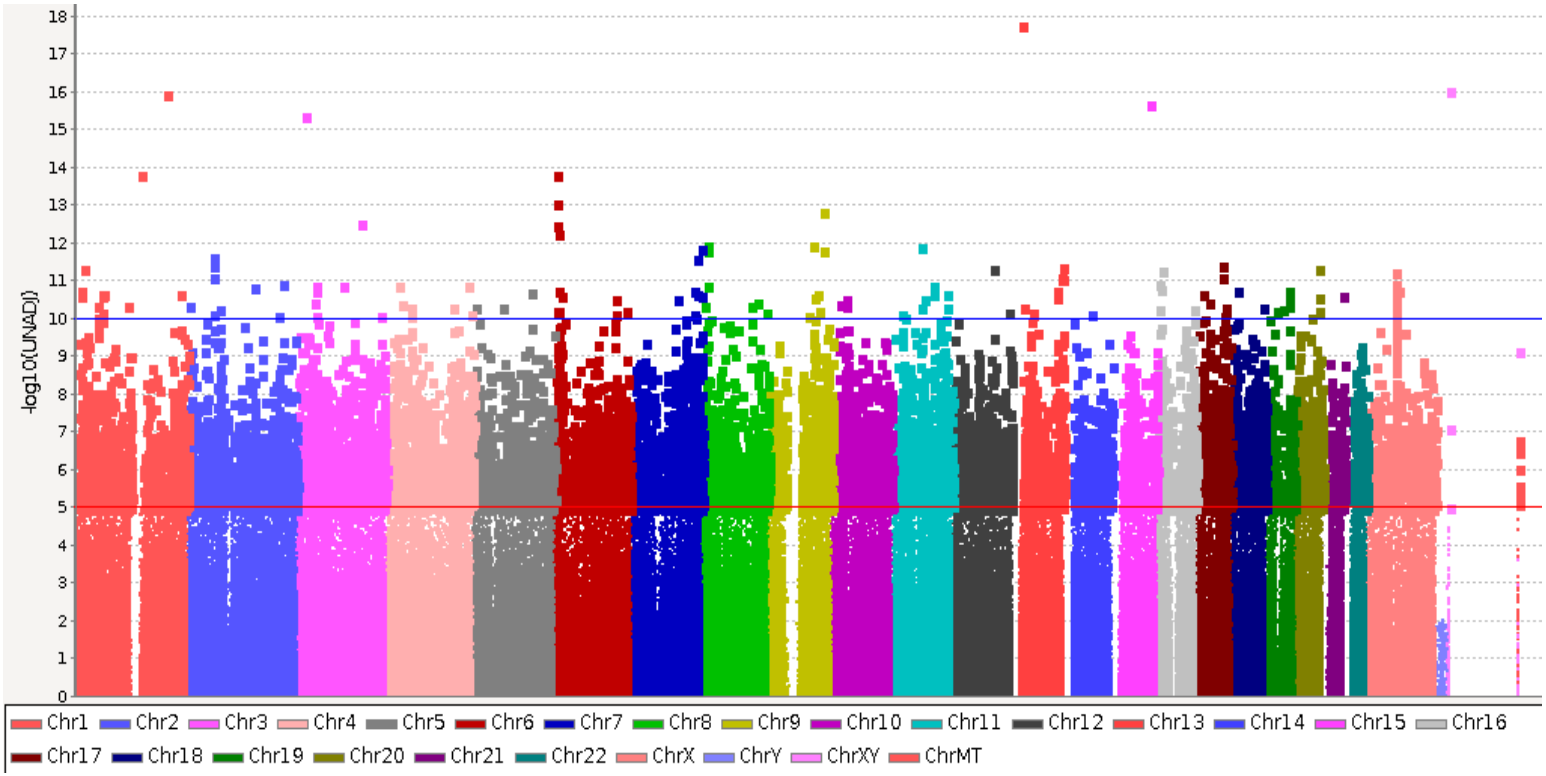

Figure 7. Manhattan Plot of PLINK results

Red line indicates 10^{-5} and there are 15,783 SNPs above it. Blue line indicates p-value of 10^{-10} and there are 201 SNPs above it. Chromosome 1 and chromosome X has the greater number of highly significant SNPs ($p < 10^{-10}$) accumulated around 10^{-11}. Chromosome 6 has 4 SNPs accumulated on higher significance level compared to other chromosomes. Chromosome 13 has the most significant SNP, following chromosome XY (pseudoautosomal region), chromosome 1, chromosome 15 and chromosome 3 with p-values smaller than 10^{-15}.

PLINK results are summarized in the Figure 3 with respect to p-values and chromosomes as a Manhattan Plot. The Y-axis shows $-\log_{10}$ of unadjusted p-values, on the X-axis SNPs are shown according to their chromosomal locations. Red line (lower line) indicates p-value of 10^{-5} and blue line (upper line) indicates p-value of 10^{-10}. Above red line there are total of 15,783 SNPs, which each one has a p-value smaller than or equals to 10^{-5} and above blue line there are 201 SNPs with a p-value smaller than or equals to 10^{-10}. Further investigation on the Manhattan plot indicates chromosome 1 and chromosome X has the greater number of highly significant SNPs. Chromosome 1 has 14 and chromosome X has 23 SNPs with their p-values mostly accumulated around 10^{-11}. Chromosome 6 has 4 SNPs which are accumulated on higher significance level compared to other chromosomes, and these are namely rs9502759, rs6916467, rs7768749, rs11967088. Chromosome 13 has the most significant SNP (rs2880301) on the Manhattan plot, following rs3869940 on chromosome XY (pseudoautosomal region), rs12741415 on chromosome 1, rs3883014 on chromosome 15 and rs17042395 on chromosome 3 with p-values smaller than 10^{-15}. The SNPs on Y chromosome show the least significant values among any other SNPs. However, there is a study reported a novel locus near the CSF2RA (colony stimulating factor, receptor 2 alpha) which is located in the pseudoautosomal region (PAR1) of the X and Y chromosomes (Xp22.32/Yp11.3)) gene [141].

First 10 SNPs that map to a gene were summarized in Table 5. The rank of the SNPs, p-value and the gene associated with the SNP and functional GO annotation of the gene were also denoted in the table. SNPs were ranked according to their p-values, where the most significantly associated ones are the top ones.

Table 5. First 10 SNPs mapping to a gene

RANK	SNP ID	p-value	Gene	GO Annotation (Function)
3	rs2563850	1.08E-19	PCDH11Y, PCDH11X	Calcium ion binding
5	rs2562967	1.43E-19	PCDH11Y, PCDH11X	Calcium ion binding
9	rs2557227	1.67E-19	PCDH11Y	Calcium ion binding
14	rs2759914	2.45E-19	PCDH11Y	Calcium ion binding
20	rs2880301	1.47E-18	TPTE2	• Ion channel activity • Phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase activity • Protein tyrosine phosphatase activity • Protein tyrosine / serine / threonine phosphatase activity
21	rs2557030	1.69E-18	PCDH11Y	Calcium ion binding
32	rs3883014	1.92E-16	PPP1R12B	Enzyme activator activity
36	rs11967088	1.37E-14	CDYL	Histone acetyltransferase activity
37	rs1778596	1.43E-14	PDE4DIP	• Enzyme binding • Protein binding
39	rs7768749	8.16E-14	CDYL	Histone acetyltransferase activity
44	rs7638929	2.73E-13	CLSTN2	Calcium ion binding
45	rs6916467	3.03E-13	CDYL	Histone acetyltransferase activity

There were 7 unique genes that are included in the list and these were PCDH11Y/X, TPTE2, PPP1R12B, CDYL, PDE4DIP and CLSTN2. Although there are still irrelevant genes within the list like PCDH11Y/X, TPTE2 and PPP1R12B, even adding one biological filter by selecting the coding SNPs to the statistical findings revealed that the genes that which implicates the association with schizophrenia, such as CDYL [191] and PDE4DIP [192].

DAVID allows query of maximum 3,000 genes, so first 3,000 gene IDs were sorted and piped into the analysis with last entity having p-value of 6.79×10^{-6}. Table 6 shows the top 15 results of pathway analysis of PLINK results through DAVID functional annotation chart (Top 100 terms can be found at appendix B and full list is provided in the electronic supplements).

Table 6. Pathways Overrepresented According to PLINK Results (DAVID analysis)

Category shows the original database where the terms orient, term column shows the enriched terms associated with the gene list, count column shows the number of genes involved in the term, % column shows the ratio of involved genes over number of query genes and p-value column shows how enriched the term is through modified fisher exact p-value.

Category	Term	Count	%	P-Value
KEGG	Glycosphingolipid biosynthesis	9	0.32502708	5.45E-04
KEGG	Nucleotide excision repair	16	0.57782593	0.0010827
KEGG	Homologous recombination	11	0.39725532	0.0049802
KEGG	Calcium signaling pathway	39	1.40845070	0.0088684
KEGG	Axon guidance	30	1.08342361	0.0120194
KEGG	Vascular smooth muscle contraction	26	0.93896713	0.0205175
KEGG	Other glycan degradation	7	0.25279884	0.0220071
BIOCARTA	Aspirin Blocks Signaling Pathway Involved in Platelet Activation	8	0.28891296	0.0220690
PANTHER	5HT2 type receptor mediated signaling pathway	18	0.65005417	0.0227012
KEGG	Cell adhesion molecules (CAMs)	29	1.04730949	0.0281569
KEGG	MAPK signaling pathway	52	1.877934272	0.0290108
BIOCARTA	Thrombin signaling and protease-activated receptors	7	0.252798844	0.0413848
KEGG	Mismatch repair	8	0.288912965	0.0432031
PANTHER	Histamine H1 receptor mediated signaling pathway	12	0.433369447	0.0446769
KEGG	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	18	0.650054171	0.0493845

DAVID analysis of PLINK results returned 15 enriched terms ($p < 0.05$) with very small percentages of genes mapping to pathways. Pathway analysis of PLINK results were found to be loosely related to each other and schizophrenia. Many of the pathways were essential biological pathways, such as ‘calcium signaling pathway’ [193], ‘cell adhesion molecules’ [194] and ‘MAPK signaling pathway’ [195], and some such as ‘glycoshingolipid biosynthesis’ term [196] and ‘axon guidance’ [197] were highly related with neurologic disorders. Another possible candidate term which was ‘other glycan degradation’ referring to a neurological function. Since it refers to a relevant function, it is most likely to be related with schizophrenia at first glance. Unfortunately,

there isn't any concrete case encountered.

A second pathway analysis website Reactome was also utilized. Top 15 results from Reactome overrepresentation analysis were presented in Table 7 (Top 100 events can be found at appendix C and full list is provided in the electronic supplements). "P-value" column is the unadjusted probability of seeing N or more genes in this event by chance, "number of genes column" is count of genes in the query which map to this event, "total number of genes" is the count of genes involved in this event and the last column shows "the name of the event" that is overrepresented.

Table 7. Events Overrepresented According to PLINK Results (Reactome analysis)

P-value	Number of genes	Total number of genes	Name of this Event
5.45E-10	8	13	FGFR4 binds to FGF
5.45E-10	8	13	FGFR4 ligand binding and activation
5.45E-10	8	13	Autocatalytic phosphorylation of FGFR4
7.75E-09	9	23	FGFR ligand binding and activation
1.21E-08	9	24	Phosphorylation of FRS2-alpha by activated FGFR
1.21E-08	9	24	Activated FGFR binds PLC-gamma
1.21E-08	9	24	PLC-gamma phosphorylation by FGFR
1.21E-08	9	24	Activated PLC gamma release by activated FGFR
1.21E-08	9	24	Phosphorylation of FRS2-beta by activated FGFR
1.21E-08	9	24	Activated FGFR binds FRS2alpha
1.21E-08	9	24	Activated FGFR binds FRS2beta
1.21E-08	9	24	Activated FGFR recruits SHC1
1.21E-08	9	24	SHC1 is phosphorylated
2.74E-08	9	26	Activated FGFR and FRS2 bind to SHP2
2.74E-08	9	26	SHP2 is phosphorylated by activated FGFR

There are total of 292 events that are significant ($p < 0.05$) according to Reactome results. These events are mostly populated with FGFR and its' variants. They were observed at least once in 73 events within 292 events (25% occurrence rate). Following PI3K was observed 22 times (7.5%). Although most of the analysis results based to

PLINK analysis showed only lose connections to schizophrenia, the Reactome results were almost directly related with the disease. FGFR2 is one of the casual genes of schizophrenia and some of the FGFR variants are known to play a role in psychological disorders [198]. PI3K pathway is also being referred in schizophrenia related studies [199], [200].

RegulomeDB, an analysis tool for identifying non-coding SNP s with regulative effects, is employed next for this study. PLINK results returned 4,732 SNP IDs and RegulomeDB scores of 179 of them were above 3 (1a is the highest score and 6 is the lowest score). So all these SNPs were selected and mapped for further analysis. 179 SNPs mapped to 222 gene IDs. Table 8 shows DAVID functional chart of these genes.

Table 8. Pathways overrepresented according to PLINK RegulomeDB results (DAVID analysis)

Category	**Term**	**Count**	**%**	**PValue**
PANTHER	Pentose phosphate pathway	3	1.442307692	0.014624855
KEGG	Starch and sucrose metabolism	3	1.442307692	0.065847064

Once again pure statistical results are shown to be inadequate to extrapolate. Pentose phosphate pathway and starch and sucrose metabolism are too way general to associate with any conclusion and establish a biological sense out of them in concordance with schizophrenia. For instance, a search regarding “pentose phosphate pathway” returns 29,400 results and "starch and sucrose metabolism" search returns 958 results.

There are total of 77,411 results in SNPnexus analysis. 3,331 of them present the keyword “schizo” (schizophrenia, schizoaffective) but when the negative associations were disregarded and total was reduced down to 2,053 SNPs from 3,331 results. Further, number of results referring to the keywords for disease class neurological, chemdependency and psych was 19,963 comprising 3,404 unique SNPs. These findings were discussed later on with AHP based prioritization findings.

3.2. METU-SNP Analysis

After calculating p-value associations with PLINK the METU-SNP software is utilized for combined p-value calculations of genes and pathways and prioritization of associated SNPs.

According to the number of SNP and their p-values, METU-SNP calculates a combined p-value for genes and biological pathways. There were total of 16,615 genes, on which associated SNPs with a p-value of smaller than 0.05, mapped in the list. 12,256 of them were significant ($p<0.05$) and 6,708 of them were highly significant ($p<0.00005$) including 2,517 genes with p-value = 0.0 (Top 100 genes can be found at appendix D and full list is provided in the electronic supplements). Since there are so many genes with a p-value of 0, it would not be wise to examine just the top significant ones one by one with respect to their GO annotations.

Table 9 shows the result of DAVID analysis results for top 15 terms (All of the terms can be found at appendix E).

Table 9. Pathways Overrepresented According to First 3000 Genes Based On Combined P-Value Calculations

Category	Term	Count	%	P-Value
KEGG	Axon guidance	49	1.668369084	3.93E-09
KEGG	Calcium signaling pathway	56	1.906707525	3.22E-07
KEGG	Long-term depression	29	0.987402111	9.62E-07
KEGG	Vascular smooth muscle contraction	39	1.327885598	2.49E-06
KEGG	Focal adhesion	59	2.008852571	2.88E-06
KEGG	Arrhythmogenic right ventricular cardiomyopathy	29	0.987402111	9.07E-06
KEGG	ECM-receptor interaction	30	1.02145046	2.67E-05
KEGG	Pathways in cancer	82	2.79196459	2.68E-05
PANTHER	Metabotropic glutamate receptor group III pathway	32	1.089547157	4.50E-05
KEGG	ABC transporters	19	0.646918624	7.52E-05

(*Table 9 cont.)*

KEGG	MAPK signaling pathway	65	2.213142663	4.70E-04
KEGG	Small cell lung cancer	27	0.919305414	5.10E-04
PANTHER	Integrin signaling pathway	65	2.213142663	5.49E-04
PANTHER	Histamine H1 receptor mediated signaling pathway	19	0.646918624	6.32E-04
KEGG	Long-term potentiation	23	0.783112019	6.94E-04

There were total of 61 more enriched terms ($p < 0.05$) overrepresented with respect to METU-SNP gene results and returned results were mostly related with neurological terms as seen from Table 5. P-values showed high enrichment and a lot smaller compared to previous DAVID results based on PLINK analysis. Furthermore, gene counts were larger in numbers as well as percentages. And there seems to be more relevant results both within terms and with schizophrenia. In addition to DAVID results of PLINK, terms related with schizophrenia include 'ECM-receptor interaction' [201], 'ABC transporters' [202] and 'long term potentiation' [203]. Other relevant neurological terms are 'long term depression' (possible relation too schizophrenia), 'metabotropic glutamate receptor group III pathway', and integrin signaling pathway (some of the integrins activate MAPK signaling pathway [204]).

Cluster analysis groups, terms having similar biological meaning since they share similar gene members. The overall enrichment score for the cluster based on the EASE scores of each term members. The higher the score is the more enriched the cluster is.

Cluster analysis for the enriched pathways with DAVID revealed total of 19 clusters in the DAVID cluster analysis, but two clusters, cluster 1 and cluster 3, from top 3 were included in the table 10, because these two clusters seemed to be the most related clusters with neurological terms and schizophrenia.

Table 10. DAVID Cluster Analysis of Combined P-Value Selected Gene Results

Cluster 1	***Enrichment Score: 2.865165214266108***		
Category	**Term**	**Count**	**%**
KEGG	Vascular smooth muscle contraction	39	1.3278855
KEGG	Gap junction	27	0.9193054
PANTHER	Endothelin signaling pathway	28	0.9533537
KEGG	GnRH signaling pathway	24	0.8171603
Cluster 3	***Enrichment Score: 2.2422371291829233***		
Category	**Term**	**Count**	**%**
PANTHER	Histamine H1 receptor mediated signaling pathway	19	0.6469186
PANTHER	Thyrotropin-releasing hormone receptor signaling pathway	24	0.8171603
PANTHER	5HT2 type receptor mediated signaling pathway	26	0.8852570
PANTHER	Oxytocin receptor mediated signaling pathway	23	0.7831120
PANTHER	Metabotropic glutamate receptor group I pathway	17	0.5788219
PANTHER	Muscarinic acetylcholine receptor 1 and 3 signaling pathway	22	0.7490636
PANTHER	Alpha adrenergic receptor signaling pathway	13	0.4426285
PANTHER	Angiotensin II-stimulated signaling through G proteins and beta-arrestin	18	0.6128702
PANTHER	Beta1 adrenergic receptor signaling pathway	16	0.5447735
PANTHER	Cortocotropin releasing factor receptor signaling pathway	12	0.4085801
PANTHER	Beta2 adrenergic receptor signaling pathway	15	0.5107252

Table 10 shows the related clusters of DAVID cluster analysis (cluster 1 and 3) of METU-SNP gene results. Cluster 1 comprises 118 genes which 63 of them are unique; cluster 3 comprises 214 genes which 52 of them are unique.

The first cluster seems to be including an unrelated term such as "vascular smooth muscle contraction" due to high number of genes contributing to these pathways it was ranked at the top. The entire term set in the latter cluster were related with neurology and brain (relations are determined via keyword search on related search engines).

GeneMANIA can be used to find out new relationships or genes among a given gene

sets. Additional analysis with GeneMANIA for the genes only comprising these two clusters revealed total interaction of the two clusters. A gene network is built on GeneMANIA with 63 genes of cluster 1 and integrating 52 genes from cluster 2 (duplicates were removed). There were 114 genes in the network including ten new genes suggested by GeneMANIA (Figure 4). Query genes are depicted as black nodes and discovered genes are depicted as gray nodes. Edges show different interactions among genes, purple for co-expression, light-blue for pathway, yellow for shared protein domains, red for physical interactions, darker blue for co-localization and green for genetic interactions. Node sizes are determined according to their weight in the network.

In order to inspect the ten genes found related in GeneMANIA first gene symbols of these genes were acquired. These are LRMP, GUCY1B3, ADCY1, ADCY4, EDN2, EDN1, ADCY7, ADCY6, ADCY10 and GRIN1. Then GO annotations and terminology related with these genes were investigated thoroughly. Except LRMP and GRIN1, the related genes from GeneMANIA were further discussed in the next section since there are genes in common.

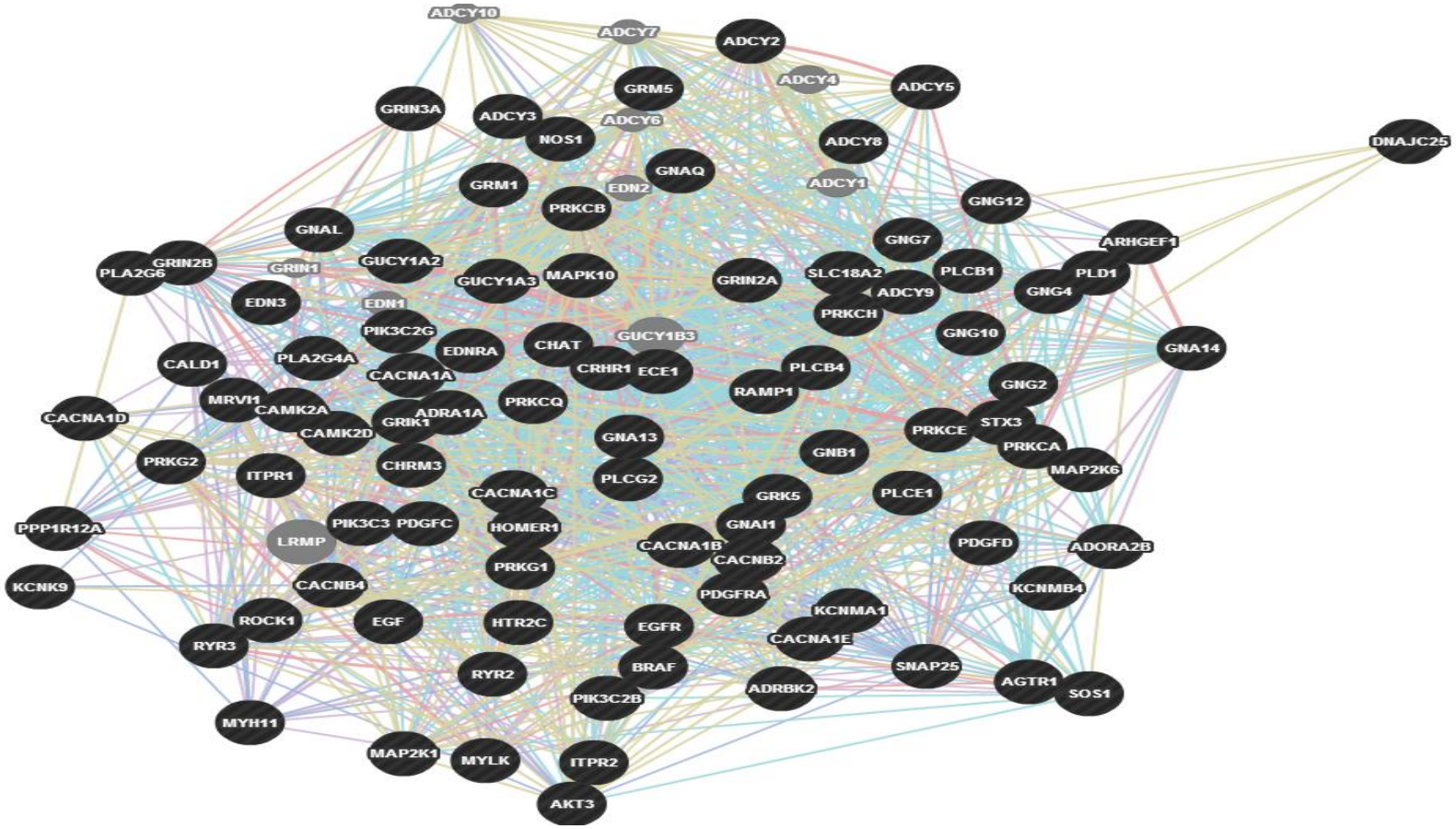

Figure 8. GeneMANIA gene network from related DAVID cluster of METU-SNP gene results

Query genes are depicted as black nodes and discovered genes are depicted as gray nodes. Edges show different interactions among genes, purple for co-expression, light-blue for pathway, yellow for shared protein domains, red for physical interactions, darker blue for co-localization and green for genetic interactions.

LRMP (lymphoid-restricted membrane protein) has GO annotation of vesicle fusion and targeting in process and endoplasmic reticulum membrane, integral to membrane and integral to plasma membrane in component with no terms retrieved from any of biological database or knowledgebase. As the findings imply LRMP is not likely to be a candidate gene that can be associated with schizophrenia.

GRIN1 (glutamate receptor, ionotropic, N-methyl D-aspartate 1) encodes a critical subunit protein of N-methyl-D-aspartate (NMDA) receptors, members of the glutamate receptor channel superfamily which are heteromeric protein complexes with multiple subunits arranged to form a ligand-gated ion channel. These subunits play a key role in the plasticity of synapses, which is believed to underlie memory and learning. It was showed that GRIN1 gene is associated with mental retardation phenotype [205]. Furthermore, there are lots of neurological terms and GO annotations associated with this gene such as Alzheimer's disease, calcium signaling pathways, long term potentiation, neuroactive ligand-receptor integration, neuronal system; neurotransmitter binding, synaptic transmission, regulation of excitatory polysynaptic membrane potential, etc. As it's shown the probability of GRIN1's being associated with schizophrenia is highly strong. Even, there are some studies showing possible association [206–209].

There is also a possibility of very distantly related genes present in the network. For instance DNAJC25 may be one of those genes as the network topology suggests. The features that DNAJC25 bears about the network such as the number of interactions, quantity of interaction types and distance of the node from network, are shown to be very weak in conjunction with general structure of the network. As GO function annotations (heat shock protein binding, unfolded protein binding) imply that DNAJC25 is the most distinctive gene among other genes that were discussed, other terms (neuronal system, neurotransmitter receptor binding) from REACTOME still validate

the relation of DNAJC25 with the network.

Top 15 results from Reactome overrepresentation analysis were presented in Table 11 (Top 100 events can be found at appendix F and full list is provided in the electronic supplements).

Table 11. Events Overrepresented Based on Combined P-Value of Gene Results (Reactome analysis)

P-value	Number of genes	Total number of genes	Name of this Event
3.16E-10	11	23	GRB2-SOS1 is recruited by activated FGFR mutants
1.62E-09	11	26	GRB2:SOS1 is recruited to activated FGFR through p-SHC1
1.62E-09	11	26	Ras nucleotide exchange mediated by GRB2-SOS1 bound to FGFR mutants
2.21E-09	10	21	Activated FGFR mutants bind FRS2alpha
2.21E-09	10	21	FGFR mutants phosphorylate FRS2alpha
2.21E-09	10	21	Activated FGFR mutants bind PLC-gamma
2.21E-09	10	21	PLC-gamma phosphorylation by FGFR mutants
2.21E-09	10	21	Activated PLC-gamma release by activated FGFR mutants
2.64E-09	11	27	GRB2:SOS1 binds to p-FRS2:activated FGFR
4.38E-09	9	17	FGFR2 ligand binding and activation
4.38E-09	9	17	Signaling by FGFR2 mutants
4.38E-09	9	17	Activated point mutants of FGFR2
4.38E-09	9	17	Point mutants of FGFR2 bind and are inactivated by tyrosine kinase inhibitors
6.52E-09	11	29	SHC-mediated cascade
6.52E-09	11	29	Ras nucleotide exchange by GRB2:SOS1 through p-SHC

Total of 341 events were found to be significant ($p < 0.05$) in Reactome analysis. The events were again mostly populated with FGFR (fibroblast growth factor receptor) and its' variants (FGFR2 occurrence increased to 15 from 10 compared to PLINK Reactome analysis). They occurred at least once in 77 events within 341 events (22.5% occurrence rate). Following SOS variants occurred 23 times (6.74%), GRB2 occurred 22 times (6.45%) and PI3K occurred 20 times (%5.86). FGFR and PI3K were already discussed

previously. GRB2 was identified and validated as a risk gene for schizophrenia in 2011 [210]. However, there isn't any verification found for SOS1. In the Reactome results SOS1 was present together with GRB2, suggesting a relation between SOS1 (or SOS1:GRB2 complex) and schizophrenia.

3.3. SNP Prioritization with METU-SNP and Analysis of Gene Lists

In the next stage the statistically associated SNPs were prioritized through METU-SNP using an algorithm based on analytical hierarchy process (AHP) integrated into METU-SNP. This procedure ranks the SNPs according to their biological relevance while preserving their statistical significance. Overall goal is to combine strength of statistics and biological importance of the SNPs with respect to the content of the study.

25,000 SNPs with a p-value smaller than $5x10^{-5}$ were prioritized using METU-SNP and AHP scores were calculated for each of them. Their scores were ranging from 0.033 to 0.7375. To further reduce the number of SNPs a threshold cut-off AHP value was determined as 0.4, which is the suggested threshold in previous studies [211]. This selection resulted in 6,202 SNPs and 2,587 unique gene IDs. Table 12 shows the top 10 prioritized genes with their associated gene names and p-values (Top 100 SNPs and their associated p-values can be found at appendix G and full list is provided in the electronic supplements).

Table 12. Top 10 Coding SNPs After AHP Prioritization

CHR	SNP ID	P-Val	AHP Score	Associated Genes
11	rs17115004	1.42E-06	0.737535	NCAM1
11	rs2229163	4.14E-06	0.73671	CHRM4
8	rs7009117	5.67E-07	0.71945	PCM1
11	rs6589360	6.64E-08	0.715562	NCAM1

(Table 12 cont.)

11	rs7128875	1.38E-06	0.715562	SLC6A5
7	rs7809317	1.77E-06	0.715562	CNTNAP2
9	rs6475523	9.05E-06	0.715562	SMARCA2
7	rs2906288	4.26E-05	0.715562	CNTNAP2
14	rs1012023	3.73E-07	0.714776	NPAS3
1	rs4846051	1.64E-05	0.714737	MTHFR

SNPs and the genes they are mapped were listed on Table 13 along with their GO annotations. Although none of the schizophrenia associated genes mentioned earlier made to the top of the list, there are some genes closely related with neurodevelopment and some that have specific neurological functions. The list is further enriched with neurological terms when component GO annotations are considered. Further investigations regarding following genes revealed neurological phenotypes such as Alzheimer's disease due to NCAM1 [212]; hyperekplexia due to SLC6A5 [213]; Alzheimer's disease [214], autism 15 [215], schizophrenia and bipolar disorder [216] due to CNTNAP2; Nicolaides-Baraitser syndrome due to SMARCA2 [217]; schizophrenia, bipolar disorder and depression due to NPAS3 [218], [219]; schizophrenia due to MTHFR [110].

Table 13. GO Annotations of selected genes after SNP prioritization

SNP ID	Genes	GO Annotation
rs17115004	NCAM1	axon guidance, cell adhesion, cytokine-mediated signaling pathway
rs2229163	CHRM4	G-protein coupled acetylcholine receptor signaling pathway, adenylate cyclase-inhibiting G-protein coupled acetylcholine receptor signaling pathway, cell proliferation, cell surface receptor signaling pathway
rs7009117	PCM1	centrosome organization, cilium assembly, interkinetic nuclear migration, microtubule anchoring, microtubule anchoring at centrosome
rs6589360	NCAM1	axon guidance, cell adhesion, cytokine-mediated signaling pathway
rs7128875	SLC6A5	glycine:sodium symporter activity, neurotransmitter:sodium symporter activity
rs7809317	CNTNAP2	enzyme binding, receptor binding, brain development, cell adhesion, cellular protein localization, cerebral cortex development
rs6475523	SMARCA2	helicase activity, protein binding, transcription coactivator activity, transcription regulatory region DNA binding

(Table 13 cont.)

rs2906288	CNTNAP2	enzyme binding, receptor binding, brain development, cell adhesion, cellular protein localization, cerebral cortex development
rs1012023	NPAS3	DNA binding, protein dimerization activity, signal transducer activity
rs4846051	MTHFR	methylenetetrahydrofolate reductase (NADPH) activity, modified amino acid binding

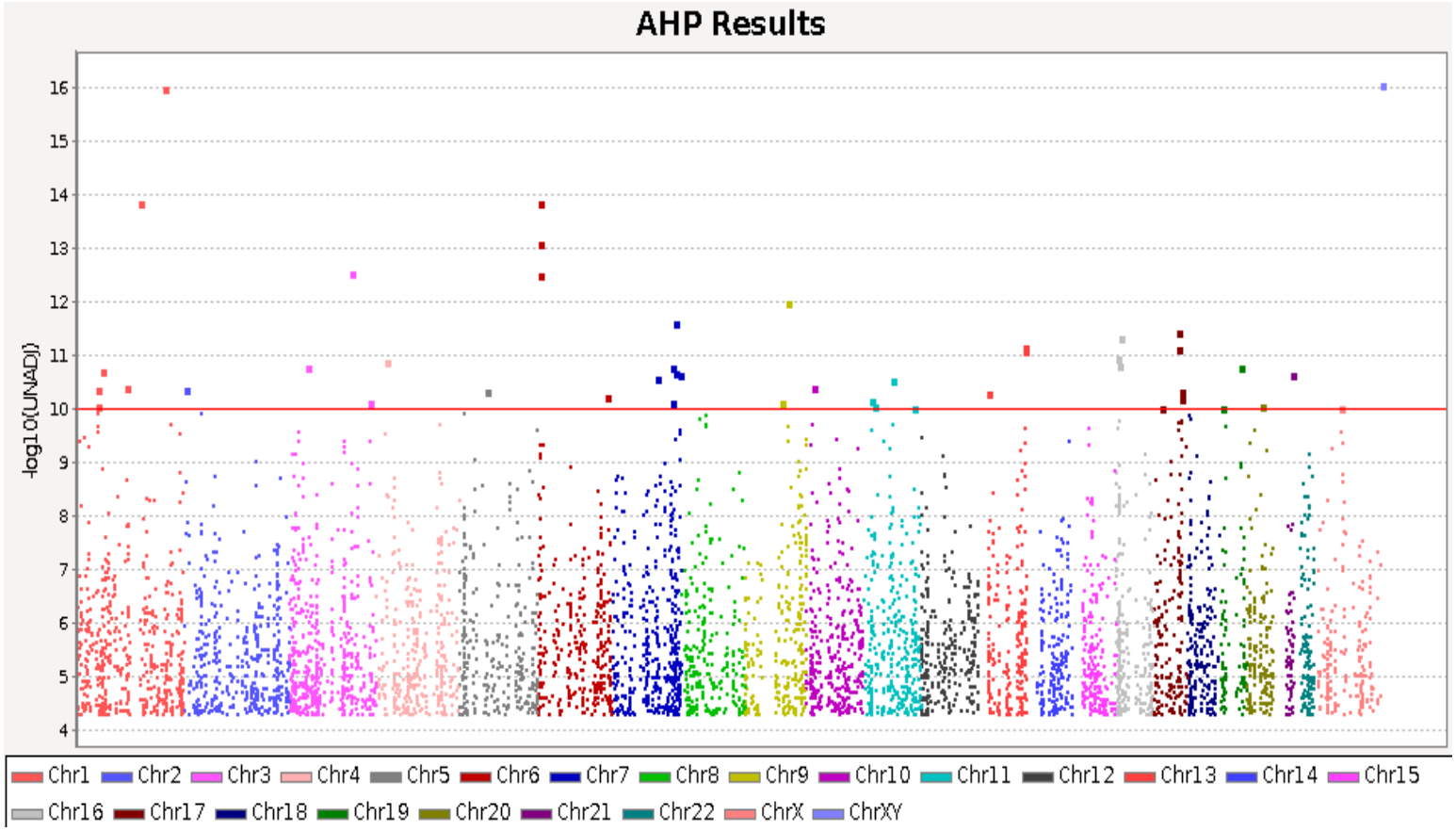

Figure 9. Manhattan plot of selected SNPs after AHP based prioritization with METU-SNP

AHP results are summarized in the figure 5 with respect to p-values and the chromosomes SNPs are located. Since the SNP list has maximum p-value of 0.00005 is used as cut off for the SNPs to be prioritized, Manhattan plot starts just before 10^{-5}. The Y-axis shows $-\log_{10}$ of unadjusted p-values, on the X-axis chromosomes are distributed in an ordered fashion. Red line indicates p-value of 10^{-10}. Above red line the number of SNPs reduces from 6,202 to 57. Further investigation on the Manhattan plot indicates chromosome 1 has the greater number of highly significant SNPs again. Chromosome 1 has 6 SNPs as well as chromosome 7. Chromosome 6 has 4 SNPs which are accumulated on higher significance level compared to other chromosomes, and these are namely rs7760946, rs6916467, rs7768749 and rs11967088. The latter 3 SNPs were also present in previous plot from PLINK results. All these SNPs map to CDYL gene which is encountered in submicroscopic 6p25 deletion. And this syndrome is characterized by intellectual disability, language impairment, hearing deficit, craniofacial, ophthalmologic, cardiac, and varying central nervous system anomalies [191]. Rs3869940 on chromosome XY (pseudoautosomal region) is the most significant SNP on the plot, following rs12741415 on chromosome 1 and the rest is less significant ($p<10^{-15}$). Although rs3869940 has the highest significance, there isn't any variation on the plot located on the pseudoautosomal region; rs3869940 is unique in that respect. Furthermore, it isn't being mapped to any gene. It might be a statistical anomaly as well as a significant marker.

The variations located on Y chromosome have less significance compared to other SNPs on Manhattan plot of PLINK. And as seen from current plot there isn't any variation located on Y chromosome made to the AHP list.

DAVID pathway analysis of AHP prioritized genes are shown on Table 14 (All of the terms can be found at appendix H). It was possible to run analysis with the entire gene IDs, since the number of entities (2,587) for AHP list was within DAVID's limits.

Table 14. Pathways Overrepresented According to AHP Results (DAVID analysis)

Category	Term	Count	%	PValue
KEGG	Axon guidance	52	2.1069692	9.94E-08
KEGG	Focal adhesion	69	2.7957860	9.22E-07
KEGG	Vascular smooth muscle contraction	44	1.7828200	2.56E-06
KEGG	Calcium signaling pathway	59	2.3905996	1.48E-05
KEGG	Long-term depression	29	1.1750405	4.52E-05
KEGG	Pathways in cancer	94	3.8087520	5.17E-05
PANTHER	Integrin signalling pathway	73	2.9578606	6.24E-05
KEGG	ABC transporters	21	0.8508914	8.48E-05
KEGG	MAPK signaling pathway	78	3.1604538	1.28E-04
BIOCARTA	Thrombin signaling and protease-activated receptors	12	0.4862236	2.04E-04
KEGG	Regulation of actin cytoskeleton	64	2.5931928	3.27E-04
BIOCARTA	PKC-catalyzed phosphorylation of inhibitory phosphoprotein of myosin phosphatase	10	0.4051863	3.88E-04
KEGG	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	28	1.1345218	8.00E-04
PANTHER	5HT2 type receptor mediated signaling pathway	27	1.0940032	1.47E-03
PANTHER	Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	49	1.9854132	1.54E-03

There were total of 59 significant terms ($p<0.05$) overrepresented with based on AHP results and were mostly related with neurological terms as seen from Table 14. Moreover, top 10 of these results are populated with terms specifically related with neurodevelopment: axon guidance, focal adhesion, integrin signaling pathway, ABC transporters, and MAPK signaling pathway. Especially associations between corpus callosum development and schizophrenia were shown in the previous studies [220–224], revealing the presence of structural differences in corpus collasum of schizophrenia patients. Hence, any pathway related to neurodevelopmental process may lead to similar structural differentiation. P-values presented higher significance and were a lot smaller compared to DAVID results of PLINK analysis. Yet, these results are almost similar to previous gene results (Table 9) based on combined p-values. In both enrichment analysis top 5 terms were same, but gene counts from AHP results were higher and percentages

(%) were also larger due to higher gene counts and less number of genes involved. Although p-values were close to each other, p-values of combined gene results showed a stronger enrichment. In addition to p-values, five different terms are involved among prioritized SNP set analysis, namely: Thrombin signaling and protease-activated receptors, regulation of actin cytoskeleton, PKC-catalyzed phosphorylation of inhibitory phosphoprotein of myosin phosphatase, 5HT2 type receptor mediated signaling pathway and heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway. Association of term 'regulation of actin cytoskeleton' with schizophrenia was predicted in previous studies as well [201], [225]. Thrombin signaling and protease-activated receptors and 5HT2 type receptor mediated signaling pathway were found to be associated with neural terminology [226–228].

Table 15 shows the first two clusters of DAVID cluster analysis of AHP results.

Table 15. DAVID Cluster Analysis of AHP Results

Cluster 1	***Enrichment Score: 2.322655045830916***		
Category	**Term**	**Count**	**%**
KEGG	Vascular smooth muscle contraction	44	1.7828200
PANTHER	Endothelin signaling pathway	32	1.2965964
KEGG	Gap junction	28	1.1345218
KEGG	GnRH signaling pathway	27	1.0940032
KEGG	Melanogenesis	24	0.9724473
Cluster 2	***Enrichment Score: 2.232236011179473***		
Category	**Term**	**Count**	**%**
PANTHER	5HT2 type receptor mediated signaling pathway	27	1.0940032
PANTHER	Oxytocin receptor mediated signaling pathway	24	0.9724473
PANTHER	Thyrotropin-releasing hormone receptor signaling pathway	24	0.9724473
PANTHER	Alpha adrenergic receptor signaling pathway	14	0.5672609
PANTHER	Histamine H1 receptor mediated signaling pathway	18	0.7293354
PANTHER	Angiotensin II-stimulated signaling through G proteins and beta-arrestin	20	0.8103727
PANTHER	Metabotropic glutamate receptor group I pathway	16	0.6482982
PANTHER	Muscarinic acetylcholine receptor 1 and 3 signaling pathway	21	0.8508914

There were total of 23 clusters in the DAVID cluster analysis, but first only two clusters are included in the table 15, as they were found to be the most related clusters with neurological terms and schizophrenia. Cluster 1 comprises 155 genes which 77 of them are unique; cluster 2 comprises 176 genes which 54 of them are unique. Although the entire term set in the second cluster related with neurology and brain (relations are determined via keyword search on related search engines), first cluster is seemed to be including unrelated terms such as "vascular smooth muscle contraction" and "melanogenesis". Yet, a deeper investigation proved otherwise. Building a gene network via GeneMANIA with 77 genes of cluster 1 revealed a network with full coverage. In other words, each one of these genes are interacting with one or more genes within the cluster. Moreover, integrating genes from cluster 2 with genes cluster one, a set of genes comprising 108 entities (there is a single gene ID that is not recognized by GeneMANIA) is formed. An analysis of this set brought out a second gene network not different from the first one. Second gene network comprising the entire genes from both of the clusters revealed full interaction, which means related and interacting genes were clustered after DAVID analysis (Figure 6).

GeneMANIA was also employed to present the gene interaction network and to search for new candidate genes. 10 new genes were returned by GeneMANIA analysis as they were linked to to the query genes in both of the networks. These genes were ADCY1, ADCY10, ADCY4, ADCY6, ADCY7, ARGLU1, EDN1, EDN2, GUCY1B3 and PLA2G12B and these ten genes were ranked in top 20 genes among 118 genes with PLA2G12B ranking first by GeneMANIA. Among these genes Brennand *et al.* previously implied association of ADCY7 with schizophrenia. In addition ADCY8 was mentioned along with ADCY7 as possible candidate [192]. Although ADCY8 was not one of the discovered genes, it was included in the network as a query gene. Besides that, 8 of these genes are in common with the results from METU-SNP genes. Since DAVID results were so similar, such an overlap would be expected. Table 16 lists these genes

with their rankings (GeneMANIA ranked the genes according to their functions) annotations of biological importance and relevance over NCBI. Among those genes GUCY1B3 is the node with most interactions, following ADCY7 and EDN1. In addition to that, these 3 genes are the only ones that have physically interactions with the network. The overall topology of GUCY1B3 (long term depression term from KEGG) and its interactions suggest that it has the highest potential to be associated with schizophrenia among other genes revealed by GeneMANIA.

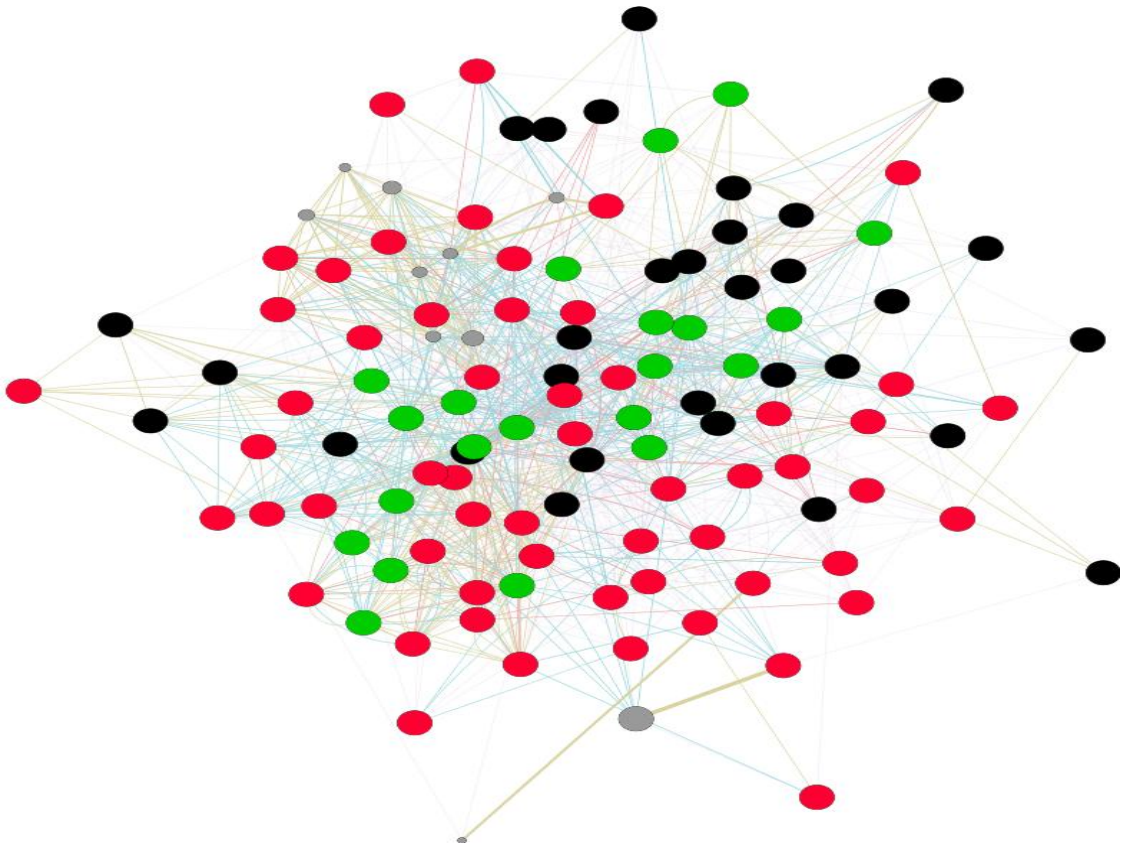

Figure 10. Merged Gene Network of DAVID of Prioritized SNPs

Black nodes are unique to second cluster (31 nodes); red ones are unique to first cluster (64 nodes), green one are common nodes (21 nodes) and grey nodes are added by GeneMANIA plugin of Cytoscape [224]. Gene names are missing due to requirement of specific zooming in which it is not possible to show whole network.

Table 16. Discovered Genes from GeneMANIA with Their Importance and Relevance

Rank	Gene Name	Biological importance & relevance
1	PLA2G12B	-
4	ADCY4	• GO Terms (process): nerve growth factor receptor signaling pathway • GO Terms (component): dendrite • Neuronal system (From REACTOME)
6	ADCY10	• GO Terms (component): axon, dendrite, neuronal cell body
8	ADCY1	• Located in brain • A similar protein in mouse is involved in pattern formation of the brain.
10	ADCY7	• GO Terms (process): nerve growth factor receptor signaling pathway, synaptic • transmission
12	ADCY6	• Expressed in brain tissue • Neuronal system (From REACTOME)
14	ARGLU1	-
18	GUCY1B3	• Long term depression (From KEGG)
19	EDN2	-
20	EDN1	• Endothelin 1 can affect the central nervous system.

An interesting matter to note that is PLA2G12B's placement (according to GeneMANIA ranking) in the network without any annotation such as the LRMP gene (with no terminology) from GeneMANIA network of METU-SNP genes. One explanation for their high ranking by GeneMANIA may be their interactions with other highly ranked genes. Even, PLA2G12B and LRMP were placed on the network comprising genes in the related clusters from both METU-SNP genes and AHP results (Figure 7). They were listed among top 7 genes in a network of 138 genes, PLA2G12B ranking second. GUCY1B3 was also involved in the list as 25^{th} gene under rest of the previously discovered genes except EDN1 and ARGLU1 (EDN1 came 26^{th} and ARGLU1 could not make to the list). It is important to note the orientations of GUCY1B3, PLA2G12B and LRMP. All of them are filling a gap in the network. Especially, GUCY1B3 is located almost in the middle of the network (Figure 7).

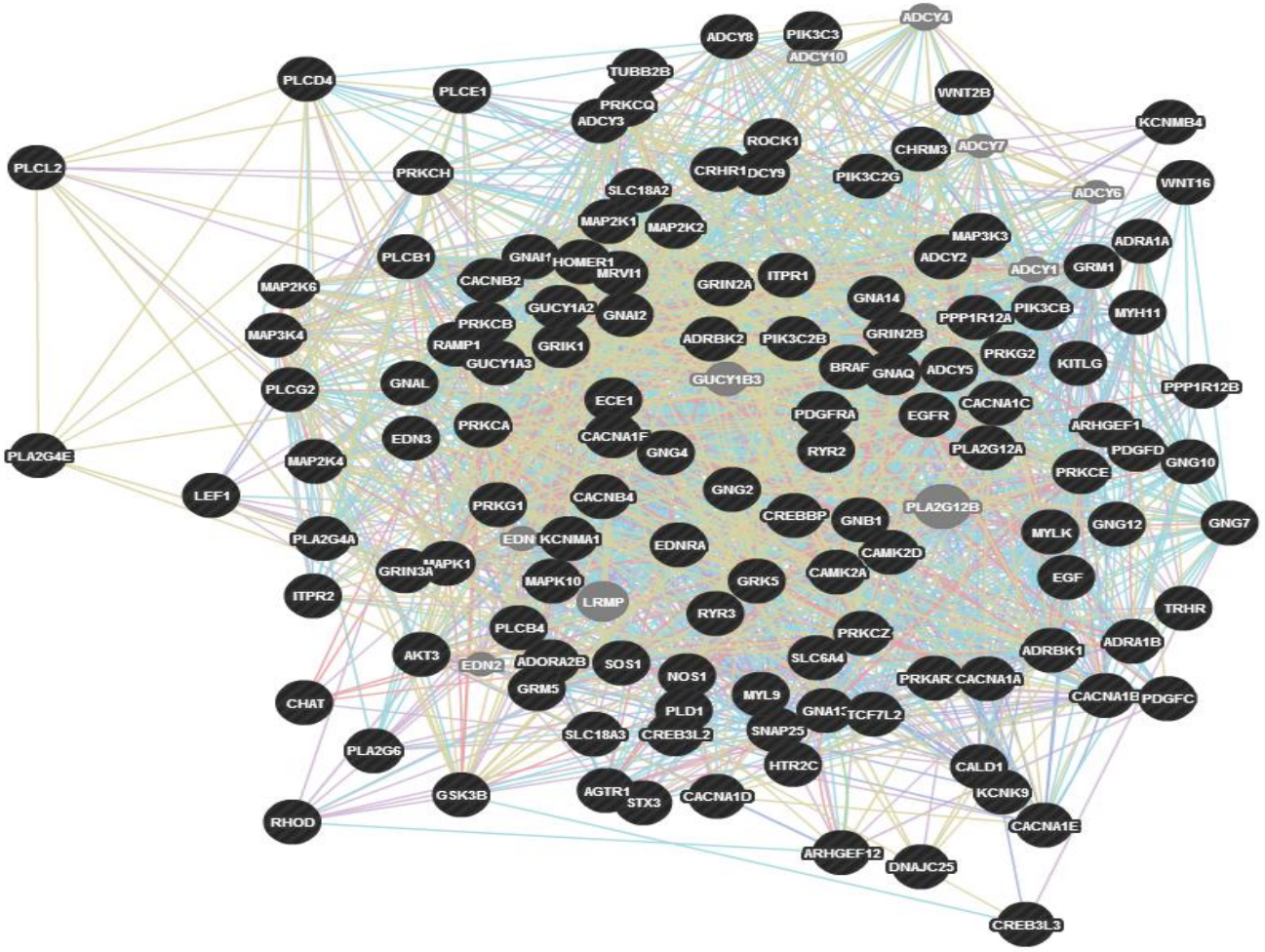

Figure 11. Combined GeneMANIA network from related DAVID cluster of significant gene based on combined p-value and genes prioritized SNPs mapped

As it is observed the networks are placing GUCY1B3 forward with backing evidence from NCBI and term annotations from databases and knowledgebases. The same networks are also highlighting PLA2G12B and LRMP, which, by the way, are not backed up by other databases, for all of these genes.

Next, the SNPnexus analyses were carried out with all the prioritized SNPs (6,202 SNPs) to investigate their phenotypic relations. The list of prioritized SNPs returned 62,584 results, as expected it was less than thePLINK results. 1,714 of them related to the keyword "schizo" (schizophrenia, schizoaffective), and disregarding the negative associations and there were 747 SNPs within 1,714 results. Further, number of results referring to disease class neurological, chemdependency and psych is 15,523 comprising 4,201 unique SNPs. These findings support our previous observations suggesting the strength of AHP based SNP prioritization which takes both biological relevance and statistical significance into account.

Table 17 shows the first cluster of DAVID analysis that was carried out with the results of SNPnexus coming from prioritized SNPs. 4,201 SNPs were mapped to 1,322 genes via ENSEMBL BioMart and these genes were analyzed with DAVID. Compared to table 15, table 17 includes two more terms: long-term depression and potentiation. Other than these two terms AHP based prioritization was found to be quite successful bringing out the relevant terms (Long-term depression and potentiation terms were also found in DAVID chart of AHP results). Moreover, PLINK (statistical) results (1,792 SNPs were mapped to 1,975 genes), which were biologically refined through SNPnexus, provided the same cluster table with table 17 (only difference was the orders of Endothelin signaling pathway and GnRH signaling pathway). Presence of 'vascular smooth muscle contraction' and 'melanogenesis' in table 17 reinforces their possible relation with schizophrenia.

Table 17. DAVID cluster analysis of SNPnexus results of prioritized SNPs

Cluster 1	***Enrichment Score: 3.24197385617593***		
Category	**Term**	**Count**	**%**
KEGG	Vascular smooth muscle contraction	32	2.523659306
KEGG	Long-term depression	23	1.813880126
KEGG	Long-term potentiation	20	1.577287066
KEGG	Gap junction	21	1.65615142
PANTHER	Endothelin signaling pathway	23	1.813880126
KEGG	GnRH signaling pathway	19	1.498422713
KEGG	Melanogenesis	16	1.261829653

3.3.1. Analysis of Non-Coding Prioritized SNPs

All of the prioritized SNPs (25,000) were analyzed through RegulomeDB as PLINK results. Among the 6,192 non-coding SNPs returned there were 177 SNPs with a RegulomeDB score above 3. These SNPs were suggested to have regulatory effects on 192 genes. We have analyzed these genes with DAVID tool using default pathway options for this study. Table 18 shows the result of DAVID analysis of these genes.

Table 18. Pathways overrepresented according to AHP RegulomeDB results (DAVID analysis)

Category	**Term**	**Count**	**%**	**PValue**
KEGG	Chemokine signaling pathway	7	3.846153846	0.059653976
KEGG	Long-term potentiation	4	2.197802198	0.080680724
KEGG	Vascular smooth muscle contraction	5	2.747252747	0.085674033
KEGG	Renal cell carcinoma	4	2.197802198	0.086315476

Long-term potentiation and vascular smooth muscle contraction terms were listed in DAVID chart for prioritized non-coding SNPs' results. In addition to vascular smooth muscle contraction term is present in the first cluster of DAVID analysis for prioritized coding SNPs. The genes involved with vascular smooth muscle contraction were PLCB1,

ARHGEF1, ADORA2B, CACNA1C and PRKCH; and they were also related to neurological terms, process and/or phenotypes. Genes involved in long-term potentiation include RPS6KA2 and CREBBP in addition to PLCB1 and CACNA1C. Among all these genes three are related with a neurological phenotype, which suggests a role in molecular etiology of schizophrenia: PLCB1 with Epileptic encephalopathy, early infantile, 12 (EIEE12) [229]; CACNA1C with major depression, bipolar disorder, and schizophrenia [151], [219], [230–232] and CREBBP with Rubinstein-Taybi syndrome (since symptoms include mental retardation) [233]. Associations of these genes with the phenotypes show that CACNA1C is the most likely candidate gene that may be associated with schizophrenia. Although PLCB1 and CREBBP have been associated with neurological phenotypes, CACNA1C's relation with psychological conditions makes it a strong candidate. In order to specify more precisely regulatory elements located on CACNA1C gene were found to be more likely associated with schizophrenia.

All of 149 events are found to be significant ($p < 0.05$) in Reactome analysis. These events are mostly populated with FGFR and its' variants like previous analysis. They were occurred at least once in 37 events within 149 events (24.8% occurrence rate). Following PI3K occurred 20 times (%13).

Top 15 results from Reactome overrepresentation analysis according to AHP results were presented in Table 18 (Top 100 events can be found at appendix I and full list is provided in the electronic supplements).

Table 19. Reactome Analysis of genes of Prioritized Non-Coding SNPs

P-value	Number of genes	Total number of genes	Name of this Event
1.22E-04	2	2	D-glyceraldehyde 3-phosphate + orthophosphate + NAD+ <=> 1,3-bisphospho-D-glycerate + NADH + H+
1.22E-04	2	2	1,3-bisphospho-D-glycerate + NADH + H+ <=> D-glyceraldehyde 3-phosphate + Orthophosphate + NAD+
2.69E-04	10	238	Metabolism of carbohydrates
2.69E-04	3	12	Keratan sulfate degradation
3.63E-04	2	3	limit dextrin-glycogenin => ((1,6)-alpha-glucosyl)poly((1,4)-alpha-glucosyl) glycogenin
3.63E-04	2	3	((1,6)-alpha-glucosyl)poly((1,4)-alpha-glucosyl)glycogenin => poly{(1,4)-alpha-glucosyl} glycogenin + alpha-D-glucose
3.63E-04	2	3	poly((1,4)-alpha-D-glucosyl) glycogenin => glycogen-glycogenin
3.63E-04	2	3	Activated NOD oligomer recruites RIP2 (RICK)
3.99E-04	4	32	Keratan sulfate/keratin metabolism
4.39E-04	3	14	RIP2 is K63 polyubiquitinated
5.83E-04	5	62	Glucose metabolism
7.20E-04	2	4	GPVI binds Fyn and Lyn
7.20E-04	2	4	RIP2 binds NEMO
7.20E-04	2	4	CARD9 binds RIP2 (and NOD2)
1.77E-03	2	6	Binding of GPVI:Fc Epsilon R1 gamma receptor complex with collagen

CHAPTER IV

4 CONCLUSION AND FUTURE STUDIES

4.1. Conclusions

- GWA studies are very well designed for identification of novel features related with a disease as they provide higher resolution for any given linkage study.
- Data produced as a result of GWA study should be well handled via statistical tools to filter out insignificant ones then it should be biologically assessed through biological data and knowledgebases.
- Statistical results are not sufficient enough for association studies most of the time. Data size may require further decrease to handle and be feasible for further studies.
- METU-SNP was shown to be a very useful tool to further decrease the size of data. As the software preserves statistical significance of the data, it also enriches data biologically.
- Integrating GWAS studies with pathway and network analysis was valuable to discover new relations and validate the findings of the study.
- 25 of the 66 genes from molecular summary of schizophrenia were observed in combined p-value of genes and AHP based SNP prioritization results.

- The results of the non-GAIN data analysis was in parallel with the previous GAIN study [155]. DAVID cluster analysis of AHP results was also pointed three of the five pathways mentioned in the GAIN study: cell adhesion molecules (CAMs), cell cycle and tight junction.
- GRIN1 (NMDA receptor) was found to be most likely related with schizophrenia according to GeneMANIA results of METU-SNP genes.
- Association of FGFR (especially FGFR2) and its related genes which is previously associated with schizophrenia was also observed in this study.
- Previously nominated GRB2 gene as causal of schizophrenia was also observed in this study.
- CNTNAP2, NPAS3 and MTHFR genes, which were previously associated, were observed in this study.
- ADCY7, whose association was also implied in [192], observed among discovered genes in the gene network built by GeneMANIA.

Novel findings of our study include:

- SOS1 forming a complex with GRB2 was found to be possible candidate according to Reactome results of METU-SNP genes.
- Melanogenesis and vascular smooth muscle contraction pathways were found to be related with schizophrenia according to DAVID cluster analysis. Although melanogenesis pathway was one of the less enriched pathways, genes belong to this pathway showed full coverage in a gene network composed of genes from neurological pathways.
- GUCY1B3 was identified as a strong candidate gene as a result of GeneMANIA analysis. There were also two other genes PLA2G12B and LRMP ranked before GUCY1B3. Although there isn't any suggestive evidence that these genes may

be related with schizophrenia, they ranked at top levels within a neurological gene network.

- CACNA1C was identified as a casual gene of schizophrenia in previous studies, this study suggests that CACNA1C's association with schizophrenia is being originated from its' regulatory elements.
- Axon guidance, focal adhesion, integrin signaling pathway, ABC transporters, and MAPK signaling pathways shown to be associated with schizophrenia in this study, might have a role in corpus collasum reshaping observed in schizophrenia patients.
- Long-term potentiation and vascular smooth muscle contraction pathways also being present in DAVID analysis of RegulomeDB results suggest interaction between the networks of the coding and non-coding SNPs.
- The final combined network can be suggested as a key molecular basis for the etiology of schizophrenia.

4.2. Future Studies

Future studies are including but not limited to:

i. Meta analysis of the two datasets including current one and additional schizophrenia genotyping data, which is planned to be accessed through dbGAP, to validate SNP profiles associated with schizophrenia.
ii. Integration of findings from genotyping analysis with phenotypic features.
iii. Building a decision support model using genotype and phenotype features by means of data mining methods.
iv. A pilot study with genotype profiles and the decision support model for the external validation of the proposed model.

REFERENCES

[1] W. S. Klug and M. R. Cummings, *Concepts of Genetics (8th Edition)*. Prentice Hall, 2005, pp. 244–248.

[2] B. Lewin, *Genes IX*. Jones & Bartlett Publishers, 2007, p. 892.

[3] T. R. Gregory, J. A. Nicol, H. Tamm, B. Kullman, K. Kullman, I. J. Leitch, B. G. Murray, D. F. Kapraun, J. Greilhuber, and M. D. Bennett, "Eukaryotic genome size databases.," *Nucleic acids research*, vol. 35, no. Database issue, pp. D332–8, Jan. 2007.

[4] "About the Human Genome Project." [Online]. Available: http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml.

[5] A. J. Marian, "Molecular genetic studies of complex phenotypes.," *Translational research : the journal of laboratory and clinical medicine*, vol. 159, no. 2, pp. 64–79, Feb. 2012.

[6] Human Genome Sequencing ConsortiumInternational, "Finishing the euchromatic sequence of the human genome.," *Nature*, vol. 431, no. 7011, pp. 931–45, Oct. 2004.

[7] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx,

S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, and J. Szustakowki, "Initial sequencing and analysis of the human genome.," *Nature*, vol. 409, no. 6822, pp. 860–921, Feb. 2001.

[8] J. M. Kidd, G. M. Cooper, W. F. Donahue, H. S. Hayden, N. Sampas, T. Graves, N. Hansen, B. Teague, C. Alkan, F. Antonacci, E. Haugen, T. Zerr, N. A. Yamada, P. Tsang, T. L. Newman, E. Tüzün, Z. Cheng, H. M. Ebling, N. Tusneem, R. David, W. Gillett, K. A. Phelps, M. Weaver, D. Saranga, A. Brand, W. Tao, E. Gustafson, K. McKernan, L. Chen, M. Malig, J. D. Smith, J. M. Korn, S. A. McCarroll, D. A. Altshuler, D. A. Peiffer, M. Dorschner, J. Stamatoyannopoulos, D. Schwartz, D. A. Nickerson, J. C. Mullikin, R. K. Wilson, L. Bruhn, M. V Olson, R. Kaul, D. R. Smith, and E. E. Eichler, "Mapping and sequencing of structural variation from eight human genomes.," *Nature*, vol. 453, no. 7191, pp. 56–64, May 2008.

[9] C. E. G. Bruder, A. Piotrowski, A. A. C. J. Gijsbers, R. Andersson, S. Erickson, T. Diaz de Ståhl, U. Menzel, J. Sandgren, D. von Tell, A. Poplawski, M. Crowley, C. Crasto, E. C. Partridge, H. Tiwari, D. B. Allison, J. Komorowski, G.-J. B. van Ommen, D. I. Boomsma, N. L. Pedersen, J. T. den Dunnen, K. Wirdefeldt, and J. P. Dumanski, "Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles.," *American journal of human genetics*, vol. 82, no. 3, pp. 763–71, Mar. 2008.

[10] K. A. Frazer, S. S. Murray, N. J. Schork, and E. J. Topol, "Human genetic variation and its contribution to complex traits.," *Nature reviews. Genetics*, vol. 10, no. 4, pp. 241–51, May 2009.

[11] S. Levy, G. Sutton, P. C. Ng, L. Feuk, A. L. Halpern, B. P. Walenz, N. Axelrod, J. Huang, E. F. Kirkness, G. Denisov, Y. Lin, J. R. MacDonald, A. W. C. Pang, M. Shago, T. B. Stockwell, A. Tsiamouri, V. Bafna, V. Bansal, S. A. Kravitz, D. A. Busam, K. Y. Beeson, T. C. McIntosh, K. A. Remington, J. F. Abril, J. Gill, J. Borman, Y.-H. Rogers, M. E. Frazier, S. W. Scherer, R. L. Strausberg, and J. C. Venter, "The diploid genome sequence of an individual human.," *PLoS biology*, vol. 5, no. 10, p. e254, Sep. 2007.

[12] D. Graur and W.-H. Li, *Fundamentals of Molecular Evolution*. Sinauer Associates, 2000, p. 29.

[13] T. Strachan and A. Read, *Human Molecular Genetics, Fourth Edition*. Garland Science, 2010, pp. 467–495.

[14] A. D. Roses, "Apolipoprotein E alleles as risk factors in Alzheimer's disease.," *Annual review of medicine*, vol. 47, pp. 387–400, Jan. 1996.

[15] L. B. Jorde, W. S. Watkins, M. J. Bamshad, M. E. Dixon, C. E. Ricker, M. T. Seielstad, and M. A. Batzer, "The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data.," *American journal of human genetics*, vol. 66, no. 3, pp. 979–88, Mar. 2000.

[16] W. E. Evans and M. V Relling, "Pharmacogenomics: translating functional genomics into rational therapeutics.," *Science (New York, N.Y.)*, vol. 286, no. 5439, pp. 487–91, Oct. 1999.

[17] M. Singh, P. Singh, P. K. Juneja, S. Singh, and T. Kaur, "SNP-SNP interactions within APOE gene influence plasma lipids in postmenopausal osteoporosis.," *Rheumatology international*, vol. 31, no. 3, pp. 421–3, Mar. 2011.

[18] "The International HapMap Project.," *Nature*, vol. 426, no. 6968, pp. 789–96, Dec. 2003.

[19] A. Kitts and S. Sherry, "The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation," in *The NCBI Handbook*, 2002.

[20] "dbSNP Summary." [Online]. Available: http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi?view+summary=view+summary&build_id=137.

[21] R. Sachidanandam, D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein, G. Marth, S. Sherry, J. C. Mullikin, B. J. Mortimore, D. L. Willey, S. E. Hunt, C. G. Cole, P. C. Coggill, C. M. Rice, Z. Ning, J. Rogers, D. R. Bentley, P. Y. Kwok, E. R. Mardis, R. T. Yeh, B. Schultz, L. Cook, R. Davenport, M. Dante, L. Fulton, L. Hillier, R. H. Waterston, J. D. McPherson, B. Gilman, S. Schaffner, W. J. Van Etten, D. Reich, J. Higgins, M. J. Daly, B. Blumenstiel, J. Baldwin, N. Stange-Thomann, M. C. Zody, L. Linton, E. S. Lander, and D. Altshuler, "A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms.," *Nature*, vol. 409, no. 6822, pp. 928–33, Feb. 2001.

[22] T. A. Manolio, "Genome-Wide Association Studies and Clinical Applications," 2011.

[23] J. Hardy and A. Singleton, "Genomewide association studies and human disease.," *The New England journal of medicine*, vol. 360, no. 17, pp. 1759–68, Apr. 2009.

[24] T. A. Manolio, "Genomewide association studies and assessment of the risk of disease.," *The New England journal of medicine*, vol. 363, no. 2, pp. 166–76, Jul. 2010.

[25] M. McGue and T. J. Bouchard, "Genetic and environmental influences on human behavioral differences.," *Annual review of neuroscience*, vol. 21, no. Degler 1991, pp. 1–24, Jan. 1998.

[26] "Gregor Mendel and Single-Gene Disorders." [Online]. Available: http://www.nature.com/scitable/topicpage/mendelian-genetics-patterns-of-inheritance-and-single-966.

[27] J. F. Gusella, N. S. Wexler, P. M. Conneally, S. L. Naylor, M. A. Anderson, R. E. Tanzi, P. C. Watkins, K. Ottina, M. R. Wallace, A. Y. Sakaguchi, A. B. Young, I.

Shoulson, E. Bonilla, and J. B. Martin, “A polymorphic DNA marker genetically linked to Huntington’s disease,” *Nature*, vol. 306, no. 5940, pp. 234–238, Nov. 1983.

[28] G. D. Schellenberg, M. A. Pericak-Vance, E. M. Wijsman, D. K. Moore, P. C. Gaskell, L. A. Yamaoka, J. L. Bebout, L. Anderson, K. A. Welsh, and C. M. Clark, “Linkage analysis of familial Alzheimer disease, using chromosome 21 markers.,” *American journal of human genetics*, vol. 48, no. 3, pp. 563–83, Mar. 1991.

[29] N. Risch and K. Merikangas, “The future of genetic studies of complex human diseases.,” *Science (New York, N.Y.)*, vol. 273, no. 5281, pp. 1516–7, Sep. 1996.

[30] B. E. Stranger, E. A. Stahl, and T. Raj, “Progress and promise of genome-wide association studies for human complex trait genetics.,” *Genetics*, vol. 187, no. 2, pp. 367–83, Feb. 2011.

[31] R. J. Klein, C. Zeiss, E. Y. Chew, J.-Y. Tsai, R. S. Sackler, C. Haynes, A. K. Henning, J. P. SanGiovanni, S. M. Mane, S. T. Mayne, M. B. Bracken, F. L. Ferris, J. Ott, C. Barnstable, and J. Hoh, “Complement factor H polymorphism in age-related macular degeneration.,” *Science (New York, N.Y.)*, vol. 308, no. 5720, pp. 385–9, Apr. 2005.

[32] R. H. Duerr, K. D. Taylor, S. R. Brant, J. D. Rioux, M. S. Silverberg, M. J. Daly, A. H. Steinhart, C. Abraham, M. Regueiro, A. Griffiths, T. Dassopoulos, A. Bitton, H. Yang, S. Targan, L. W. Datta, E. O. Kistner, L. P. Schumm, A. T. Lee, P. K. Gregersen, M. M. Barmada, J. I. Rotter, D. L. Nicolae, and J. H. Cho, “A genome-wide association study identifies IL23R as an inflammatory bowel disease gene.,” *Science (New York, N.Y.)*, vol. 314, no. 5804, pp. 1461–3, Dec. 2006.

[33] K. Ozaki and T. Tanaka, “Genome-wide association study to identify SNPs conferring risk of myocardial infarction and their functional analyses.,” *Cellular and molecular life sciences : CMLS*, vol. 62, no. 16, pp. 1804–13, Aug. 2005.

[34] K. Yamazaki, D. McGovern, J. Ragoussis, M. Paolucci, H. Butler, D. Jewell, L. Cardon, M. Takazoe, T. Tanaka, T. Ichimori, S. Saito, A. Sekine, A. Iida, A. Takahashi, T. Tsunoda, M. Lathrop, and Y. Nakamura, “Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn’s disease.,” *Human molecular genetics*, vol. 14, no. 22, pp. 3499–506, Nov. 2005.

[35] R. Sladek, G. Rocheleau, J. Rung, C. Dina, L. Shen, D. Serre, P. Boutin, D. Vincent, A. Belisle, S. Hadjadj, B. Balkau, B. Heude, G. Charpentier, T. J. Hudson, A. Montpetit, A. V Pshezhetsky, M. Prentki, B. I. Posner, D. J. Balding, D. Meyre, C. Polychronakos, and P. Froguel, "A genome-wide association study identifies novel risk loci for type 2 diabetes.," *Nature*, vol. 445, no. 7130, pp. 881–5, Feb. 2007.

[36] "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.," *Nature*, vol. 447, no. 7145, pp. 661–78, Jun. 2007.

[37] G. A. Thorisson, O. Lancaster, R. C. Free, R. K. Hastings, P. Sarmah, D. Dash, S. K. Brahmachari, and A. J. Brookes, "HGVbaseG2P: a central genetic association database.," *Nucleic acids research*, vol. 37, no. Database issue, pp. D797–802, Jan. 2009.

[38] M. J. Li, P. Wang, X. Liu, E. L. Lim, Z. Wang, M. Yeager, M. P. Wong, P. C. Sham, S. J. Chanock, and J. Wang, "GWASdb: a database for human genetic variants identified by genome-wide association studies.," *Nucleic acids research*, vol. 40, no. Database issue, pp. D1047–54, Jan. 2012.

[39] L. A. Hindorff, P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, and T. A. Manolio, "Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 23, pp. 9362–7, Jun. 2009.

[40] K. G. Becker, K. C. Barnes, T. J. Bright, and S. A. Wang, "The genetic association database.," *Nature genetics*, vol. 36, no. 5, pp. 431–2, May 2004.

[41] T. E. Klein and R. B. Altman, "PharmGKB: the pharmacogenetics and pharmacogenomics knowledge base.," *The pharmacogenomics journal*, vol. 4, no. 1, p. 1, Jan. 2004.

[42] M. D. Mailman, M. Feolo, Y. Jin, M. Kimura, K. Tryka, R. Bagoutdinov, L. Hao, A. Kiang, J. Paschall, L. Phan, N. Popova, S. Pretel, L. Ziyabari, M. Lee, Y. Shao, Z. Y. Wang, K. Sirotkin, M. Ward, M. Kholodov, K. Zbicz, J. Beck, M. Kimelman, S. Shevelev, D. Preuss, E. Yaschenko, A. Graeff, J. Ostell, and S. T. Sherry, "The NCBI dbGaP database of genotypes and phenotypes.," *Nature genetics*, vol. 39, no. 10, pp. 1181–6, Oct. 2007.

[43] A. T. Beck, N. A. Rector, N. Stolar, and P. Grant, *Schizophrenia: Cognitive Theory, Research, and Therapy*. New York: The Guilford Press, 2009.

[44] L. Ciompi, "The natural history of schizophrenia in the long term," *The British Journal of Psychiatry*, vol. 136, no. 5, pp. 413–420, May 1980.

[45] C. M. Harding, "Course types in schizophrenia: an analysis of European and American studies.," *Schizophrenia bulletin*, vol. 14, no. 4, pp. 633–43, Jan. 1988.

[46] J. Shi, D. F. Levinson, J. Duan, A. R. Sanders, Y. Zheng, F. Dudbridge, P. A. Holmans, A. S. Whittemore, B. J. Mowry, F. Amin, C. R. Cloninger, J. M. Silverman, G. Nancy, W. F. Byerley, D. W. Black, R. R. Crowe, and R. Jorge, "Common variants on chromosome 6p22.1 are associated with schizophrenia," vol. 460, no. 7256, pp. 753–757, 2010.

[47] E. Shur, "The epidemiology of schizophrenia.," *British journal of hospital medicine*, vol. 40, no. 1, pp. 38–40, 42–5, Jul. 1988.

[48] P. B. Mortensen, C. B. Pedersen, T. Westergaard, J. Wohlfahrt, H. Ewald, O. Mors, P. K. Andersen, and M. Melbye, "Effects of family history and place and season of birth on the risk of schizophrenia.," *The New England journal of medicine*, vol. 340, no. 8, pp. 603–8, Feb. 1999.

[49] J. Boydell and R. M. Murray, *The Epidemiology of Schizophrenia*. Cambridge, UK: Cambridge University Press, 2003, pp. 49–67.

[50] M. Bresnahan, M. D. Begg, A. Brown, C. Schaefer, N. Sohler, B. Insel, L. Vella, and E. Susser, "Race and risk of schizophrenia in a US birth cohort: another example of health disparity?," *International journal of epidemiology*, vol. 36, no. 4, pp. 751–8, Aug. 2007.

[51] B. A. Palmer, V. S. Pankratz, and J. M. Bostwick, "The lifetime risk of suicide in schizophrenia: a reexamination.," *Archives of general psychiatry*, vol. 62, no. 3, pp. 247–53, Mar. 2005.

[52] S. Brown, "Excess mortality of schizophrenia. A meta-analysis.," *The British journal of psychiatry : the journal of mentalscience*, vol. 171, pp. 502–8, Dec. 1997.

[53] P. J. McKenna, *Schizophrenia and Related Syndromes*. Routledge, 2007, pp. 39–43.

[54] E. Kraepelin, *Dementia praecox and paraphrenia*. Edinburgh: Livingstone, 1913.

[55] E. Bleuler, *Dementia praecox or the group of schizophrenias*. New York: International Universities Press, 1911.

[56] M. Cannon, A. Caspi, T. E. Moffitt, H. Harrington, A. Taylor, R. M. Murray, and R. Poulton, "Evidence for early-childhood, pan-developmental impairment specific to schizophreniform disorder: results from a longitudinal birth cohort.," *Archives of general psychiatry*, vol. 59, no. 5, pp. 449–56, May 2002.

[57] M. Cannon, R. Kendell, E. Susser, and P. Jones, *The Epidemiology of Schizophrenia*. Cambridge: Cambridge University Press, 2003, pp. 74–99.

[58] J. McGrath, S. Saha, J. Welham, O. El Saadi, C. MacCauley, and D. Chant, "A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology.," *BMC medicine*, vol. 2, no. 1, p. 13, Apr. 2004.

[59] M. Weiser, J. van Os, A. Reichenberg, J. Rabinowitz, D. Nahon, E. Kravitz, G. Lubin, M. Shmushkevitz, H. Y. Knobler, S. Noy, and M. Davidson, "Social and cognitive functioning, urbanicity and risk for schizophrenia.," *The British journal of psychiatry : the journal of mental science*, vol. 191, pp. 320–4, Oct. 2007.

[60] C. Henquet, R. Murray, D. Linszen, and J. van Os, "The environment and schizophrenia: the role of cannabis use.," *Schizophrenia bulletin*, vol. 31, no. 3, pp. 608–12, Jul. 2005.

[61] M. Hambrecht and H. Häfner, "Substance abuse and the onset of schizophrenia.," *Biological psychiatry*, vol. 40, no. 11, pp. 1155–63, Dec. 1996.

[62] R. A. Chambers, J. H. Krystal, and D. W. Self, "A neurobiological basis for substance abuse comorbidity in schizophrenia," *Biological Psychiatry*, vol. 50, no. 2, pp. 71–83, Jul. 2001.

[63] S. Arndt, G. Tyrrell, M. Flaum, and N. C. Andreasen, "Comorbidity of substance abuse and schizophrenia: the role of pre-morbid adjustment," *Psychological Medicine*, vol. 22, no. 02, pp. 379–388, May 1992.

[64] R. Tandon, M. S. Keshavan, and H. A. Nasrallah, "Schizophrenia, 'Just the Facts': what we know in 2008 part 1: overview.," *Schizophrenia research*, vol. 100, no. 1–3, pp. 4–19, Mar. 2008.

[65] B. Riley and K. S. Kendler, “Molecular genetic studies of schizophrenia.,” *European journal of human genetics : EJHG*, vol. 14, no. 6, pp. 669–80, Jun. 2006.

[66] A. G. Cardno and I. I. Gottesman, “Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics.,” *American journal of medical genetics*, vol. 97, no. 1, pp. 12–7, Jan. 2000.

[67] I. I. Gottesman, *Schizophrenia Genesis: The Origins of Madness*. New York: W. H. Freeman, 1991, p. 296.

[68] P. Tienari, “Genotype-environment interaction in schizophrenia-spectrum disorder: Long-term follow-up study of Finnish adoptees,” *The British Journal of Psychiatry*, vol. 184, no. 3, pp. 216–222, Mar. 2004.

[69] M. F. Egan, T. E. Goldberg, B. S. Kolachana, J. H. Callicott, C. M. Mazzanti, R. E. Straub, D. Goldman, and D. R. Weinberger, “Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 12, pp. 6917–22, Jun. 2001.

[70] S. Shifman, M. Bronstein, M. Sternfeld, A. Pisanté-Shalom, E. Lev-Lehman, A. Weizman, I. Reznik, B. Spivak, N. Grisaru, L. Karp, R. Schiffer, M. Kotler, R. D. Strous, M. Swartz-Vanetik, H. Y. Knobler, E. Shinar, J. S. Beckmann, B. Yakir, N. Risch, N. B. Zak, and A. Darvasi, “A highly significant association between a COMT haplotype and schizophrenia.,” *American journal of human genetics*, vol. 71, no. 6, pp. 1296–302, Dec. 2002.

[71] S. J. Glatt, “Association Between a Functional Catechol O-Methyltransferase Gene Polymorphism and Schizophrenia: Meta-Analysis of Case-Control and Family-Based Studies,” *American Journal of Psychiatry*, vol. 160, no. 3, pp. 469–476, Mar. 2003.

[72] H. Liu, S. C. Heath, C. Sobin, J. L. Roos, B. L. Galke, M. L. Blundell, M. Lenane, B. Robertson, E. M. Wijsman, J. L. Rapoport, J. A. Gogos, and M. Karayiorgou, “Genetic variation at the 22q11 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 6, pp. 3717–22, Mar. 2002.

[73] J. Mukai, H. Liu, R. A. Burt, D. E. Swor, W.-S. Lai, M. Karayiorgou, and J. A. Gogos, “Evidence that the gene encoding ZDHHC8 contributes to the risk of schizophrenia.,” *Nature genetics*, vol. 36, no. 7, pp. 725–31, Jul. 2004.

[74] H. Stefansson, E. Sigurdsson, V. Steinthorsdottir, S. Bjornsdottir, T. Sigmundsson, S. Ghosh, J. Brynjolfsson, S. Gunnarsdottir, O. Ivarsson, T. T. Chou, O. Hjaltason, B. Birgisdottir, H. Jonsson, V. G. Gudnadottir, E. Gudmundsdottir, A. Bjornsson, B. Ingvarsson, A. Ingason, S. Sigfusson, H. Hardardottir, R. P. Harvey, D. Lai, M. Zhou, D. Brunner, V. Mutel, A. Gonzalo, G. Lemke, J. Sainz, G. Johannesson, T. Andresson, D. Gudbjartsson, A. Manolescu, M. L. Frigge, M. E. Gurney, A. Kong, J. R. Gulcher, H. Petursson, and K. Stefansson, “Neuregulin 1 and susceptibility to schizophrenia.,” *American journal of human genetics*, vol. 71, no. 4, pp. 877–92, Oct. 2002.

[75] H. Stefansson, J. Sarginson, A. Kong, P. Yates, V. Steinthorsdottir, E. Gudfinnsson, S. Gunnarsdottir, N. Walker, H. Petursson, C. Crombie, A. Ingason, J. R. Gulcher, K. Stefansson, and D. S. Clair, “Association of Neuregulin 1 with Schizophrenia Confirmed in a Scottish Population,” *The American Journal of Human Genetics*, vol. 72, no. 1, pp. 83–87, Jan. 2003.

[76] A. P. Corvin, D. W. Morris, K. McGhee, S. Schwaiger, P. Scully, J. Quinn, D. Meagher, D. S. Clair, J. L. Waddington, and M. Gill, “Confirmation and refinement of an ‘at-risk’ haplotype for schizophrenia suggests the EST cluster, Hs.97362, as a potential susceptibility gene at the Neuregulin-1 locus.,” *Molecular psychiatry*, vol. 9, no. 2, pp. 208–13, Feb. 2004.

[77] N. M. Williams, A. Preece, G. Spurlock, N. Norton, H. J. Williams, S. Zammit, M. C. O’Donovan, and M. J. Owen, “Support for genetic variation in neuregulin 1 and susceptibility to schizophrenia.,” *Molecular psychiatry*, vol. 8, no. 5, pp. 485–7, May 2003.

[78] D. J. Gerber, D. Hall, T. Miyakawa, S. Demars, J. A. Gogos, M. Karayiorgou, and S. Tonegawa, “Evidence for association of schizophrenia with genetic variation in the 8p21.3 gene, PPP3CC, encoding the calcineurin gamma subunit.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 15, pp. 8993–8, Jul. 2003.

[79] T. Miyakawa, L. M. Leiter, D. J. Gerber, R. R. Gainetdinov, T. D. Sotnikova, H. Zeng, M. G. Caron, and S. Tonegawa, “Conditional calcineurin knockout mice exhibit multiple abnormal behaviors related to schizophrenia.,” *Proceedings of*

the National Academy of Sciences of the United States of America, vol. 100, no. 15, pp. 8987–92, Jul. 2003.

[80] R. E. Straub, Y. Jiang, C. J. MacLean, Y. Ma, B. T. Webb, M. V Myakishev, C. Harris-Kerr, B. Wormley, H. Sadek, B. Kadambi, A. J. Cesare, A. Gibberman, X. Wang, F. A. O'Neill, D. Walsh, and K. S. Kendler, "Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia.," *American journal of human genetics*, vol. 71, no. 2, pp. 337–48, Aug. 2002.

[81] E. J. C. G. van den Oord, P. F. Sullivan, Y. Jiang, D. Walsh, F. A. O'Neill, K. S. Kendler, and B. P. Riley, "Identification of a high-risk haplotype for the dystrobrevin binding protein 1 (DTNBP1) gene in the Irish study of high-density schizophrenia families.," *Molecular psychiatry*, vol. 8, no. 5, pp. 499–510, May 2003.

[82] S. G. Schwab, M. Knapp, S. Mondabon, J. Hallmayer, M. Borrmann-Hassenbach, M. Albus, B. Lerer, M. Rietschel, M. Trixler, W. Maier, and D. B. Wildenauer, "Support for association of schizophrenia with genetic variation in the 6p22.3 gene, dysbindin, in sib-pair families with linkage and in an additional sample of triad families.," *American journal of human genetics*, vol. 72, no. 1, pp. 185–90, Jan. 2003.

[83] J. X. Tang, J. Zhou, J. B. Fan, X. W. Li, Y. Y. Shi, N. F. Gu, G. Y. Feng, Y. L. Xing, J. G. Shi, and L. He, "Family-based association study of DTNBP1 in 6p22.3 and schizophrenia.," *Molecular psychiatry*, vol. 8, no. 8, pp. 717–8, Aug. 2003.

[84] A. Van Den Bogaert, J. Schumacher, T. G. Schulze, A. C. Otte, S. Ohlraun, S. Kovalenko, T. Becker, J. Freudenberg, E. G. Jönsson, M. Mattila-Evenden, G. C. Sedvall, P. M. Czerski, P. Kapelski, J. Hauser, W. Maier, M. Rietschel, P. Propping, M. M. Nöthen, and S. Cichon, "The DTNBP1 (dysbindin) gene contributes to schizophrenia, depending on family history of the disease.," *American journal of human genetics*, vol. 73, no. 6, pp. 1438–43, Dec. 2003.

[85] B. Funke, C. T. Finn, A. M. Plocik, S. Lake, P. DeRosse, J. M. Kane, R. Kucherlapati, and A. K. Malhotra, "Association of the DTNBP1 locus with schizophrenia in a U.S. population.," *American journal of human genetics*, vol. 75, no. 5, pp. 891–8, Nov. 2004.

[86] G. Kirov, D. Ivanov, N. M. Williams, A. Preece, I. Nikolov, R. Milev, S. Koleva, A. Dimitrova, D. Toncheva, M. C. O'Donovan, and M. J. Owen, "Strong evidence for association between the dystrobrevin binding protein 1 gene (DTNBP1) and schizophrenia in 488 parent-offspring trios from Bulgaria.," *Biological psychiatry*, vol. 55, no. 10, pp. 971–5, May 2004.

[87] T. Numakawa, Y. Yagasaki, T. Ishimoto, T. Okada, T. Suzuki, N. Iwata, N. Ozaki, T. Taguchi, M. Tatsumi, K. Kamijima, R. E. Straub, D. R. Weinberger, H. Kunugi, and R. Hashimoto, "Evidence of novel neuronal functions of dysbindin, a susceptibility gene for schizophrenia.," *Human molecular genetics*, vol. 13, no. 21, pp. 2699–708, Nov. 2004.

[88] N. M. Williams, A. Preece, D. W. Morris, G. Spurlock, N. J. Bray, M. Stephens, N. Norton, H. Williams, M. Clement, S. Dwyer, C. Curran, J. Wilkinson, V. Moskvina, J. L. Waddington, M. Gill, A. P. Corvin, S. Zammit, G. Kirov, M. J. Owen, and M. C. O'Donovan, "Identification in 2 independent samples of a novel schizophrenia risk haplotype of the dystrobrevin binding protein gene (DTNBP1).," *Archives of general psychiatry*, vol. 61, no. 4, pp. 336–44, Apr. 2004.

[89] T. Li, F. Zhang, X. Liu, X. Sun, P. C. Sham, C. Crombie, X. Ma, Q. Wang, H. Meng, W. Deng, P. Yates, X. Hu, N. Walker, R. M. Murray, D. St Clair, and D. A. Collier, "Identifying potential risk haplotypes for schizophrenia at the DTNBP1 locus in Han Chinese and Scottish populations.," *Molecular psychiatry*, vol. 10, no. 11, pp. 1037–44, Nov. 2005.

[90] I. Chumakov, M. Blumenfeld, O. Guerassimenko, L. Cavarec, M. Palicio, H. Abderrahim, L. Bougueleret, C. Barry, H. Tanaka, P. La Rosa, A. Puech, N. Tahri, A. Cohen-Akenine, S. Delabrosse, S. Lissarrague, F.-P. Picard, K. Maurice, L. Essioux, P. Millasseau, P. Grel, V. Debailleul, A.-M. Simon, D. Caterina, I. Dufaure, K. Malekzadeh, M. Belova, J.-J. Luan, M. Bouillot, J.-L. Sambucy, G. Primas, M. Saumier, N. Boubkiri, S. Martin-Saumier, M. Nasroune, H. Peixoto, A. Delaye, V. Pinchot, M. Bastucci, S. Guillou, M. Chevillon, R. Sainz-Fuertes, S. Meguenni, J. Aurich-Costa, D. Cherif, A. Gimalac, C. Van Duijn, D. Gauvreau, G. Ouellette, I. Fortier, J. Raelson, T. Sherbatich, N. Riazanskaia, E. Rogaev, P. Raeymaekers, J. Aerssens, F. Konings, W. Luyten, F. Macciardi, P. C. Sham, R. E. Straub, D. R. Weinberger, N. Cohen, D. Cohen, G. Ouelette, and J. Realson, "Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 21, pp. 13675–80, Oct. 2002.

[91] M. Korostishevsky, M. Kaganovich, A. Cholostoy, M. Ashkenazi, Y. Ratner, D. Dahary, J. Bernstein, U. Bening-Abu-Shach, E. Ben-Asher, D. Lancet, M. Ritsner, and R. Navon, "Is the G72/G30 locus associated with schizophrenia? single nucleotide polymorphisms, haplotypes, and gene expression analysis.," *Biological psychiatry*, vol. 56, no. 3, pp. 169–76, Aug. 2004.

[92] F. Zou, C. Li, S. Duan, Y. Zheng, N. Gu, G. Feng, Y. Xing, J. Shi, and L. He, "A family-based study of the association between the G72/G30 genes and schizophrenia in the Chinese population.," *Schizophrenia research*, vol. 73, no. 2–3, pp. 257–61, Mar. 2005.

[93] M. Korostishevsky, I. Kremer, M. Kaganovich, A. Cholostoy, I. Murad, M. Muhaheed, I. Bannoura, M. Rietschel, M. Dobrusin, U. Bening-Abu-Shach, R. H. Belmaker, W. Maier, R. P. Ebstein, and R. Navon, "Transmission disequilibrium and haplotype analyses of the G72/G30 locus: suggestive linkage to schizophrenia in Palestinian Arabs living in the North of Israel.," *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, vol. 141B, no. 1, pp. 91–5, Jan. 2006.

[94] X. Wang, G. He, N. Gu, J. Yang, J. Tang, Q. Chen, X. Liu, Y. Shen, X. Qian, W. Lin, Y. Duan, G. Feng, and L. He, "Association of G72/G30 with schizophrenia in the Chinese population.," *Biochemical and biophysical research communications*, vol. 319, no. 4, pp. 1281–6, Jul. 2004.

[95] A. M. Addington, M. Gornick, A. L. Sporn, N. Gogtay, D. Greenstein, M. Lenane, P. Gochman, N. Baker, R. Balkissoon, R. K. Vakkalanka, D. R. Weinberger, R. E. Straub, and J. L. Rapoport, "Polymorphisms in the 13q33.2 gene G72/G30 are associated with childhood-onset schizophrenia and psychosis not otherwise specified.," *Biological psychiatry*, vol. 55, no. 10, pp. 976–80, May 2004.

[96] J. Schumacher, R. A. Jamra, J. Freudenberg, T. Becker, S. Ohlraun, A. C. J. Otte, M. Tullius, S. Kovalenko, A. Van Den Bogaert, W. Maier, M. Rietschel, P. Propping, M. M. Nöthen, and S. Cichon, "Examination of G72 and D-amino-acid oxidase as genetic risk factors for schizophrenia and bipolar affective disorder.," *Molecular psychiatry*, vol. 9, no. 2, pp. 203–7, Feb. 2004.

[97] J. Duan, M. Martinez, A. R. Sanders, C. Hou, N. Saitou, T. Kitano, B. J. Mowry, R. R. Crowe, J. M. Silverman, D. F. Levinson, and P. V. Gejman, "Polymorphisms in the Trace Amine Receptor 4 (TRAR4) Gene on Chromosome

6q23.2 Are Associated with Susceptibility to Schizophrenia," *null*, vol. null, no. null.

[98] W. Hennah, T. Varilo, M. Kestilä, T. Paunio, R. Arajärvi, J. Haukka, A. Parker, R. Martin, S. Levitzky, T. Partonen, J. Meyer, J. Lönnqvist, L. Peltonen, and J. Ekelund, "Haplotype transmission analysis provides evidence of association for DISC1 to schizophrenia and suggests sex-dependent effects.," *Human molecular genetics*, vol. 12, no. 23, pp. 3151–9, Dec. 2003.

[99] C. A. Hodgkinson, D. Goldman, J. Jaeger, S. Persaud, J. M. Kane, R. H. Lipsky, and A. K. Malhotra, "Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder.," *American journal of human genetics*, vol. 75, no. 5, pp. 862–72, Nov. 2004.

[100] P. A. Thomson, N. R. Wray, J. K. Millar, K. L. Evans, S. Le Hellard, A. Condie, W. J. Muir, D. H. R. Blackwood, and D. J. Porteous, "Association between the TRAX/DISC locus and both bipolar disorder and schizophrenia in the Scottish population.," *Molecular psychiatry*, vol. 10, no. 7, pp. 657–68, 616, Jul. 2005.

[101] J. H. Callicott, R. E. Straub, L. Pezawas, M. F. Egan, V. S. Mattay, A. R. Hariri, B. A. Verchinski, A. Meyer-Lindenberg, R. Balkissoon, B. Kolachana, T. E. Goldberg, and D. R. Weinberger, "Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 24, pp. 8627–32, Jun. 2005.

[102] K. Mirnics, F. A. Middleton, D. A. Lewis, and P. Levitt, "Analysis of complex brain disorders with gene expression microarrays: schizophrenia as a disease of the synapse," *null*, vol. null, no. null.

[103] K. V. Chowdari, "Association and linkage analyses of RGS4 polymorphisms in schizophrenia," *Human Molecular Genetics*, vol. 11, no. 12, pp. 1373–1380, Jun. 2002.

[104] X. Chen, C. Dunham, S. Kendler, X. Wang, F. A. O'Neill, D. Walsh, and K. S. Kendler, "Regulator of G-protein signaling 4 (RGS4) gene is associated with schizophrenia in Irish high density families.," *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, vol. 129B, no. 1, pp. 23–6, Aug. 2004.

[105] E. S. Emamian, D. Hall, M. J. Birnbaum, M. Karayiorgou, and J. A. Gogos, "Convergent evidence for impaired AKT1-GSK3beta signaling in schizophrenia.," *Nature genetics*, vol. 36, no. 2, pp. 131–7, Feb. 2004.

[106] M. Ikeda, N. Iwata, T. Suzuki, T. Kitajima, Y. Yamanouchi, Y. Kinoshita, T. Inada, and N. Ozaki, "Association of AKT1 with schizophrenia confirmed in a Japanese population.," *Biological psychiatry*, vol. 56, no. 9, pp. 698–700, Nov. 2004.

[107] S. G. Schwab, B. Hoefgen, C. Hanses, M. B. Hassenbach, M. Albus, B. Lerer, M. Trixler, W. Maier, and D. B. Wildenauer, "Further evidence for association of variants in the AKT1 gene with schizophrenia in a sample of European sib-pair families.," *Biological psychiatry*, vol. 58, no. 6, pp. 446–50, Sep. 2005.

[108] T. Ohtsuki, T. Inada, and T. Arinami, "Failure to confirm association between AKT1 haplotype and schizophrenia in a Japanese case-control population.," *Molecular psychiatry*, vol. 9, no. 11, pp. 981–3, Nov. 2004.

[109] A. K. Tiwari, C. C. Zai, D. J. Müller, and J. L. Kennedy, "Genetics in schizophrenia: where are we and what next?," *Dialogues in clinical neuroscience*, vol. 12, no. 3, pp. 289–303, Jan. 2010.

[110] N. C. Allen, S. Bagade, M. B. McQueen, J. P. A. Ioannidis, F. K. Kavvoura, M. J. Khoury, R. E. Tanzi, and L. Bertram, "Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database.," *Nature genetics*, vol. 40, no. 7, pp. 827–34, Jul. 2008.

[111] E. F. Torrey, B. M. Barci, M. J. Webster, J. J. Bartko, J. H. Meador-Woodruff, and M. B. Knable, "Neurochemical markers for schizophrenia, bipolar disorder, and major depression in postmortem brains.," *Biological psychiatry*, vol. 57, no. 3, pp. 252–60, Feb. 2005.

[112] V. Z. Chong, M. Thompson, S. Beltaifa, M. J. Webster, A. J. Law, and C. S. Weickert, "Elevated neuregulin-1 and ErbB4 protein in the prefrontal cortex of schizophrenic patients.," *Schizophrenia research*, vol. 100, no. 1–3, pp. 270–80, Mar. 2008.

[113] C.-G. Hahn, H.-Y. Wang, D.-S. Cho, K. Talbot, R. E. Gur, W. H. Berrettini, K. Bakshi, J. Kamins, K. E. Borgmann-Winter, S. J. Siegel, R. J. Gallop, and S. E. Arnold, "Altered neuregulin 1-erbB4 signaling contributes to NMDA receptor

hypofunction in schizophrenia.," *Nature medicine*, vol. 12, no. 7, pp. 824–8, Jul. 2006.

[114] L. E. Hong, I. Wonodi, O. C. Stine, B. D. Mitchell, and G. K. Thaker, "Evidence of missense mutations on the neuregulin 1 gene affecting function of prepulse inhibition.," *Biological psychiatry*, vol. 63, no. 1, pp. 17–23, Jan. 2008.

[115] C. Zhao, Z. Xu, J. Chen, Z. Yu, K.-L. Tong, W.-S. Lo, F. W. Pun, S.-K. Ng, S.-Y. Tsang, and H. Xue, "Two isoforms of GABA(A) receptor beta2 subunit with different electrophysiological properties: Differential expression and genotypical correlations in schizophrenia.," *Molecular psychiatry*, vol. 11, no. 12, pp. 1092–105, Dec. 2006.

[116] D. Li, Y. Duan, and L. He, "Association study of serotonin 2A receptor (5-HT2A) gene with schizophrenia and suicidal behavior using systematic meta-analysis.," *Biochemical and biophysical research communications*, vol. 340, no. 3, pp. 1006–15, Feb. 2006.

[117] A. L. López-Figueroa, C. S. Norton, M. O. López-Figueroa, D. Armellini-Dodel, S. Burke, H. Akil, J. F. López, and S. J. Watson, "Serotonin 5-HT1A, 5-HT1B, and 5-HT2A receptor mRNA expression in subjects with major depression, bipolar disorder, and schizophrenia.," *Biological psychiatry*, vol. 55, no. 3, pp. 225–33, Feb. 2004.

[118] L. Martucci, A. H. C. Wong, V. De Luca, O. Likhodi, G. W. H. Wong, N. King, and J. L. Kennedy, "N-methyl-D-aspartate receptor NR2B subunit gene GRIN2B in schizophrenia and bipolar disorder: Polymorphisms and mRNA levels.," *Schizophrenia research*, vol. 84, no. 2–3, pp. 214–21, Jun. 2006.

[119] J. Shi, E. S. Gershon, and C. Liu, "Genetic associations with schizophrenia: meta-analyses of 12 candidate genes.," *Schizophrenia research*, vol. 104, no. 1–3, pp. 96–107, Sep. 2008.

[120] J. Söderlund, J. Schröder, C. Nordin, M. Samuelsson, L. Walther-Jallow, H. Karlsson, S. Erhardt, and G. Engberg, "Activation of brain interleukin-1beta in schizophrenia.," *Molecular psychiatry*, vol. 14, no. 12, pp. 1069–71, Dec. 2009.

[121] H. Stefansson, R. a Ophoff, S. Steinberg, O. a Andreassen, S. Cichon, D. Rujescu, T. Werge, O. P. H. Pietiläinen, O. Mors, P. B. Mortensen, E. Sigurdsson, O. Gustafsson, M. Nyegaard, A. Tuulio-Henriksson, A. Ingason, T. Hansen, J. Suvisaari, J. Lonnqvist, T. Paunio, A. D. Børglum, A. Hartmann, A. Fink-Jensen,

M. Nordentoft, D. Hougaard, B. Norgaard-Pedersen, Y. Böttcher, J. Olesen, R. Breuer, H.-J. Möller, I. Giegling, H. B. Rasmussen, S. Timm, M. Mattheisen, I. Bitter, J. M. Réthelyi, B. B. Magnusdottir, T. Sigmundsson, P. Olason, G. Masson, J. R. Gulcher, M. Haraldsson, R. Fossdal, T. E. Thorgeirsson, U. Thorsteinsdottir, M. Ruggeri, S. Tosato, B. Franke, E. Strengman, L. a Kiemeney, I. Melle, S. Djurovic, L. Abramova, V. Kaleda, J. Sanjuan, R. de Frutos, E. Bramon, E. Vassos, G. Fraser, U. Ettinger, M. Picchioni, N. Walker, T. Toulopoulou, A. C. Need, D. Ge, J. L. Yoon, K. V Shianna, N. B. Freimer, R. M. Cantor, R. Murray, A. Kong, V. Golimbet, A. Carracedo, C. Arango, J. Costas, E. G. Jönsson, L. Terenius, I. Agartz, H. Petursson, M. M. Nöthen, M. Rietschel, P. M. Matthews, P. Muglia, L. Peltonen, D. St Clair, D. B. Goldstein, K. Stefansson, and D. a Collier, "Common variants conferring risk of schizophrenia.," *Nature*, vol. 460, no. 7256, pp. 744–7, Aug. 2009.

[122] S. J. Glatt, R. S. Wang, Y.-C. Yeh, M. T. Tsuang, and S. V Faraone, "Five NOTCH4 polymorphisms show weak evidence for association with schizophrenia: evidence from meta-analyses.," *Schizophrenia research*, vol. 73, no. 2–3, pp. 281–90, Mar. 2005.

[123] Z. Wang, J. Wei, X. Zhang, Y. Guo, Q. Xu, S. Liu, J. Shi, Y. Yu, G. Ju, Y. Li, and Y. Shen, "A review and re-evaluation of an association between the NOTCH4 locus and schizophrenia.," *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, vol. 141B, no. 8, pp. 902–6, Dec. 2006.

[124] M. R. Etherton, C. A. Blaiss, C. M. Powell, and T. C. Südhof, "Mouse neurexin-1alpha deletion causes correlated electrophysiological and behavioral changes consistent with cognitive impairments.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 42, pp. 17998–8003, Oct. 2009.

[125] S. H. Fatemi, D. P. King, T. J. Reutiman, T. D. Folsom, J. A. Laurence, S. Lee, Y.-T. Fan, S. A. Paciga, M. Conti, and F. S. Menniti, "PDE4B polymorphisms and decreased PDE4B expression are associated with schizophrenia.," *Schizophrenia research*, vol. 101, no. 1–3, pp. 36–49, Apr. 2008.

[126] A. K. Kähler, M. K. Otnaess, K. V Wirgenes, T. Hansen, E. G. Jönsson, I. Agartz, H. Hall, T. Werge, G. Morken, O. Mors, E. Mellerup, H. Dam, P. Koefod, I. Melle, V. M. Steen, O. A. Andreassen, and S. Djurovic, "Association study of PDE4B gene variants in Scandinavian schizophrenia and bipolar disorder multicenter case-control samples.," *American journal of medical genetics. Part B,*

Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics, vol. 153B, no. 1, pp. 86–96, Jan. 2010.

[127] J. A. Siuciak, S. A. McCarthy, D. S. Chapin, and A. N. Martin, "Behavioral and neurochemical characterization of mice deficient in the phosphodiesterase-4B (PDE4B) enzyme.," *Psychopharmacology*, vol. 197, no. 1, pp. 115–26, Mar. 2008.

[128] L. Tomppo, W. Hennah, P. Lahermo, A. Loukola, A. Tuulio-Henriksson, J. Suvisaari, T. Partonen, J. Ekelund, J. Lönnqvist, and L. Peltonen, "Association between genes of Disrupted in schizophrenia 1 (DISC1) interactors and schizophrenia supports the role of the DISC1 pathway in the etiology of major mental illnesses.," *Biological psychiatry*, vol. 65, no. 12, pp. 1055–62, Jun. 2009.

[129] J. A. Gogos, M. Santha, Z. Takacs, K. D. Beck, V. Luine, L. R. Lucas, J. V Nadler, and M. Karayiorgou, "The gene encoding proline dehydrogenase modulates sensorimotor gating in mice.," *Nature genetics*, vol. 21, no. 4, pp. 434–9, Apr. 1999.

[130] E. Tunbridge, P. W. J. Burnet, M. S. Sodhi, and P. J. Harrison, "Catechol-o-methyltransferase (COMT) and proline dehydrogenase (PRODH) mRNAs in the dorsolateral prefrontal cortex in schizophrenia, bipolar disorder, and major depression.," *Synapse (New York, N.Y.)*, vol. 51, no. 2, pp. 112–8, Feb. 2004.

[131] C. Erbel-Sieler, C. Dudley, Y. Zhou, X. Wu, S. J. Estill, T. Han, R. Diaz-Arrastia, E. W. Brunskill, S. S. Potter, and S. L. McKnight, "Behavioral and regulatory abnormalities in mice deficient in the NPAS1 and NPAS3 transcription factors.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 37, pp. 13648–53, Sep. 2004.

[132] A. Guidotti, J. Auta, J. M. Davis, V. Di-Giorgi-Gerevini, Y. Dwivedi, D. R. Grayson, F. Impagnatiello, G. Pandey, C. Pesold, R. Sharma, D. Uzunov, E. Costa, and V. DiGiorgi Gerevini, "Decrease in reelin and glutamic acid decarboxylase67 (GAD67) expression in schizophrenia and bipolar disorder: a postmortem brain study.," *Archives of general psychiatry*, vol. 57, no. 11, pp. 1061–9, Nov. 2000.

[133] B. K. Lipska, T. Peters, T. M. Hyde, N. Halim, C. Horowitz, S. Mitkus, C. S. Weickert, M. Matsumoto, A. Sawa, R. E. Straub, R. Vakkalanka, M. M. Herman, D. R. Weinberger, and J. E. Kleinman, "Expression of DISC1 binding partners is

reduced in schizophrenia and associated with DISC1 SNPs.," *Human molecular genetics*, vol. 15, no. 8, pp. 1245–58, Apr. 2006.

[134] C. M. Lewis, D. F. Levinson, L. H. Wise, L. E. DeLisi, R. E. Straub, I. Hovatta, N. M. Williams, S. G. Schwab, A. E. Pulver, S. V Faraone, L. M. Brzustowicz, C. A. Kaufmann, D. L. Garver, H. M. D. Gurling, E. Lindholm, H. Coon, H. W. Moises, W. Byerley, S. H. Shaw, A. Mesen, R. Sherrington, F. A. O'Neill, D. Walsh, K. S. Kendler, J. Ekelund, T. Paunio, J. Lönnqvist, L. Peltonen, M. C. O'Donovan, M. J. Owen, D. B. Wildenauer, W. Maier, G. Nestadt, J.-L. Blouin, S. E. Antonarakis, B. J. Mowry, J. M. Silverman, R. R. Crowe, C. R. Cloninger, M. T. Tsuang, D. Malaspina, J. M. Harkavy-Friedman, D. M. Svrakic, A. S. Bassett, J. Holcomb, G. Kalsi, A. McQuillin, J. Brynjolfson, T. Sigmundsson, H. Petursson, E. Jazin, T. Zoëga, and T. Helgason, "Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia.," *American journal of human genetics*, vol. 73, no. 1, pp. 34–48, Jul. 2003.

[135] M. Y. M. Ng, D. F. Levinson, S. V Faraone, B. K. Suarez, L. E. DeLisi, T. Arinami, B. Riley, T. Paunio, A. E. Pulver, Irmansyah, P. A. Holmans, M. Escamilla, D. B. Wildenauer, N. M. Williams, C. Laurent, B. J. Mowry, L. M. Brzustowicz, M. Maziade, P. Sklar, D. L. Garver, G. R. Abecasis, B. Lerer, M. D. Fallin, H. M. D. Gurling, P. V Gejman, E. Lindholm, H. W. Moises, W. Byerley, E. M. Wijsman, P. Forabosco, M. T. Tsuang, H.-G. Hwu, Y. Okazaki, K. S. Kendler, B. Wormley, A. Fanous, D. Walsh, F. A. O'Neill, L. Peltonen, G. Nestadt, V. K. Lasseter, K. Y. Liang, G. M. Papadimitriou, D. G. Dikeos, S. G. Schwab, M. J. Owen, M. C. O'Donovan, N. Norton, E. Hare, H. Raventos, H. Nicolini, M. Albus, W. Maier, V. L. Nimgaonkar, L. Terenius, J. Mallet, M. Jay, S. Godard, D. Nertney, M. Alexander, R. R. Crowe, J. M. Silverman, A. S. Bassett, M.-A. Roy, C. Mérette, C. N. Pato, M. T. Pato, J. L. Roos, Y. Kohn, D. Amann-Zalcenstein, G. Kalsi, A. McQuillin, D. Curtis, J. Brynjolfson, T. Sigmundsson, H. Petursson, A. R. Sanders, J. Duan, E. Jazin, M. Myles-Worsley, M. Karayiorgou, and C. M. Lewis, "Meta-analysis of 32 genome-wide linkage studies of schizophrenia.," *Molecular psychiatry*, vol. 14, no. 8, pp. 774–85, Aug. 2009.

[136] N. M. Williams, N. Norton, H. Williams, B. Ekholm, M. L. Hamshere, Y. Lindblom, K. V Chowdari, A. G. Cardno, S. Zammit, L. A. Jones, K. C. Murphy, R. D. Sanders, G. McCarthy, M. Y. Gray, G. Jones, P. Holmans, V. Nimgaonkar, R. Adolfson, U. Osby, L. Terenius, G. Sedvall, M. C. O'Donovan, and M. J. Owen, "A systematic genomewide linkage study in 353 sib pairs with schizophrenia.," *American journal of human genetics*, vol. 73, no. 6, pp. 1355–67, Dec. 2003.

[137] R. Khan, “Gene Expression: Linkage versus association: a mini-primer.” [Online]. Available: http://www.gnxp.com/blog/2006/07/linkage-versus-association-mini-primer.php.

[138] T. Keith, “GEN | Magazine Articles: Human Genome-Wide Association Studies.” [Online]. Available: http://www.genengnews.com/gen-articles/human-genome-wide-association-studies/1970/.

[139] B. K. Suarez, J. Duan, A. R. Sanders, A. L. Hinrichs, C. H. Jin, C. Hou, N. G. Buccola, N. Hale, A. N. Weilbaecher, D. a Nertney, A. Olincy, S. Green, A. W. Schaffer, C. J. Smith, D. E. Hannah, J. P. Rice, N. J. Cox, M. Martinez, B. J. Mowry, F. Amin, J. M. Silverman, D. W. Black, W. F. Byerley, R. R. Crowe, R. Freedman, C. R. Cloninger, D. F. Levinson, and P. V Gejman, “Genomewide linkage scan of 409 European-ancestry and African American families with schizophrenia: suggestive evidence of linkage at 8p23.3-p21.2 and 11p13.1-q14.1 in the combined sample.,” *American journal of human genetics*, vol. 78, no. 2, pp. 315–33, Feb. 2006.

[140] S. Mah, M. R. Nelson, L. E. Delisi, R. H. Reneland, N. Markward, M. R. James, D. R. Nyholt, N. Hayward, H. Handoko, B. Mowry, S. Kammerer, and A. Braun, “Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia.,” *Molecular psychiatry*, vol. 11, no. 5, pp. 471–8, May 2006.

[141] T. Lencz, T. V Morgan, M. Athanasiou, B. Dain, C. R. Reed, J. M. Kane, R. Kucherlapati, and A. K. Malhotra, “Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia.,” *Molecular psychiatry*, vol. 12, no. 6, pp. 572–80, Jun. 2007.

[142] S. Sun, F. Wang, J. Wei, L. Y. Cao, G. Y. Wu, L. Lu, T. A. Kosten, T. R. Kosten, and X. Y. Zhang, “Association between interleukin-3 receptor alpha polymorphism and schizophrenia in the Chinese population.,” *Neuroscience letters*, vol. 440, no. 1, pp. 35–7, Jul. 2008.

[143] S. Sun, J. Wei, H. Li, S. Jin, P. Li, G. Ju, Y. Liu, and X. Y. Zhang, “A family-based study of the IL3RA gene on susceptibility to schizophrenia in a Chinese Han population.,” *Brain research*, vol. 1268, pp. 13–6, May 2009.

[144] H. Stefansson, D. Rujescu, S. Cichon, O. P. H. Pietiläinen, A. Ingason, S. Steinberg, R. Fossdal, E. Sigurdsson, T. Sigmundsson, J. E. Buizer-Voskamp, T. Hansen, K. D. Jakobsen, P. Muglia, C. Francks, P. M. Matthews, A. Gylfason, B.

V Halldorsson, D. Gudbjartsson, T. E. Thorgeirsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, A. Bjornsson, S. Mattiasdottir, T. Blondal, M. Haraldsson, B. B. Magnusdottir, I. Giegling, H.-J. Möller, A. Hartmann, K. V Shianna, D. Ge, A. C. Need, C. Crombie, G. Fraser, N. Walker, J. Lonnqvist, J. Suvisaari, A. Tuulio-Henriksson, T. Paunio, T. Toulopoulou, E. Bramon, M. Di Forti, R. Murray, M. Ruggeri, E. Vassos, S. Tosato, M. Walshe, T. Li, C. Vasilescu, T. W. Mühleisen, A. G. Wang, H. Ullum, S. Djurovic, I. Melle, J. Olesen, L. A. Kiemeney, B. Franke, C. Sabatti, N. B. Freimer, J. R. Gulcher, U. Thorsteinsdottir, A. Kong, O. A. Andreassen, R. A. Ophoff, A. Georgi, M. Rietschel, T. Werge, H. Petursson, D. B. Goldstein, M. M. Nöthen, L. Peltonen, D. A. Collier, D. St Clair, and K. Stefansson, "Large recurrent microdeletions associated with schizophrenia.," *Nature*, vol. 455, no. 7210, pp. 232–6, Sep. 2008.

[145] H. C. Mefford, A. J. Sharp, C. Baker, A. Itsara, Z. Jiang, K. Buysse, S. Huang, V. K. Maloney, J. a Crolla, D. Baralle, A. Collins, C. Mercer, K. Norga, T. de Ravel, K. Devriendt, E. M. H. F. Bongers, N. de Leeuw, W. Reardon, S. Gimelli, F. Bena, R. C. Hennekam, A. Male, L. Gaunt, J. Clayton-Smith, I. Simonic, S. M. Park, S. G. Mehta, S. Nik-Zainal, C. G. Woods, H. V Firth, G. Parkin, M. Fichera, S. Reitano, M. Lo Giudice, K. E. Li, I. Casuga, A. Broomer, B. Conrad, M. Schwerzmann, L. Räber, S. Gallati, P. Striano, A. Coppola, J. L. Tolmie, E. S. Tobias, C. Lilley, L. Armengol, Y. Spysschaert, P. Verloo, A. De Coene, L. Goossens, G. Mortier, F. Speleman, E. van Binsbergen, M. R. Nelen, R. Hochstenbach, M. Poot, L. Gallagher, M. Gill, J. McClellan, M.-C. King, R. Regan, C. Skinner, R. E. Stevenson, S. E. Antonarakis, C. Chen, X. Estivill, B. Menten, G. Gimelli, S. Gribble, S. Schwartz, J. S. Sutcliffe, T. Walsh, S. J. L. Knight, J. Sebat, C. Romano, C. E. Schwartz, J. a Veltman, B. B. a de Vries, J. R. Vermeesch, J. C. K. Barber, L. Willatt, M. Tassabehji, and E. E. Eichler, "Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes.," *The New England journal of medicine*, vol. 359, no. 16, pp. 1685–99, Oct. 2008.

[146] The International Schizophrenia Consortium, "Rare chromosomal deletions and duplications increase risk of schizophrenia.," *Nature*, vol. 455, no. 7210, pp. 237–41, Sep. 2008.

[147] S. M. Purcell, N. R. Wray, J. L. Stone, P. M. Visscher, M. C. O'Donovan, P. F. Sullivan, and P. Sklar, "Common polygenic variation contributes to risk of schizophrenia and bipolar disorder.," *Nature*, vol. 460, no. 7256, pp. 748–52, Aug. 2009.

[148] M. C. O'Donovan, N. Craddock, N. Norton, H. Williams, T. Peirce, V. Moskvina, I. Nikolov, M. Hamshere, L. Carroll, L. Georgieva, S. Dwyer, P. Holmans, J. L. Marchini, C. C. a Spencer, B. Howie, H.-T. Leung, A. M. Hartmann, H.-J. Möller, D. W. Morris, Y. Shi, G. Feng, P. Hoffmann, P. Propping, C. Vasilescu, W. Maier, M. Rietschel, S. Zammit, J. Schumacher, E. M. Quinn, T. G. Schulze, N. M. Williams, I. Giegling, N. Iwata, M. Ikeda, A. Darvasi, S. Shifman, L. He, J. Duan, A. R. Sanders, D. F. Levinson, P. V Gejman, S. Cichon, M. M. Nöthen, M. Gill, A. Corvin, D. Rujescu, G. Kirov, M. J. Owen, N. G. Buccola, B. J. Mowry, R. Freedman, F. Amin, D. W. Black, J. M. Silverman, W. F. Byerley, and C. R. Cloninger, "Identification of loci associated with schizophrenia by genome-wide association and follow-up.," *Nature genetics*, vol. 40, no. 9, pp. 1053–5, Sep. 2008.

[149] M. C. O'Donovan, N. Norton, H. Williams, T. Peirce, V. Moskvina, I. Nikolov, M. Hamshere, L. Carroll, L. Georgieva, S. Dwyer, P. Holmans, J. L. Marchini, C. C. A. Spencer, B. Howie, H.-T. Leung, I. Giegling, A. M. Hartmann, H.-J. Möller, D. W. Morris, Y. Shi, G. Feng, P. Hoffmann, P. Propping, C. Vasilescu, W. Maier, M. Rietschel, S. Zammit, J. Schumacher, E. M. Quinn, T. G. Schulze, N. Iwata, M. Ikeda, A. Darvasi, S. Shifman, L. He, J. Duan, A. R. Sanders, D. F. Levinson, R. Adolfsson, U. Osby, L. Terenius, E. G. Jönsson, S. Cichon, M. M. Nöthen, M. Gill, A. P. Corvin, D. Rujescu, P. V Gejman, G. Kirov, N. Craddock, N. M. Williams, and M. J. Owen, "Analysis of 10 independent samples provides evidence for association between schizophrenia and a SNP flanking fibroblast growth factor receptor 2.," *Molecular psychiatry*, vol. 14, no. 1, pp. 30–6, Jan. 2009.

[150] L. Athanasiu, M. Mattingsdal, A. K. Kähler, A. Brown, O. Gustafsson, I. Agartz, I. Giegling, P. Muglia, S. Cichon, M. Rietschel, O. P. H. Pietiläinen, L. Peltonen, E. Bramon, D. Collier, D. S. Clair, E. Sigurdsson, H. Petursson, D. Rujescu, I. Melle, V. M. Steen, S. Djurovic, and O. A. Andreassen, "Gene variants associated with schizophrenia in a Norwegian genome-wide study are replicated in a large European cohort.," *Journal of psychiatric research*, vol. 44, no. 12, pp. 748–53, Sep. 2010.

[151] "Genome-wide association study identifies five new schizophrenia loci.," *Nature genetics*, vol. 43, no. 10, pp. 969–76, Oct. 2011.

[152] P. Jia, L. Wang, H. Y. Meltzer, and Z. Zhao, "Common variants conferring risk of schizophrenia : A pathway analysis of GWAS data,"*Schizophrenia Research*, vol. 122, no. 1–3, pp. 38–42, 2010.

[153] P. Jia, L. Wang, H. Y. Meltzer, and Z. Zhao, “Pathway-based analysis of GWAS datasets: effective but caution required.,” *The international journal of neuropsychopharmacology / official scientific journal of the Collegium Internationale Neuropsychopharmacologicum (CINP)*, vol. 14, no. 4, pp. 567–72, May 2011.

[154] K. Zhang, L. Zhang, W. Zhang, and J. Wang, “Pathway-based analysis for genome-wide association studies of schizophrenia to provide new insight in schizophrenia study,” *Chinese Science Bulletin*, vol. 56, no. 32, pp. 3398–3402, Sep. 2011.

[155] C. O’Dushlaine, E. Kenny, E. Heron, G. Donohoe, M. Gill, D. Morris, and A. Corvin, “Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia.” 2010.

[156] L. M. Brzustowicz, A. S. Bassett, R. Silva, and U. Simón, “miRNA-mediated risk for schizophrenia in 22q11 . 2 deletion syndrome,” vol. 3, no. December, pp. 1–9, 2012.

[157] A. S. Bassett, E. W. C. Chow, P. AbdelMalik, M. Gheorghiu, J. Husted, and R. Weksberg, “The schizophrenia phenotype in 22q11 deletion syndrome.,” *The American journal of psychiatry*, vol. 160, no. 9, pp. 1580–6, Sep. 2003.

[158] A. S. Bassett, K. Hodgkinson, E. W. Chow, S. Correia, L. E. Scutt, and R. Weksberg, “22q11 deletion syndrome in adults with schizophrenia.,” *American journal of medical genetics*, vol. 81, no. 4, pp. 328–37, Jul. 1998.

[159] A. Hamosh, A. F. Scott, J. S. Amberger, C. A. Bocchini, and V. A. McKusick, “Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders.,” *Nucleic acids research*, vol. 33, no. Database issue, pp. D514–7, Jan. 2005.

[160] *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR Fourth Edition (Text Revision)*. American Psychiatric Association, 2000, p. 943.

[161] P. V. Gejman, “Cases: Clinical Procedures.” [Online]. Available: http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?id=phd000188.

[162] P. V. Gejman, “Controls: Clinical Procedures.” [Online]. Available: http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?id=phd000187.

[163] “dbGaP | phs000167.v1.p1 | Molecular Genetics of Schizophrenia - nonGAIN Sample (MGS_nonGAIN).” [Online]. Available: http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000167.v1.p1&phv=71751&phd=189&pha=&pht=714&phvf=&phdf=&phaf=&phtf=&dssp=6&consent=&temp=1.

[164] “Genetic Analysis | SNP Analysis | Affymetrix.” [Online]. Available: http://www.affymetrix.com/browse/products.jsp?productId=131533&navMode=34000&navAction=jump&aId=productsNav#1_1.

[165] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham, “PLINK: a tool set for whole-genome association and population-based linkage analyses.,” *American journal of human genetics*, vol. 81, no. 3, pp. 559–75, Sep. 2007.

[166] S. Purcell, “PLINK: Whole genome data analysis toolset.” [Online]. Available: http://pngu.mgh.harvard.edu/~purcell/plink/.

[167] D. Maglott, J. Ostell, K. D. Pruitt, and T. Tatusova, “Entrez Gene: gene-centered information at NCBI.,” *Nucleic acids research*, vol. 35, no. Database issue, pp. D26–31, Jan. 2007.

[168] M. Kanehisa and S. Goto, “KEGG: kyoto encyclopedia of genes and genomes.,” *Nucleic acids research*, vol. 28, no. 1, pp. 27–30, Jan. 2000.

[169] M. A. Harris, J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, K. Eilbeck, S. Lewis, B. Marshall, C. Mungall, J. Richter, G. M. Rubin, J. A. Blake, C. Bult, M. Dolan, H. Drabkin, J. T. Eppig, D. P. Hill, L. Ni, M. Ringwald, R. Balakrishnan, J. M. Cherry, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. Engel, D. G. Fisk, J. E. Hirschman, E. L. Hong, R. S. Nash, A. Sethuraman, C. L. Theesfeld, D. Botstein, K. Dolinski, B. Feierbach, T. Berardini, S. Mundodi, S. Y. Rhee, R. Apweiler, D. Barrell, E. Camon, E. Dimmer, V. Lee, R. Chisholm, P. Gaudet, W. Kibbe, R. Kishore, E. M. Schwarz, P. Sternberg, M. Gwinn, L. Hannick, J. Wortman, M. Berriman, V. Wood, N. de la Cruz, P. Tonellato, P. Jaiswal, T. Seigfried, and R. White, “The Gene Ontology (GO) database and informatics resource.,” *Nucleic acids research*, vol. 32, no. Database issue, pp. D258–61, Jan. 2004.

[170] R. Fisher and S. Genetiker, *Statistical methods for research workers*. 1970.

[171] L. Cheng and X. Sheng, “Combination of ‘Combinations of P-values’.”

[172] R. Fisher, “On the interpretation of χ2 from contingency tables, and the calculation of P,” *Journal of the Royal Statistical Society*, 1922.

[173] E. W. Weisstein, “Fisher’s Exact Test -- from Wolfram MathWorld.” Wolfram Research, Inc.

[174] T. L. Saaty, “The Analytic Hierarchy Process,” 1980.

[175] K. Teknomo, “Analytic Hierarchy Process (AHP) Tutorial.” [Online]. Available: http://people.revoledu.com/kardi/tutorial/AHP/purchase.html.

[176] O. S. Vaidya and S. Kumar, “Analytic hierarchy process: An overview of applications,” *European Journal of Operational Research*, vol. 169, no. 1, pp. 1–29, Feb. 2006.

[177] G. Ustünkar and Y. Aydın Son, “METU-SNP: an integrated software system for SNP-complex disease association analysis.,” *Journal of integrative bioinformatics*, vol. 8, no. 1, p. 187, Jan. 2011.

[178] A. P. Boyle, E. L. Hong, M. Hariharan, Y. Cheng, M. A. Schaub, M. Kasowski, K. J. Karczewski, J. Park, B. C. Hitz, S. Weng, J. M. Cherry, and M. Snyder, “Annotation of functional variation in personal genomes using RegulomeDB.,” *Genome research*, vol. 22, no. 9, pp. 1790–7, Sep. 2012.

[179] A. Z. Dayem Ullah, N. R. Lemoine, and C. Chelala, “SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update).,” *Nucleic acids research*, vol. 40, no. Web Server issue, pp. W65–70, Jul. 2012.

[180] S. A. Forbes, N. Bindal, S. Bamford, C. Cole, C. Y. Kok, D. Beare, M. Jia, R. Shepherd, K. Leung, A. Menzies, J. W. Teague, P. J. Campbell, M. R. Stratton, and P. A. Futreal, “COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer.,” *Nucleic acids research*, vol. 39, no. Database issue, pp. D945–50, Jan. 2011.

[181] D. W. Huang, B. T. Sherman, and R. A. Lempicki, “Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.,” *Nature protocols*, vol. 4, no. 1, pp. 44–57, Jan. 2009.

[182] D. Croft, G. O’Kelly, G. Wu, R. Haw, M. Gillespie, L. Matthews, M. Caudy, P. Garapati, G. Gopinath, B. Jassal, S. Jupe, I. Kalatskaya, S. Mahajan, B. May, N.

Ndegwa, E. Schmidt, V. Shamovsky, C. Yung, E. Birney, H. Hermjakob, P. D'Eustachio, and L. Stein, "Reactome: a database of reactions, pathways and biological processes.," *Nucleic acids research*, vol. 39, no. Database issue, pp. D691–7, Jan. 2011.

[183] D. Nishimura, "BioCarta," *Biotech Software <html_ent glyph="@amp;" ascii="&"/> Internet Report*, vol. 2, no. 3, pp. 117–120, Jun. 2001.

[184] H. Mi and P. Thomas, "PANTHER pathway: an ontology-based pathway database coupled with data analysis tools.," *Methods in molecular biology (Clifton, N.J.)*, vol. 563, pp. 123–40, Jan. 2009.

[185] G. D. Jr, B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, and R. A. Lempicki, "DAVID: Database for Annotation, Visualization, and Integrated Discovery."

[186] D. Warde-Farley, S. L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, C. Grouios, F. Kazi, C. T. Lopes, A. Maitland, S. Mostafavi, J. Montojo, Q. Shao, G. Wright, G. D. Bader, and Q. Morris, "The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function.," *Nucleic acids research*, vol. 38, no. Web Server issue, pp. W214–20, Jul. 2010.

[187] P. Flicek, M. R. Amode, D. Barrell, K. Beal, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fairley, S. Fitzgerald, L. Gil, L. Gordon, M. Hendrix, T. Hourlier, N. Johnson, A. K. Kähäri, D. Keefe, S. Keenan, R. Kinsella, M. Komorowska, G. Koscielny, E. Kulesha, P. Larsson, I. Longden, W. McLaren, M. Muffato, B. Overduin, M. Pignatelli, B. Pritchard, H. S. Riat, G. R. S. Ritchie, M. Ruffier, M. Schuster, D. Sobral, Y. A. Tang, K. Taylor, S. Trevanion, J. Vandrovcova, S. White, M. Wilson, S. P. Wilder, B. L. Aken, E. Birney, F. Cunningham, I. Dunham, R. Durbin, X. M. Fernández-Suarez, J. Harrow, J. Herrero, T. J. P. Hubbard, A. Parker, G. Proctor, G. Spudich, J. Vogel, A. Yates, A. Zadissa, and S. M. J. Searle, "Ensembl 2012.," *Nucleic acids research*, vol. 40, no. Database issue, pp. D84–90, Jan. 2012.

[188] R. J. Kinsella, A. Kähäri, S. Haider, J. Zamora, G. Proctor, G. Spudich, J. Almeida-King, D. Staines, P. Derwent, A. Kerhornou, P. Kersey, and P. Flicek, "Ensembl BioMarts: a hub for data retrieval across taxonomic space.," *Database : the journal of biological databases and curation*, vol. 2011, p. bar030, Jan. 2011.

[189] C. M. Durand, C. Kappeler, C. Betancur, R. Delorme, H. Quach, H. Goubran-Botros, J. Melke, G. Nygren, N. Chabane, F. Bellivier, A. Szoke, F. Schurhoff, M. Rastam, H. Anckarsäter, C. Gillberg, M. Leboyer, and T. Bourgeron, "Expression and genetic variability of PCDH11Y, a gene specific to Homo sapiens and candidate for susceptibility to psychiatric disorders.," *American journal of medical genetics. Part B, Neuropsychiatric genetics□ : the official publication of the International Society of Psychiatric Genetics*, vol. 141B, no. 1, pp. 67–70, Jan. 2006.

[190] M. Giouzeli, N. A. Williams, L. J. Lonie, L. E. DeLisi, and T. J. Crow, "ProtocadherinX/Y, a candidate gene-pair for schizophrenia and schizoaffective disorder: a DHPLC investigation of genomic sequence.," *American journal of medical genetics. Part B, Neuropsychiatric genetics□ : the official publication of the International Society of Psychiatric Genetics*, vol. 129B, no. 1, pp. 1–9, Aug. 2004.

[191] M. Bozza, L. Bernardini, A. Novelli, P. Brovedani, E. Moretti, R. Canapicchi, V. Doccini, T. Filippi, and A. Battaglia, "6p25 Interstitial deletion in two dizygotic twins with gyral pattern anomaly and speech and language disorder.," *European journal of paediatric neurology□ : EJPN□ : official journal of the European Paediatric Neurology Society*, vol. null, no. null, Oct. 2012.

[192] K. J. Brennand, A. Simone, J. Jou, C. Gelboin-Burkhart, N. Tran, S. Sangar, Y. Li, Y. Mu, G. Chen, D. Yu, S. McCarthy, J. Sebat, and F. H. Gage, "Modelling schizophrenia using human induced pluripotent stem cells.," *Nature*, vol. 473, no. 7346, pp. 221–5, May 2011.

[193] M. S. Lidow, "Calcium signaling dysfunction in schizophrenia: a unifying approach," *Brain Research Reviews*, vol. 43, no. 1, pp. 70–84, Sep. 2003.

[194] M. Poltorak, R. Wright, J. J. Hemperly, E. F. Torrey, F. Issa, R. J. Wyatt, and W. J. Freed, "Monozygotic twins discordant for schizophrenia are discordant for N-CAM and L1 in CSF.," *Brain research*, vol. 751, no. 1, pp. 152–4, Mar. 1997.

[195] J. J. Haddad, "N-methyl-D-aspartate (NMDA) and the regulation of mitogen-activated protein kinase (MAPK) signaling pathways: a revolving neurochemical axis for therapeutic intervention?," *Progress in neurobiology*, vol. 77, no. 4, pp. 252–82, Nov. 2005.

[196] S. Narayan, S. R. Head, T. J. Gilmartin, B. Dean, and E. A. Thomas, “Evidence for disruption of sphingolipid metabolism in schizophrenia.,” *Journal of neuroscience research*, vol. 87, no. 1, pp. 278–88, Jan. 2009.

[197] P. F. Maness and M. Schachner, “Neural recognition molecules of the immunoglobulin superfamily: signaling transducers of axon guidance and neuronal migration.,” *Nature neuroscience*, vol. 10, no. 1, pp. 19–26, Jan. 2007.

[198] F. Gaughran, J. Payne, P. M. Sedgwick, D. Cotter, and M. Berry, “Hippocampal FGF-2 and FGFR1 mRNA expression in major depression, schizophrenia and bipolar disorder.,” *Brain research bulletin*, vol. 70, no. 3, pp. 221–7, Jul. 2006.

[199] T. F. Franke, “PI3K/Akt: getting it right matters.,” *Oncogene*, vol. 27, no. 50, pp. 6473–88, Oct. 2008.

[200] A. J. Law, J. E. Kleinman, D. R. Weinberger, and C. S. Weickert, “Disease-associated intronic variants in the ErbB4 gene are related to altered ErbB4 splice-variant expression in the brain in schizophrenia.,” *Human molecular genetics*, vol. 16, no. 2, pp. 129–41, Jan. 2007.

[201] D. O. Perkins, C. D. Jeffries, L. F. Jarskog, J. M. Thomson, K. Woods, M. A. Newman, J. S. Parker, J. Jin, and S. M. Hammond, “microRNA expression in the prefrontal cortex of individuals with schizophrenia and schizoaffective disorder.,” *Genome biology*, vol. 8, no. 2, p. R27, Jan. 2007.

[202] S. Shen and W. Zhang, “ABC transporters and drug efflux at the blood-brain barrier.,” *Reviews in the neurosciences*, vol. 21, no. 1, pp. 29–53, Jan. 2010.

[203] C. J. Carter, “Schizophrenia susceptibility genes converge on interlinked pathways related to glutamatergic transmission and long-term potentiation, oxidative stress and oligodendrocyte viability.,” *Schizophrenia research*, vol. 86, no. 1–3, pp. 1–14, Sep. 2006.

[204] L. S. Campos, D. P. Leone, J. B. Relvas, C. Brakebusch, R. Fässler, U. Suter, and C. ffrench-Constant, “Beta1 integrins activate a MAPK signalling pathway in neural stem cells that contributes to their maintenance.,” *Development (Cambridge, England)*, vol. 131, no. 14, pp. 3433–44, Jul. 2004.

[205] F. F. Hamdan, J. Gauthier, Y. Araki, D.-T. Lin, Y. Yoshizawa, K. Higashi, A.-R. Park, D. Spiegelman, S. Dobrzeniecka, A. Piton, H. Tomitori, H. Daoud, C. Massicotte, E. Henrion, O. Diallo, M. Shekarabi, C. Marineau, M. Shevell, B.

Maranda, G. Mitchell, A. Nadeau, G. D'Anjou, M. Vanasse, M. Srour, R. G. Lafrenière, P. Drapeau, J. C. Lacaille, E. Kim, J.-R. Lee, K. Igarashi, R. L. Huganir, G. A. Rouleau, and J. L. Michaud, "Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability.," *American journal of human genetics*, vol. 88, no. 3, pp. 306–16, Mar. 2011.

[206] S. Qin, X. Zhao, Y. Pan, J. Liu, G. Feng, J. Fu, J. Bao, Z. Zhang, and L. He, "An association study of the N-methyl-D-aspartate receptor NR1 subunit gene (GRIN1) and NR2B subunit gene (GRIN2B) in schizophrenia with universal DNA microarray.," *European journal of human genetics : EJHG*, vol. 13, no. 7, pp. 807–14, Jul. 2005.

[207] S. Begni, S. Moraschi, S. Bignotti, F. Fumagalli, L. Rillosi, J. Perez, and M. Gennarelli, "Association between the G1001C polymorphism in the GRIN1 gene promoter region and schizophrenia," *Biological Psychiatry*, vol. 53, no. 7, pp. 617–619, Apr. 2003.

[208] A. Georgi, R. A. Jamra, K. Klein, A. W. Villela, J. Schumacher, T. Becker, T. Paul, C. Schmael, S. Höfels, N. Klopp, T. Illig, P. Propping, S. Cichon, M. M. Nöthen, T. G. Schulze, and M. Rietschel, "Possible association between genetic variants at the GRIN1 gene and schizophrenia with lifetime history of depressive symptoms in a German sample.," *Psychiatric genetics*, vol. 17, no. 5, pp. 308–10, Oct. 2007.

[209] X. Zhao, H. Li, Y. Shi, R. Tang, W. Chen, J. Liu, G. Feng, J. Shi, L. Yan, H. Liu, and L. He, "Significant association between the genetic variations in the 5' end of the N-methyl-D-aspartate receptor subunit gene GRIN1 and schizophrenia.," *Biological psychiatry*, vol. 59, no. 8, pp. 747–53, Apr. 2006.

[210] J. Sun, C. Wan, P. Jia, A. H. Fanous, K. S. Kendler, B. P. Riley, and Z. Zhao, "Application of systems biology approach identifies and validates GRB2 as a risk gene for schizophrenia in the Irish Case Control Study of Schizophrenia (ICCSS) sample.," *Schizophrenia research*, vol. 125, no. 2–3, pp. 201–8, Feb. 2011.

[211] O. KADIOGLU, "DETERMINATION OF PERFORMANCE PARAMETERS FOR AHP BASED SINGLE NUCLEOTIDE POLYMORPHISM (SNP) PRIORITIZATION APPROACH ON ALZHEIMERS'S DISEASE DATA," METU, 2011.

[212] S. J. Furney, A. Simmons, G. Breen, I. Pedroso, K. Lunnon, P. Proitsi, A. Hodges, J. Powell, L.-O. Wahlund, I. Kloszewska, P. Mecocci, H. Soininen, M. Tsolaki, B. Vellas, C. Spenger, M. Lathrop, L. Shen, S. Kim, A. J. Saykin, M. W. Weiner, and S. Lovestone, "Genome-wide association with MRI atrophy measures as a quantitative trait locus for Alzheimer's disease.," *Molecular psychiatry*, vol. 16, no. 11, pp. 1130–8, Nov. 2011.

[213] M. I. Rees, K. Harvey, B. R. Pearce, S.-K. Chung, I. C. Duguid, P. Thomas, S. Beatty, G. E. Graham, L. Armstrong, R. Shiang, K. J. Abbott, S. M. Zuberi, J. B. P. Stephenson, M. J. Owen, M. A. J. Tijssen, A. M. J. M. van den Maagdenberg, T. G. Smart, S. Supplisson, and R. J. Harvey, "Mutations in the gene encoding GlyT2 (SLC6A5) define a presynaptic component of human startle disease.," *Nature genetics*, vol. 38, no. 7, pp. 801–6, Jul. 2006.

[214] M. W. Logue, M. Schu, B. N. Vardarajan, J. Buros, R. C. Green, R. C. P. Go, P. Griffith, T. O. Obisesan, R. Shatz, A. Borenstein, L. A. Cupples, K. L. Lunetta, M. D. Fallin, C. T. Baldwin, and L. A. Farrer, "A comprehensive genetic association study of Alzheimer disease in African Americans.," *Archives of neurology*, vol. 68, no. 12, pp. 1569–79, Dec. 2011.

[215] M. Poot, V. Beyer, I. Schwaab, N. Damatova, R. Van't Slot, J. Prothero, S. E. Holder, and T. Haaf, "Disruption of CNTNAP2 and additional structural genome changes in a boy with speech delay and autism spectrum disorder.," *Neurogenetics*, vol. 11, no. 1, pp. 81–9, Feb. 2010.

[216] K.-S. Wang, X.-F. Liu, and N. Aragam, "A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder.," *Schizophrenia research*, vol. 124, no. 1–3, pp. 192–9, Dec. 2010.

[217] J. K. J. Van Houdt, B. A. Nowakowska, S. B. Sousa, B. D. C. van Schaik, E. Seuntjens, N. Avonce, A. Sifrim, O. A. Abdul-Rahman, M.-J. H. van den Boogaard, A. Bottani, M. Castori, V. Cormier-Daire, M. A. Deardorff, I. Filges, A. Fryer, J.-P. Fryns, S. Gana, L. Garavelli, G. Gillessen-Kaesbach, B. D. Hall, D. Horn, D. Huylebroeck, J. Klapecki, M. Krajewska-Walasek, A. Kuechler, M. A. Lines, S. Maas, K. D. Macdermot, S. McKee, A. Magee, S. A. de Man, Y. Moreau, F. Morice-Picard, E. Obersztyn, J. Pilch, E. Rosser, N. Shannon, I. Stolte-Dijkstra, P. Van Dijck, C. Vilain, A. Vogels, E. Wakeling, D. Wieczorek, L. Wilson, O. Zuffardi, A. H. C. van Kampen, K. Devriendt, R. Hennekam, and J. R. Vermeesch, "Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome.," *Nature genetics*, vol. 44, no. 4, pp. 445–9, S1, Apr. 2012.

[218] J. Huang, R. H. Perlis, P. H. Lee, A. J. Rush, M. Fava, G. S. Sachs, J. Lieberman, S. P. Hamilton, P. Sullivan, P. Sklar, S. Purcell, and J. W. Smoller, "Cross-disorder genomewide analysis of schizophrenia, bipolar disorder, and depression.," *The American journal of psychiatry*, vol. 167, no. 10, pp. 1254–63, Oct. 2010.

[219] M. A. R. Ferreira, M. C. O'Donovan, Y. A. Meng, I. R. Jones, D. M. Ruderfer, L. Jones, J. Fan, G. Kirov, R. H. Perlis, E. K. Green, J. W. Smoller, D. Grozeva, J. Stone, I. Nikolov, K. Chambert, M. L. Hamshere, V. L. Nimgaonkar, V. Moskvina, M. E. Thase, S. Caesar, G. S. Sachs, J. Franklin, K. Gordon-Smith, K. G. Ardlie, S. B. Gabriel, C. Fraser, B. Blumenstiel, M. Defelice, G. Breen, M. Gill, D. W. Morris, A. Elkin, W. J. Muir, K. A. McGhee, R. Williamson, D. J. MacIntyre, A. W. MacLean, C. D. St, M. Robinson, M. Van Beck, A. C. P. Pereira, R. Kandaswamy, A. McQuillin, D. A. Collier, N. J. Bass, A. H. Young, J. Lawrence, I. N. Ferrier, A. Anjorin, A. Farmer, D. Curtis, E. M. Scolnick, P. McGuffin, M. J. Daly, A. P. Corvin, P. A. Holmans, D. H. Blackwood, H. M. Gurling, M. J. Owen, S. M. Purcell, P. Sklar, and N. Craddock, "Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder.," *Nature genetics*, vol. 40, no. 9, pp. 1056–8, Sep. 2008.

[220] V. M. Goghari, D. J. Lang, S. W. Flynn, A. L. Mackay, and W. G. Honer, "Smaller corpus callosum subregions containing motor fibers in schizophrenia.," *Schizophrenia research*, vol. 73, no. 1, pp. 59–68, Feb. 2005.

[221] E. Zaidel and M. Iacoboni, *The Parallel Brain: The Cognitive Neuroscience of the Corpus Callosum*. 2003.

[222] S. F. WITELSON, "HAND AND SEX DIFFERENCES IN THE ISTHMUS AND GENU OF THE HUMAN CORPUS CALLOSUM," *Brain*, vol. 112, no. 3, pp. 799–835, Jun. 1989.

[223] C. N. Vidal, R. Nicolson, T. J. DeVito, K. M. Hayashi, J. A. Geaga, D. J. Drost, P. C. Williamson, N. Rajakumar, Y. Sui, R. A. Dutton, A. W. Toga, and P. M. Thompson, "Mapping Corpus Callosum Deficits in Autism: An Index of Aberrant Cortical Connectivity."

[224] S. A. Mitelman, Y. K. Nikiforova, E. L. Canfield, E. A. Hazlett, A. M. Brickman, L. Shihabuddin, and M. S. Buchsbaum, "A longitudinal study of the corpus callosum in chronic schizophrenia.," *Schizophrenia research*, vol. 114, no. 1–3, pp. 144–53, Oct. 2009.

[225] R. J. Webster, K. M. Giles, K. J. Price, P. M. Zhang, J. S. Mattick, and P. J. Leedman, "Regulation of epidermal growth factor receptor signaling in human cancer cells by microRNA-7.," *The Journal of biological chemistry*, vol. 284, no. 9, pp. 5731–41, Mar. 2009.

[226] H. Guo, D. Liu, H. Gelbard, T. Cheng, R. Insalaco, J. A. Fernández, J. H. Griffin, and B. V Zlokovic, "Activated Protein C Prevents Neuronal Apoptosis via Protease Activated Receptors 1 and 3," *Neuron*, vol. 41, no. 4, pp. 563–572, Feb. 2004.

[227] J. Y. Ban, H. C. Youn, H.-K. Park, G. Gwak, and B. S. Kim, "Gene expression profiles regulated by Hspa1b in MPTP-induced dopaminergic neurotoxicity using knockout mice," *Molecular & Cellular Toxicology*, vol. 8, no. 3, pp. 281–287, Oct. 2012.

[228] J. L. Spence and S. Wallihan, "Computational prediction of the PolyQ and CAG repeat spinocerebellar ataxia network based on sequence identity to untranslated regions.," *Gene*, vol. 509, no. 2, pp. 273–81, Nov. 2012.

[229] M. A. Kurian, E. Meyer, G. Vassallo, N. V Morgan, N. Prakash, S. Pasha, N. A. Hai, S. Shuib, F. Rahman, E. Wassmer, J. H. Cross, F. J. O'Callaghan, J. P. Osborne, I. E. Scheffer, P. Gissen, and E. R. Maher, "Phospholipase C beta 1 deficiency is associated with early-onset epileptic encephalopathy.," *Brain : a journal of neurology*, vol. 133, no. 10, pp. 2964–70, Oct. 2010.

[230] Y. Liu, D. H. Blackwood, S. Caesar, E. J. C. de Geus, A. Farmer, M. A. R. Ferreira, I. N. Ferrier, C. Fraser, K. Gordon-Smith, E. K. Green, D. Grozeva, H. M. Gurling, M. L. Hamshere, P. Heutink, P. A. Holmans, W. J. Hoogendijk, J. J. Hottenga, L. Jones, I. R. Jones, G. Kirov, D. Lin, P. McGuffin, V. Moskvina, W. A. Nolen, R. H. Perlis, D. Posthuma, E. M. Scolnick, A. B. Smit, J. H. Smit, J. W. Smoller, D. St Clair, R. van Dyck, M. Verhage, G. Willemsen, A. H. Young, T. Zandbelt, D. I. Boomsma, N. Craddock, M. C. O'Donovan, M. J. Owen, B. W. J. H. Penninx, S. Purcell, P. Sklar, and P. F. Sullivan, "Meta-analysis of genome-wide association data of bipolar disorder and major depressive disorder.," *Molecular psychiatry*, vol. 16, no. 1, pp. 2–4, Jan. 2011.

[231] "Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4.," *Nature genetics*, vol. 43, no. 10, pp. 977–83, Oct. 2011.

[232] S. E. Bergen, C. T. O'Dushlaine, S. Ripke, P. H. Lee, D. M. Ruderfer, S. Akterin, J. L. Moran, K. D. Chambert, R. E. Handsaker, L. Backlund, U. Ösby, S. McCarroll, M. Landen, E. M. Scolnick, P. K. E. Magnusson, P. Lichtenstein, C. M. Hultman, S. M. Purcell, P. Sklar, and P. F. Sullivan, "Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder.," *Molecular psychiatry*, vol. 17, no. 9, pp. 880–6, Sep. 2012.

[233] F. Petrij, R. H. Giles, H. G. Dauwerse, J. J. Saris, R. C. Hennekam, M. Masuno, N. Tommerup, G. J. van Ommen, R. H. Goodman, and D. J. Peters, "Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP.," *Nature*, vol. 376, no. 6538, pp. 348–51, Jul. 1995.

APPENDICES

4.3. APPENDIX A: TOP100 SNPs FROM PLINK RESULTS

Rank	SNP	P-Val	Rank	SNP	P-Val
1	rs2574608	4.16E-20	**51**	rs11995470	1.40E-12
2	rs2352301	8.30E-20	**52**	rs1927887	1.42E-12
3	rs2771913	8.78E-20	**53**	rs6713559	2.18E-12
4	rs2563850	1.08E-19	**54**	rs10239811	2.40E-12
5	rs2123561	1.36E-19	**55**	rs6711688	3.59E-12
6	rs2562967	1.43E-19	**56**	rs7214631	3.67E-12
7	rs2558989	1.43E-19	**57**	rs9604109	3.74E-12
8	rs425231	1.57E-19	**58**	rs2426553	4.37E-12
9	rs2557227	1.67E-19	**59**	rs1458528	4.49E-12
10	rs2755895	2.01E-19	**60**	rs12068752	4.50E-12
11	rs2750987	2.12E-19	**61**	rs7404615	4.72E-12
12	rs2563654	2.45E-19	**62**	rs884562	5.30E-12
13	rs2759914	2.45E-19	**63**	rs2146327	7.02E-12
14	rs1435898	2.65E-19	**64**	rs2558593	7.19E-12
15	rs2352691	3.08E-19	**65**	rs2904311	7.34E-12
16	rs1435910	4.35E-19	**66**	rs8078633	7.46E-12
17	rs2752696	5.25E-19	**67**	rs9514925	7.62E-12
18	rs2522781	1.26E-18	**68**	rs2318050	8.30E-12
19	rs2880301	1.47E-18	**69**	rs2734743	1.10E-11
20	rs2557030	1.69E-18	**70**	rs6735695	1.12E-11
21	rs2441262	2.86E-18	**71**	rs7055488	1.13E-11
22	rs2653037	4.94E-18	**72**	rs7052126	1.14E-11
23	rs2534048	7.05E-18	**73**	rs1566060	1.15E-11
24	rs2534362	1.14E-17	**74**	rs11984987	1.19E-11
25	rs3865868	1.25E-17	**75**	rs10000609	1.23E-11
26	rs2197554	1.34E-17	**76**	rs7652014	1.24E-11

27	rs2559142	1.41E-17	**77**	rs2333645	1.25E-11
28	rs1435909	2.38E-17	**78**	rs1680499	1.26E-11
29	rs3869940	8.38E-17	**79**	rs4662425	1.27E-11
30	rs12741415	1.03E-16	**80**	rs2524632	1.40E-11
31	rs3883014	1.92E-16	**81**	rs2795549	1.44E-11
32	rs2198374	2.67E-16	**82**	rs5964897	1.47E-11
33	rs2578818	2.96E-16	**83**	rs11880554	1.60E-11
34	rs17042395	3.88E-16	**84**	rs2352134	1.63E-11
35	rs11967088	1.37E-14	**85**	rs2003397	1.63E-11
36	rs1778596	1.43E-14	**86**	rs6625960	1.65E-11
37	rs2557661	2.91E-14	**87**	rs7231185	1.66E-11
38	rs7768749	8.16E-14	**88**	rs7748724	1.68E-11
39	rs6478322	1.31E-13	**89**	rs6599101	1.68E-11
40	rs2755680	1.31E-13	**90**	rs7993976	1.71E-11
41	rs7859338	1.39E-13	**91**	rs2563034	1.75E-11
42	rs2556984	1.58E-13	**92**	rs7711172	1.84E-11
43	rs7638929	2.73E-13	**93**	rs7047325	1.92E-11
44	rs6916467	3.03E-13	**94**	rs4448540	1.94E-11
45	rs6618424	4.84E-13	**95**	rs10830849	1.95E-11
46	rs9502759	5.14E-13	**96**	rs10790299	1.96E-11
47	rs10086065	9.98E-13	**97**	rs12070887	1.99E-11
48	rs10125239	1.00E-12	**98**	rs7805995	2.04E-11
49	rs17146088	1.09E-12	**99**	rs12120383	2.06E-11
50	rs17837718	1.22E-12	**100**	rs8075550	2.08E-11

4.4. APPENDIX B: DAVID FUNCTIONAL ANNOTATION CHART OF PLINK RESULTS (TOP100)

Category	Term	Count	%	PValue
KEGG	Glycosphingolipid biosynthesis	9	0.325027086	5.45E-04
KEGG	Nucleotide excision repair	16	0.57782593	0.001082709
KEGG	Homologous recombination	11	0.397255327	0.004980238
KEGG	Calcium signaling pathway	39	1.408450704	0.008868431
KEGG	Axon guidance	30	1.083423619	0.012019383
KEGG	Vascular smooth muscle contraction	26	0.938967136	0.020517533
KEGG	Other glycan degradation	7	0.252798844	0.02200715
BIOCARTA	Aspirin Blocks Signaling Pathway Involved in Platelet Activation	8	0.288912965	0.022069069
PANTHER	5HT2 type receptor mediated signaling pathway	18	0.650054171	0.022701283
KEGG	Cell adhesion molecules (CAMs)	29	1.047309498	0.028156905
KEGG	MAPK signaling pathway	52	1.877934272	0.029010801
BIOCARTA	Thrombin signaling and protease-activated receptors	7	0.252798844	0.041384851
KEGG	Mismatch repair	8	0.288912965	0.043203185
PANTHER	Histamine H1 receptor mediated signaling pathway	12	0.433369447	0.044676939
KEGG	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	18	0.650054171	0.049384564
KEGG	ABC transporters	12	0.433369447	0.051566312
PANTHER	5HT4 type receptor mediated signaling pathway	10	0.361141206	0.052613823
PANTHER	Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	31	1.119537739	0.054580545
PANTHER	Oxytocin receptor mediated signaling pathway	15	0.541711809	0.054877268
KEGG	Lysosome	25	0.902853016	0.056386669
PANTHER	p38 MAPK pathway	14	0.505597689	0.064659838
BIOCARTA	p38 MAPK Signaling Pathway	10	0.361141206	0.067632814
PANTHER	Alzheimer disease-amyloid secretase pathway	17	0.613940051	0.070684189
KEGG	DNA replication	10	0.361141206	0.07431895
PANTHER	EGF receptor signaling pathway	29	1.047309498	0.076829584
BIOCARTA	ALK in cardiac myocytes	10	0.361141206	0.079671813
PANTHER	Inflammation mediated by chemokine and	56	2.022390755	0.084656409

	cytokine signaling pathway			
KEGG	Tight junction	27	0.975081257	0.08490533
PANTHER	Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway	34	1.227880101	0.091847131
BIOCARTA	Bioactive Peptide Induced Signaling Pathway	9	0.325027086	0.092442178
KEGG	Non-small cell lung cancer	13	0.469483568	0.093015077

4.5. APPENDIX C: REACTOME ANALYSIS OF PLINK RESULTS (TOP100)

P-value	Number of genes	Total number of genes	Name of the event
5.45E-10	8	13	FGFR4 binds to FGF
5.45E-10	8	13	FGFR4 ligand binding and activation
5.45E-10	8	13	Autocatalytic phosphorylation of FGFR4
7.75E-09	9	23	FGFR ligand binding and activation
1.21E-08	9	24	Phosphorylation of FRS2-alpha by activated FGFR
1.21E-08	9	24	Activated FGFR binds PLC-gamma
1.21E-08	9	24	PLC-gamma phosphorylation by FGFR
1.21E-08	9	24	Activated PLC gamma release by activated FGFR
1.21E-08	9	24	Phosphorylation of FRS2-beta by activated FGFR
1.21E-08	9	24	Activated FGFR binds FRS2alpha
1.21E-08	9	24	Activated FGFR binds FRS2beta
1.21E-08	9	24	Activated FGFR recruits SHC1
1.21E-08	9	24	SHC1 is phosphorylated
2.74E-08	9	26	Activated FGFR and FRS2 bind to SHP2
2.74E-08	9	26	SHP2 is phosphorylated by activated FGFR
2.74E-08	9	26	GRB2:SOS1 is recruited to activated FGFR through p-SHC1
2.74E-08	9	26	Activated ERK1/2 threonine-phosphorylate FRS2alpha.
4.01E-08	9	27	GRB2:GAB1:PI3Kreg binds directly to p-FRS2:activated FGFR
4.01E-08	9	27	GRB2:SOS1 binds to p-FRS2:activated FGFR
4.01E-08	9	27	p-CBL:GRB2 binds p-FRS2alpha:activated FGFR
5.76E-08	9	28	GRB2:GAB1:PI3Kreg binds to p-SHP2 on p-FRS2:activated FGFR
5.76E-08	9	28	FGFR-associated PI3K phosphorylates PIP2 to PIP3
5.76E-08	9	28	PI3K catalytic subunit is recruited by FGFR-associated PI3KR1
7.04E-08	8	21	Activated FGFR mutants bind FRS2alpha
7.04E-08	8	21	FGFR mutants phosphorylate FRS2alpha
7.04E-08	8	21	Activated FGFR mutants bind PLC-gamma

7.04E-08	8	21	PLC-gamma phosphorylation by FGFR mutants
7.04E-08	8	21	Activated PLC-gamma release by activated FGFR mutants
8.15E-08	9	29	SHC-mediated cascade
8.15E-08	9	29	PI3K catalytic subunit binds to GRB2:GAB1:PI3Kreg (indirect)
8.15E-08	9	29	Ras nucleotide exchange by GRB2:SOS1 through p-SHC
8.15E-08	9	29	FGFR associated PI3K phosphorylates PIP2 to PIP3
1.13E-07	9	30	Ras nucleotide exchange by GRB2:SOS1 bound to p-FRS2:activated FGFR
1.56E-07	9	31	CBL ubiquitinates FRS2 and FGFR
1.61E-07	8	23	GRB2-SOS1 is recruited by activated FGFR mutants
2.36E-07	8	24	GRB2:GAB1:PI3Kreg binds to p-FRS2:activated FGFR mutants
2.57E-07	7	17	FGFR2 ligand binding and activation
2.57E-07	7	17	Signaling by FGFR2 mutants
2.57E-07	7	17	Activated point mutants of FGFR2
2.57E-07	7	17	Point mutants of FGFR2 bind and are inactivated by tyrosine kinase inhibitors
3.38E-07	8	25	Recruitment of the catalytic subunit of PI3K by activated FGFR1 mutants
3.38E-07	8	25	Conversion of PIP2 to PIP3 by PI3K associated with FGFR mutants
4.74E-07	6	12	FGFR3 ligand binding and activation
4.74E-07	6	12	FGFR3c binds to FGF
4.74E-07	6	12	FGFR3c ligand binding and activation
4.74E-07	6	12	Autocatalytic phosphorylation of FGFR3c
4.74E-07	6	12	Signaling by activated point mutants of FGFR3
4.74E-07	6	12	Autocatalytic phosphorylation of FGFR3c P250R mutant
4.74E-07	6	12	FGFR3c P250R mutant binds to ligand with enhanced affinity
4.74E-07	6	12	Signaling by FGFR3 mutants
4.74E-07	6	12	FGFR3 mutants bind and are inactivated by tyrosine kinase inhibitors
4.76E-07	8	26	Ras nucleotide exchange mediated by GRB2-SOS1 bound to FGFR mutants
8.25E-07	9	37	PIP2 conversion to PIP3
8.60E-07	6	13	FGFR2c binds to FGF
8.60E-07	6	13	FGFR2c ligand binding and activation
8.60E-07	6	13	Autocatalytic phosphorylation of FGFR2c
1.05E-06	9	38	FRS2-mediated cascade
1.22E-06	11	61	PI3K phosphorylates PIP2 to PIP3

1.22E-06	11	61	PI3K inhibitors block PI3K catalytic activity
1.47E-06	6	14	FGFR1 ligand binding and activation
1.68E-06	9	40	Negative regulation of FGFR signaling
2.39E-06	6	15	FGFR2c mutants bind an expanded range of ligands
2.39E-06	6	15	Autocatalytic phosphorylation of FGFR2c mutants with enhanced ligand binding
8.31E-06	5	11	FGFR1c binds to FGF
8.31E-06	5	11	FGFR1c ligand binding and activation
8.31E-06	5	11	Autocatalytic phosphorylation of FGFR1c
8.31E-06	5	11	Signaling by activated point mutants of FGFR1
8.31E-06	5	11	FGFR1 P252X mutants bind ligand with enhanced affinity
8.31E-06	5	11	Autocatalytic phosphorylation of FGFR1 P252X mutant dimers
2.29E-05	9	54	Phospholipase C-mediated cascade
2.36E-05	4	7	FGFR3b binds to FGF
2.36E-05	4	7	FGFR3b ligand binding and activation
2.36E-05	4	7	Autocatalytic phosphorylation of FGFR3b
2.50E-05	3	3	Interaction of integrin alpha8beta1 with Fibronectin
2.50E-05	3	3	Interaction of integrin alphaVbeta1 with Fibronectin
3.38E-05	11	85	Integrin cell surface interactions
3.54E-05	12	101	PI-3K cascade
3.54E-05	12	101	PI3K events in ERBB4 signaling
3.54E-05	12	101	PIP3 activates AKT signaling
3.54E-05	12	101	PI3K events in ERBB2 signaling
3.54E-05	12	101	PI3K/AKT Signaling in Cancer
3.77E-05	11	86	Constitutive PI3K/AKT Signaling in Cancer
4.07E-05	8	45	Signaling by FGFR mutants
4.61E-05	4	8	Latent TGF-beta-1 binds integrins
4.75E-05	12	104	PI3K/AKT activation
5.23E-05	12	105	GAB1 signalosome
9.77E-05	3	4	CARD9 binds RIP2 (and NOD2)
9.77E-05	3	4	Carbonic Anhydrase Dehydrates Bicarbonate (plasma membrane)
9.77E-05	3	4	Carbonic Anhydrase Hydrates Carbon Dioxide (plasma membrane)
1.43E-04	10	83	IRS-related events triggered by IGF1R
1.49E-04	9	68	PI3K Cascade
1.93E-04	10	86	Signaling by Type 1 Insulin-like Growth Factor 1 Receptor (IGF1R)

1.93E-04	10	86	IGF1R signaling cascade
2.20E-04	13	140	Signaling by SCF-KIT
2.97E-04	4	12	Reversible Hydration of Carbon Dioxide
3.28E-04	15	185	Signaling by PDGF
4.68E-04	3	6	Klotho-mediated ligand binding
4.70E-04	14	171	Downstream Signaling Events Of B Cell Receptor (BCR)
4.70E-04	9	79	IRS-mediated signalling
4.74E-04	6	35	NOD1/2 Signaling Pathway

4.6. APPENDIX D: TOP 100 GENES FROM COMBINED GENE RESULTS

Rank	Gene ID	P-Val	Rank	Gene ID	P-Val
1	55277	0	**51**	136	0
2	3604	0	**52**	132	0
3	3603	0	**53**	8406	0
4	220963	0	**54**	54768	0
5	337876	0	**55**	4045	0
6	132884	0	**56**	120	0
7	283652	0	**57**	114134	0
8	8499	0	**58**	54756	0
9	8496	0	**59**	4036	0
10	8495	0	**60**	4033	0
11	55222	0	**61**	119	0
12	8491	0	**62**	115	0
13	120892	0	**63**	114	0
14	23677	0	**64**	111	0
15	441116	0	**65**	151963	0
16	23671	0	**66**	4026	0
17	23670	0	**67**	4025	0
18	140469	0	**68**	109	0
19	55214	0	**69**	108	0
20	55213	0	**70**	105	0
21	144817	0	**71**	104	0
22	199920	0	**72**	102	0
23	55206	0	**73**	145282	0
24	8470	0	**74**	4009	0
25	8464	0	**75**	4008	0
26	23648	0	**76**	7991	0
27	185	0	**77**	340595	0
28	152404	0	**78**	149628	0
29	283601	0	**79**	54715	0

30	8451	0	**80**	54714	0
31	176	0	**81**	7982	0
32	175	0	**82**	93492	0
33	23637	0	**83**	619343	0
34	23635	0	**84**	284058	0
35	4086	0	**85**	151903	0
36	162	0	**86**	284040	0
37	54798	0	**87**	221322	0
38	54796	0	**88**	340554	0
39	157	0	**89**	3572	0
40	23613	0	**90**	3556	0
41	8427	0	**91**	121256	0
42	8424	0	**92**	441097	0
43	4067	0	**93**	55193	0
44	4065	0	**94**	121227	0
45	148	0	**95**	55182	0
46	23608	0	**96**	79109	0
47	23604	0	**97**	441061	0
48	8412	0	**98**	55160	0
49	4057	0	**99**	646021	0
50	4052	0	**100**	55140	0

4.7. APPENDIX E: DAVID FUNCTIONAL ANNOTATION CHART OF COMBINED GENE RESULTS (TOTAL 78)

Category	Term	Count	%	PValue
KEGG	Axon guidance	49	1.668369084	3.93E-09
KEGG	Calcium signaling pathway	56	1.906707525	3.22E-07
KEGG	Long-term depression	29	0.987402111	9.62E-07
KEGG	Vascular smooth muscle contraction	39	1.327885598	2.49E-06
KEGG	Focal adhesion	59	2.008852571	2.88E-06
KEGG	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	29	0.987402111	9.07E-06
KEGG	ECM-receptor interaction	30	1.02145046	2.67E-05
KEGG	Pathways in cancer	82	2.79196459	2.68E-05
PANTHER	Metabotropic glutamate receptor group III pathway	32	1.089547157	4.50E-05
KEGG	ABC transporters	19	0.646918624	7.52E-05
KEGG	MAPK signaling pathway	65	2.213142663	4.70E-04
KEGG	Small cell lung cancer	27	0.919305414	5.10E-04
PANTHER	Integrin signalling pathway	65	2.213142663	5.49E-04
PANTHER	Histamine H1 receptor mediated signaling pathway	19	0.646918624	6.32E-04
KEGG	Long-term potentiation	23	0.783112019	6.94E-04
PANTHER	Thyrotropin-releasing hormone receptor signaling pathway	24	0.817160368	8.00E-04
PANTHER	5HT2 type receptor mediated signaling pathway	26	0.885257065	9.85E-04
KEGG	Phosphatidylinositol signaling system	24	0.817160368	9.91E-04
KEGG	Dilated cardiomyopathy	28	0.953353762	1.01E-03
KEGG	Gap junction	27	0.919305414	1.34E-03
BIOCARTA	Synaptic Proteins at the Synaptic Junction	10	0.340483487	1.48E-03
PANTHER	Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	46	1.566224038	1.55E-03
PANTHER	Oxytocin receptor mediated signaling pathway	23	0.783112019	1.59E-03
PANTHER	Ionotropic glutamate receptor pathway	22	0.74906367	1.86E-03
KEGG	ErbB signaling pathway	26	0.885257065	2.12E-03
BIOCARTA	Aspirin Blocks Signaling Pathway Involved in Platelet Activation	10	0.340483487	2.40E-03
PANTHER	Endogenous_cannabinoid_signaling	13	0.442628533	2.40E-03

PANTHER	Axon guidance mediated by Slit/Robo	15	0.51072523	2.83E-03
BIOCARTA	PKC-catalyzed phosphorylation of inhibitory phosphoprotein of myosin phosphatase	8	0.272386789	3.03E-03
KEGG	Hypertrophic cardiomyopathy (HCM)	25	0.851208716	3.32E-03
PANTHER	Metabotropic glutamate receptor group I pathway	17	0.578821927	4.01E-03
KEGG	O-Glycan biosynthesis	12	0.408580184	5.24E-03
PANTHER	Muscarinic acetylcholine receptor 1 and 3 signaling pathway	22	0.74906367	6.07E-03
PANTHER	Axon guidance mediated by netrin	16	0.544773578	6.19E-03
KEGG	Regulation of actin cytoskeleton	50	1.702417433	6.23E-03
KEGG	Prostate cancer	25	0.851208716	6.28E-03
PANTHER	Alpha adrenergic receptor signaling pathway	13	0.442628533	7.13E-03
KEGG	Glioma	19	0.646918624	9.03E-03
KEGG	Fc gamma R-mediated phagocytosis	25	0.851208716	1.45E-02
PANTHER	GABA-B_receptor_II_signaling	15	0.51072523	1.58E-02
BIOCARTA	Thrombin signaling and protease-activated receptors	8	0.272386789	1.82E-02
BIOCARTA	Role of ERBB2 in Signal Transduction and Oncology	9	0.306435138	1.96E-02
KEGG	Neuroactive ligand-receptor interaction	55	1.872659176	2.02E-02
KEGG	Fc epsilon RI signaling pathway	21	0.715015322	2.09E-02
KEGG	Non-small cell lung cancer	16	0.544773578	2.12E-02
KEGG	Tight junction	32	1.089547157	2.15E-02
PANTHER	Angiotensin II-stimulated signaling through G proteins and beta-arrestin	18	0.612870276	2.21E-02
KEGG	Melanogenesis	25	0.851208716	2.37E-02
KEGG	B cell receptor signaling pathway	20	0.680966973	2.70E-02
PANTHER	Endothelin signaling pathway	28	0.953353762	2.77E-02
PANTHER	Opioid proopiomelanocortin pathway	13	0.442628533	2.83E-02
PANTHER	Beta1 adrenergic receptor signaling pathway	16	0.544773578	3.13E-02
BIOCARTA	Nuclear Receptors in Lipid Metabolism and Toxicity	12	0.408580184	3.69E-02
KEGG	GnRH signaling pathway	24	0.817160368	3.75E-02
PANTHER	Alzheimer disease-amyloid secretase pathway	22	0.74906367	4.17E-02
KEGG	Nucleotide excision repair	13	0.442628533	4.24E-02
PANTHER	Cortocotropin releasing factor receptor signaling pathway	12	0.408580184	4.24E-02
KEGG	Inositol phosphate metabolism	15	0.51072523	4.44E-02
PANTHER	Metabotropic glutamate receptor group II pathway	17	0.578821927	4.49E-02
PANTHER	Angiogenesis	56	1.906707525	4.72E-02
KEGG	Cell adhesion molecules (CAMs)	30	1.02145046	4.77E-02

BIOCARTA	Erk1/Erk2 Mapk Signaling pathway	10	0.340483487	5.18E-02
PANTHER	5HT4 type receptor mediated signaling pathway	12	0.408580184	5.24E-02
PANTHER	Opioid proenkephalin pathway	12	0.408580184	5.24E-02
PANTHER	Beta2 adrenergic receptor signaling pathway	15	0.51072523	5.38E-02
BIOCARTA	Role of PPAR-gamma Coactivators in Obesity and Thermogenesis	5	0.170241743	5.53E-02
KEGG	Melanoma	18	0.612870276	5.72E-02
KEGG	Adherens junction	19	0.646918624	6.29E-02
BIOCARTA	Bioactive Peptide Induced Signaling Pathway	10	0.340483487	6.31E-02
BIOCARTA	g-Secretase mediated ErbB4 Signaling Pathway	4	0.136193395	6.99E-02
KEGG	Purine metabolism	33	1.123595506	6.99E-02
PANTHER	B cell activation	24	0.817160368	7.39E-02
BIOCARTA	Eph Kinases and ephrins support platelet aggregation	5	0.170241743	7.99E-02
KEGG	Thyroid cancer	9	0.306435138	8.39E-02
KEGG	Fructose and mannose metabolism	10	0.340483487	8.62E-02
PANTHER	Nicotinic acetylcholine receptor signaling pathway	24	0.817160368	9.19E-02
PANTHER	Beta3 adrenergic receptor signaling pathway	10	0.340483487	9.27E-02

4.8. APPENDIX F: REACTOME ANALYSIS OF COMBINED GENE RESULTS (TOP100)

P-value	Number of genes	Total number of genes	Name of the event
3.16E-10	11	23	GRB2-SOS1 is recruited by activated FGFR mutants
1.62E-09	11	26	GRB2:SOS1 is recruited to activated FGFR through p-SHC1
1.62E-09	11	26	Ras nucleotide exchange mediated by GRB2-SOS1 bound to FGFR mutants
2.21E-09	10	21	Activated FGFR mutants bind FRS2alpha
2.21E-09	10	21	FGFR mutants phosphorylate FRS2alpha
2.21E-09	10	21	Activated FGFR mutants bind PLC-gamma
2.21E-09	10	21	PLC-gamma phosphorylation by FGFR mutants
2.21E-09	10	21	Activated PLC-gamma release by activated FGFR mutants
2.64E-09	11	27	GRB2:SOS1 binds to p-FRS2:activated FGFR
4.38E-09	9	17	FGFR2 ligand binding and activation
4.38E-09	9	17	Signaling by FGFR2 mutants
4.38E-09	9	17	Activated point mutants of FGFR2
4.38E-09	9	17	Point mutants of FGFR2 bind and are inactivated by tyrosine kinase inhibitors
6.52E-09	11	29	SHC-mediated cascade
6.52E-09	11	29	Ras nucleotide exchange by GRB2:SOS1 through p-SHC
6.62E-09	8	13	FGFR4 binds to FGF
6.62E-09	8	13	FGFR4 ligand binding and activation
6.62E-09	8	13	Autocatalytic phosphorylation of FGFR4
6.68E-09	10	23	FGFR ligand binding and activation
9.93E-09	11	30	Ras nucleotide exchange by GRB2:SOS1 bound to p-FRS2:activated FGFR
1.10E-08	10	24	Phosphorylation of FRS2-alpha by activated FGFR
1.10E-08	10	24	Activated FGFR binds PLC-gamma
1.10E-08	10	24	PLC-gamma phosphorylation by FGFR
1.10E-08	10	24	Activated PLC gamma release by activated FGFR
1.10E-08	10	24	Phosphorylation of FRS2-beta by activated FGFR

1.10E-08	10	24	Activated FGFR binds FRS2alpha
1.10E-08	10	24	Activated FGFR binds FRS2beta
1.10E-08	10	24	Activated FGFR recruits SHC1
1.10E-08	10	24	SHC1 is phosphorylated
1.10E-08	10	24	GRB2:GAB1:PI3Kreg binds to p-FRS2:activated FGFR mutants
1.49E-08	8	14	FGFR1 ligand binding and activation
1.78E-08	10	25	Recruitment of the catalytic subunit of PI3K by activated FGFR1 mutants
1.78E-08	10	25	Conversion of PIP2 to PIP3 by PI3K associated with FGFR mutants
2.79E-08	10	26	Activated FGFR and FRS2 bind to SHP2
2.79E-08	10	26	SHP2 is phosphorylated by activated FGFR
2.79E-08	10	26	Activated ERK1/2 threonine-phosphorylate FRS2alpha.
4.27E-08	10	27	GRB2:GAB1:PI3Kreg binds directly to p-FRS2:activated FGFR
4.27E-08	10	27	p-CBL:GRB2 binds p-FRS2alpha:activated FGFR
6.41E-08	10	28	GRB2:GAB1:PI3Kreg binds to p-SHP2 on p-FRS2:activated FGFR
6.41E-08	10	28	FGFR-associated PI3K phosphorylates PIP2 to PIP3
6.41E-08	10	28	PI3K catalytic subunit is recruited by FGFR-associated PI3KR1
9.44E-08	10	29	PI3K catalytic subunit binds to GRB2:GAB1:PI3Kreg (indirect)
9.44E-08	10	29	FGFR associated PI3K phosphorylates PIP2 to PIP3
1.04E-07	7	12	FGFR3 ligand binding and activation
1.04E-07	7	12	FGFR3c binds to FGF
1.04E-07	7	12	FGFR3c ligand binding and activation
1.04E-07	7	12	Autocatalytic phosphorylation of FGFR3c
1.04E-07	7	12	Signaling by activated point mutants of FGFR3
1.04E-07	7	12	Autocatalytic phosphorylation of FGFR3c P250R mutant
1.04E-07	7	12	FGFR3c P250R mutant binds to ligand with enhanced affinity
1.04E-07	7	12	Signaling by FGFR3 mutants
1.04E-07	7	12	FGFR3 mutants bind and are inactivated by tyrosine kinase inhibitors
1.64E-07	11	38	FRS2-mediated cascade
1.91E-07	16	85	Integrin cell surface interactions
1.95E-07	10	31	CBL ubiquitinates FRS2 and FGFR
2.18E-07	7	13	FGFR2c binds to FGF
2.18E-07	7	13	FGFR2c ligand binding and activation

2.18E-07	7	13	Autocatalytic phosphorylation of FGFR2c
6.18E-07	13	61	PI3K phosphorylates PIP2 to PIP3
6.18E-07	13	61	PI3K inhibitors block PI3K catalytic activity
7.63E-07	7	15	FGFR2c mutants bind an expanded range of ligands
7.63E-07	7	15	Autocatalytic phosphorylation of FGFR2c mutants with enhanced ligand binding
1.08E-06	11	45	Signaling by FGFR mutants
1.23E-06	10	37	PIP2 conversion to PIP3
1.29E-06	15	86	Constitutive PI3K/AKT Signaling in Cancer
1.29E-06	15	86	Signaling by Type 1 Insulin-like Growth Factor 1 Receptor (IGF1R)
1.29E-06	15	86	IGF1R signaling cascade
1.55E-06	6	11	FGFR1c binds to FGF
1.55E-06	6	11	FGFR1c ligand binding and activation
1.55E-06	6	11	Autocatalytic phosphorylation of FGFR1c
1.55E-06	6	11	Signaling by activated point mutants of FGFR1
1.55E-06	6	11	FGFR1 P252X mutants bind ligand with enhanced affinity
1.55E-06	6	11	Autocatalytic phosphorylation of FGFR1 P252X mutant dimers
1.98E-06	5	7	FGFR3b binds to FGF
1.98E-06	5	7	FGFR3b ligand binding and activation
1.98E-06	5	7	Autocatalytic phosphorylation of FGFR3b
2.69E-06	19	140	Signaling by SCF-KIT
2.70E-06	10	40	Negative regulation of FGFR signaling
4.45E-06	14	83	IRS-related events triggered by IGF1R
1.03E-05	15	101	PI-3K cascade
1.03E-05	15	101	PI3K events in ERBB4 signaling
1.03E-05	15	101	PIP3 activates AKT signaling
1.03E-05	15	101	PI3K events in ERBB2 signaling
1.03E-05	15	101	PI3K/AKT Signaling in Cancer
1.48E-05	15	104	PI3K/AKT activation
1.66E-05	15	105	GAB1 signalosome
3.33E-05	13	86	Insulin receptor signalling cascade
3.82E-05	5	11	Formation of collagen fibrils
3.82E-05	5	11	Formation of collagen fibres
3.91E-05	10	53	GABA receptor activation
4.63E-05	10	54	Phospholipase C-mediated cascade
6.35E-05	20	189	Transmission across Chemical Synapses

6.38E-05	12	79	IRS-mediated signalling
7.15E-05	11	68	PI3K Cascade
8.20E-05	12	81	IRS-related events
1.06E-04	17	151	Signaling by ERBB4
1.45E-04	19	185	Signaling by PDGF
1.71E-04	17	157	DAP12 signaling
1.97E-04	16	144	Downstream signaling of activated FGFR
2.17E-04	5	15	Removal of fibrillar collagen N-propeptides

4.9. APPENDIX G: TOP100 SNPs FROM AHP RESULTS

Rank	SNP	P-Val	AHP Score	Rank	SNP	P-Val	AHP Score
1	rs17115004	1.42E-06	0.737535	**51**	rs7799600	3.41E-09	0.654374
2	rs2229163	4.14E-06	0.73671	**52**	rs8190621	5.19E-09	0.654374
3	rs7009117	5.67E-07	0.71945	**53**	rs10231699	1.69E-08	0.654374
4	rs6589360	6.64E-08	0.715562	**54**	rs16917234	2.05E-08	0.654374
5	rs7128875	1.38E-06	0.715562	**55**	rs1525256	2.27E-08	0.654374
6	rs7809317	1.77E-06	0.715562	**56**	rs1177930	4.17E-08	0.654374
7	rs6475523	9.05E-06	0.715562	**57**	rs9903602	6.23E-08	0.654374
8	rs2906288	4.26E-05	0.715562	**58**	rs7692461	8.64E-08	0.654374
9	rs1012023	3.73E-07	0.714776	**59**	rs4349	9.94E-08	0.654374
10	rs4846051	1.64E-05	0.714737	**60**	rs12273539	1.08E-07	0.654374
11	rs16895119	3.11E-05	0.714737	**61**	rs11015002	1.09E-07	0.654374
12	rs6348	3.61E-05	0.714737	**62**	rs6980130	1.16E-07	0.654374
13	rs1953439	1.85E-05	0.692803	**63**	rs10156202	1.37E-07	0.654374
14	rs10483448	2.49E-05	0.692803	**64**	rs11974602	1.52E-07	0.654374
15	rs2982712	1.23E-05	0.692215	**65**	rs3842842	1.92E-07	0.654374
16	rs17085734	8.97E-07	0.677721	**66**	rs17170624	2.39E-07	0.654374
17	rs12295969	9.78E-07	0.677721	**67**	rs10241143	2.64E-07	0.654374
18	rs11819808	1.56E-06	0.677721	**68**	rs6966555	3.01E-07	0.654374
19	rs12336931	2.6E-06	0.677721	**69**	rs363333	3.48E-07	0.654374
20	rs1494446	2.74E-06	0.677721	**70**	rs6442824	4E-07	0.654374
21	rs11568942	4.14E-06	0.677721	**71**	rs6961141	4.4E-07	0.654374
22	rs720024	4.6E-06	0.677721	**72**	rs11208836	5.01E-07	0.654374
23	rs7074934	5.19E-06	0.677721	**73**	rs3027403	5.74E-07	0.654374
24	rs7111410	6.31E-06	0.677721	**74**	rs17170037	6.1E-07	0.654374
25	rs16848098	6.78E-06	0.677721	**75**	rs6942437	6.38E-07	0.654374
26	rs5030351	1.09E-05	0.677721	**76**	rs16976623	7.89E-07	0.654374
27	rs17021884	1.36E-05	0.677721	**77**	rs17133801	8.38E-07	0.654374
28	rs6669695	1.72E-05	0.677721	**78**	rs9479126	8.5E-07	0.654374
29	rs741351	1.84E-05	0.677721	**79**	rs2045472	9.5E-07	0.654374
30	rs3787303	3.52E-05	0.677721	**80**	rs6723039	9.78E-07	0.654374

31	rs17128450	0.000037	0.677721	**81**	rs1112416	1.2E-06	0.654374
32	rs7785335	3.95E-05	0.677721	**82**	rs11214533	1.23E-06	0.654374
33	rs10155998	4.77E-05	0.677721	**83**	rs10811491	1.27E-06	0.654374
34	rs2172721	4.83E-05	0.677721	**84**	rs801942	1.51E-06	0.654374
35	rs7948159	3.14E-09	0.655025	**85**	rs2808356	1.62E-06	0.654374
36	rs7140732	2.25E-07	0.654962	**86**	rs17039997	1.69E-06	0.654374
37	rs2256712	2.51E-07	0.654962	**87**	rs6953850	1.69E-06	0.654374
38	rs13436623	5.96E-07	0.654962	**88**	rs12242971	1.71E-06	0.654374
39	rs9313450	8.71E-07	0.654962	**89**	rs2049590	2.07E-06	0.654374
40	rs7722406	9.32E-07	0.654962	**90**	rs952732	2.32E-06	0.654374
41	rs41499245	2.03E-06	0.654962	**91**	rs11804879	2.37E-06	0.654374
42	rs17070453	1.27E-05	0.654962	**92**	rs13411953	2.42E-06	0.654374
43	rs1150219	1.49E-06	0.654691	**93**	rs1402717	2.58E-06	0.654374
44	rs1065035	4.32E-06	0.654691	**94**	rs2903236	2.59E-06	0.654374
45	rs3867250	5.01E-06	0.654691	**95**	rs6505673	2.6E-06	0.654374
46	rs9983698	1.22E-05	0.654691	**96**	rs16847806	2.63E-06	0.654374
47	rs10280038	2.14E-05	0.654691	**97**	rs7300145	2.65E-06	0.654374
48	rs10239811	2.4E-12	0.654374	**98**	rs7955388	2.83E-06	0.654374
49	rs7805995	2.04E-11	0.654374	**99**	rs13438715	2.89E-06	0.654374
50	rs10259955	3.52E-10	0.654374	**100**	rs6435649	2.9E-06	0.654374

4.10. APPENDIX H: DAVID FUNCTIONAL ANNOTATION CHART OF COMBINED GENE RESULTS (TOTAL 83)

Category	Term	Count	%	PValue
KEGG	Axon guidance	52	2.106969	9.94E-08
KEGG	Focal adhesion	69	2.795786	9.22E-07
KEGG	Vascular smooth muscle contraction	44	1.782820	2.56E-06
KEGG	Calcium signaling pathway	59	2.390600	1.48E-05
KEGG	Long-term depression	29	1.175041	4.52E-05
KEGG	Pathways in cancer	94	3.808752	5.17E-05
PANTHER	Integrin signalling pathway	73	2.957861	6.24E-05
KEGG	ABC transporters	21	0.850891	8.48E-05
KEGG	MAPK signaling pathway	78	3.160454	1.28E-04
BIOCARTA	Thrombin signaling and protease-activated receptors	12	0.486224	2.04E-04
KEGG	Regulation of actin cytoskeleton	64	2.593193	3.27E-04
BIOCARTA	PKC-catalyzed phosphorylation of inhibitory phosphoprotein of myosin phosphatase	10	0.405186	3.88E-04
KEGG	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	28	1.134522	8.00E-04
PANTHER	5HT2 type receptor mediated signaling pathway	27	1.094003	1.47E-03
PANTHER	Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	49	1.985413	1.54E-03
PANTHER	Alzheimer disease-amyloid secretase pathway	28	1.134522	1.68E-03
KEGG	ECM-receptor interaction	29	1.175041	1.98E-03
KEGG	Small cell lung cancer	29	1.175041	1.98E-03
PANTHER	Oxytocin receptor mediated signaling pathway	24	0.972447	2.03E-03
KEGG	Dilated cardiomyopathy	31	1.256078	2.05E-03
BIOCARTA	Synaptic Proteins at the Synaptic Junction	11	0.445705	2.15E-03
PANTHER	Thyrotropin-releasing hormone receptor signaling pathway	24	0.972447	2.62E-03
BIOCARTA	g-Secretase mediated ErbB4 Signaling Pathway	6	0.243112	2.70E-03
KEGG	ErbB signaling pathway	29	1.175041	3.53E-03
BIOCARTA	Aspirin Blocks Signaling Pathway Involved in Platelet Activation	11	0.445705	3.65E-03

PANTHER	Alpha adrenergic receptor signaling pathway	14	0.567261	4.80E-03
PANTHER	Histamine H1 receptor mediated signaling pathway	18	0.729335	4.86E-03
KEGG	Prostate cancer	29	1.175041	5.05E-03
KEGG	Hypertrophic cardiomyopathy (HCM)	28	1.134522	5.05E-03
PANTHER	Inflammation mediated by chemokine and cytokine signaling pathway	86	3.484603	5.11E-03
KEGG	Non-small cell lung cancer	20	0.810373	5.19E-03
KEGG	Thyroid cancer	13	0.526742	5.80E-03
KEGG	Mismatch repair	11	0.445705	7.67E-03
PANTHER	Endothelin signaling pathway	32	1.296596	7.82E-03
KEGG	Long-term potentiation	23	0.931929	8.55E-03
BIOCARTA	Role of ERBB2 in Signal Transduction and Oncology	11	0.445705	9.02E-03
KEGG	Gap junction	28	1.134522	9.91E-03
PANTHER	Angiotensin II-stimulated signaling through G proteins and beta-arrestin	20	0.810373	1.09E-02
KEGG	TGF-beta signaling pathway	27	1.094003	1.38E-02
KEGG	Cell cycle	36	1.458671	1.40E-02
KEGG	Melanoma	23	0.931929	1.46E-02
PANTHER	Nicotinic acetylcholine receptor signaling pathway	29	1.175041	1.59E-02
KEGG	Glycosphingolipid biosynthesis	8	0.324149	1.62E-02
KEGG	Nucleotide excision repair	16	0.648298	1.68E-02
KEGG	Endometrial cancer	18	0.729335	1.76E-02
BIOCARTA	IL-7 Signal Transduction	8	0.324149	2.10E-02
BIOCARTA	Bioactive Peptide Induced Signaling Pathway	13	0.526742	2.18E-02
PANTHER	Metabotropic glutamate receptor group I pathway	16	0.648298	2.22E-02
KEGG	Phosphatidylinositol signaling system	23	0.931929	2.37E-02
KEGG	Chronic myeloid leukemia	23	0.931929	2.75E-02
KEGG	Glioma	20	0.810373	2.95E-02
BIOCARTA	Role of PPAR-gamma Coactivators in Obesity and Thermogenesis	6	0.243112	3.10E-02
PANTHER	Muscarinic acetylcholine receptor 1 and 3 signaling pathway	21	0.850891	3.15E-02
PANTHER	Axon guidance mediated by netrin	15	0.607780	3.15E-02
KEGG	Adherens junction	23	0.931929	3.66E-02
KEGG	DNA replication	13	0.526742	3.67E-02
PANTHER	Wnt signaling pathway	87	3.525122	3.76E-02
BIOCARTA	Apoptotic Signaling in Response to DNA Damage	10	0.405186	3.89E-02
BIOCARTA	Nuclear Receptors in Lipid Metabolism and Toxicity	14	0.567261	4.13E-02

BIOCARTA	Phospholipids as signalling intermediaries	10	0.405186	5.18E-02
KEGG	O-Glycan biosynthesis	11	0.445705	5.42E-02
KEGG	GnRH signaling pathway	27	1.094003	5.75E-02
BIOCARTA	Integrin Signaling Pathway	13	0.526742	5.81E-02
KEGG	Pancreatic cancer	21	0.850891	5.86E-02
KEGG	Tight junction	35	1.418152	5.94E-02
BIOCARTA	Agrin in Postsynaptic Differentiation	11	0.445705	5.94E-02
BIOCARTA	Angiotensin II mediated activation of JNK Pathway via Pyk2 dependent signaling	11	0.445705	5.94E-02
PANTHER	Beta1 adrenergic receptor signaling pathway	16	0.648298	6.06E-02
BIOCARTA	Chromatin Remodeling by hSWI/SNF ATP-dependent Complexes	8	0.324149	6.41E-02
BIOCARTA	MAPKinase Signaling Pathway	25	1.012966	6.52E-02
BIOCARTA	Human Cytomegalovirus and Map Kinase Pathways	7	0.283630	7.06E-02
KEGG	Fc epsilon RI signaling pathway	22	0.891410	7.17E-02
KEGG	Cell adhesion molecules (CAMs)	34	1.377634	7.45E-02
BIOCARTA	CCR3 signaling in Eosinophils	9	0.364668	7.57E-02
PANTHER	EGF receptor signaling pathway	39	1.580227	8.60E-02
KEGG	Purine metabolism	38	1.539708	9.27E-02
BIOCARTA	Nuclear receptors coordinate the activities of chromatin remodeling complexes and coactivators to facilitate initiation of transcription in carcinoma cells	7	0.283630	9.62E-02
PANTHER	Beta2 adrenergic receptor signaling pathway	15	0.607780	9.62E-02
BIOCARTA	Erk and PI-3 Kinase Are Necessary for Collagen Binding in Corneal Epithelia	9	0.364668	9.67E-02
PANTHER	Axon guidance mediated by semaphorins	16	0.648298	9.79E-02
BIOCARTA	p38 MAPK Signaling Pathway	12	0.486224	9.82E-02
KEGG	Glycosaminoglycan degradation	8	0.324149	9.97E-02

4.11. APPENDIX I: REACTOME ANALYSIS OF AHP RESULTS (TOP100)

P-value	Number of genes	Total number of genes	Name of the event
1.22E-04	2	2	D-glyceraldehyde 3-phosphate + orthophosphate + NAD+ <=> 1,3-bisphospho-D-glycerate + NADH + H+
1.22E-04	2	2	1,3-bisphospho-D-glycerate + NADH + H+ <=> D-glyceraldehyde 3-phosphate + Orthophosphate + NAD+
2.69E-04	10	238	Metabolism of carbohydrates
2.69E-04	3	12	Keratan sulfate degradation
3.63E-04	2	3	limit dextrin-glycogenin => ((1,6)-alpha-glucosyl)poly((1,4)-alpha-glucosyl) glycogenin
3.63E-04	2	3	((1,6)-alpha-glucosyl)poly((1,4)-alpha-glucosyl)glycogenin => poly{(1,4)-alpha-glucosyl} glycogenin + alpha-D-glucose
3.63E-04	2	3	poly((1,4)-alpha-D-glucosyl) glycogenin => glycogen-glycogenin
3.63E-04	2	3	Activated NOD oligomer recruites RIP2 (RICK)
3.99E-04	4	32	Keratan sulfate/keratin metabolism
4.39E-04	3	14	RIP2 is K63 polyubiquitinated
5.83E-04	5	62	Glucose metabolism
7.20E-04	2	4	GPVI binds Fyn and Lyn
7.20E-04	2	4	RIP2 binds NEMO
7.20E-04	2	4	CARD9 binds RIP2 (and NOD2)
1.77E-03	2	6	Binding of GPVI:Fc Epsilon R1 gamma receptor complex with collagen
1.77E-03	2	6	Fyn/Lyn-mediated phosphorylation of FcR1 gamma
1.77E-03	2	6	Grb2 binds CD28
1.77E-03	2	6	Gads binds CD28
2.47E-03	2	7	Binding of Syk tyrosine kinase
2.47E-03	2	7	Translocation of Vav1 to CD28
2.47E-03	2	7	Activation of Vav1
2.47E-03	2	7	Syk is released
2.47E-03	2	7	Syk autophosphorylates
3.27E-03	2	8	Glycogen synthesis

3.27E-03	2	8	Phosphorylation of CD28
3.27E-03	2	8	Activation of Rac1 by pVav1
3.27E-03	2	8	Activation of Cdc42 by pVav1
4.17E-03	2	9	SHP2 phosphatase binds CTLA-4
4.17E-03	2	9	PI3K binds CD28
4.17E-03	2	9	CD28 bound PI3K phosphorylates PIP2 to PIP3
4.17E-03	2	9	A20 deubiquitinates RIP2
4.17E-03	2	9	CYLD deubiquitinates NEMO
4.44E-03	6	140	Signaling by SCF-KIT
4.54E-03	4	61	PI3K phosphorylates PIP2 to PIP3
4.54E-03	4	61	PI3K inhibitors block PI3K catalytic activity
5.11E-03	5	101	PI-3K cascade
5.11E-03	5	101	PI3K events in ERBB4 signaling
5.11E-03	5	101	PIP3 activates AKT signaling
5.11E-03	5	101	PI3K events in ERBB2 signaling
5.11E-03	5	101	PI3K/AKT Signaling in Cancer
5.78E-03	5	104	PI3K/AKT activation
6.02E-03	5	105	GAB1 signalosome
6.28E-03	2	11	CD28 dependent Vav1 pathway
6.28E-03	2	11	RIP2 induces K63-linked ubiquitination of NEMO
6.68E-03	3	35	NOD1/2 Signaling Pathway
7.48E-03	6	156	Downstream signal transduction
7.48E-03	2	12	K63 polyubiquitinated RIP2 associates with the TAK1 complex
7.48E-03	2	12	TAK1 is activated
8.78E-03	2	13	Platelet Adhesion to exposed collagen
8.78E-03	2	13	FGFR4 binds to FGF
8.78E-03	2	13	FGFR4 ligand binding and activation
8.78E-03	2	13	Autocatalytic phosphorylation of FGFR4
1.02E-02	2	14	Glycogen breakdown (glycogenolysis)
1.02E-02	2	14	GPVI stimulates PI3K beta, gamma
1.16E-02	2	15	Formation of ATP by chemiosmotic coupling
1.16E-02	2	15	ATP is synthesized from ADP and Pi by ATPase
1.16E-02	2	15	Enzyme-bound ATP is released
1.16E-02	2	15	ADP and Pi bind to ATPase
1.24E-02	6	174	Signaling by EGFR
1.31E-02	6	176	Signaling by EGFR in Cancer

1.32E-02	2	16	Connection of adjacent cells through calcium-dependent trans-dimerization of cadherin
1.32E-02	2	16	Regulation of KIT signaling
1.34E-02	10	405	GPCR ligand binding
1.35E-02	8	286	Signalling by NGF
1.49E-02	2	17	Activated TAK1 mediates phosphorylation of the IKK Complex
1.49E-02	2	17	Phosphorylation of CTLA-4
1.49E-02	2	17	Activated TAK1 phosphorylates MKK4/MKK7
1.49E-02	2	17	activated human TAK1 phosphorylates MKK3/MKK6
1.50E-02	4	86	Constitutive PI3K/AKT Signaling in Cancer
1.64E-02	6	185	Signaling by PDGF
1.66E-02	2	18	Na+/Cl- dependent neurotransmitter transporters
1.84E-02	2	19	CD28 dependent PI3K/Akt signaling
1.92E-02	8	305	Class A/1 (Rhodopsin-like receptors)
2.03E-02	2	20	Interaction of cadherin with Beta/gamma catenin, alpha catenin and p120 catenin
2.03E-02	2	20	JNK (c-Jun kinases) phosphorylation and activation mediated by activated human TAK1
2.15E-02	5	144	Downstream signaling of activated FGFR
2.15E-02	5	144	Liganded Gq/11-activating GPCRs act as GEFs for Gq/11
2.15E-02	5	144	Liganded Gq-activating GPCRs bind inactive heterotrimeric Gq
2.15E-02	5	144	The Ligand:GPCR:Gq complex dissociates
2.23E-02	2	21	activated TAK1 mediates p38 MAPK activation
2.23E-02	2	21	Activated FGFR mutants bind FRS2alpha
2.23E-02	2	21	FGFR mutants phosphorylate FRS2alpha
2.23E-02	2	21	Activated FGFR mutants bind PLC-gamma
2.23E-02	2	21	PLC-gamma phosphorylation by FGFR mutants
2.23E-02	2	21	Activated PLC-gamma release by activated FGFR mutants
2.29E-02	3	55	Nucleotide-binding domain, leucine rich repeat containing receptor (NLR) signaling pathways
2.37E-02	6	201	NGF signalling via TRKA from the plasma membrane
2.44E-02	2	22	CTLA4 inhibitory signaling
2.44E-02	2	22	TAK1 activates NFkB by phosphorylation and activation of IKKs complex
2.44E-02	2	22	Regulation of signaling by CBL
2.58E-02	5	151	Signaling by ERBB4
2.65E-02	2	23	FGFR ligand binding and activation
2.65E-02	2	23	GRB2-SOS1 is recruited by activated FGFR mutants
2.87E-02	2	24	Phosphorylation of FRS2-alpha by activated FGFR

2.87E-02	2	24	Activated FGFR binds PLC-gamma
2.87E-02	2	24	PLC-gamma phosphorylation by FGFR
2.87E-02	2	24	Activated PLC gamma release by activated FGFR
2.87E-02	2	24	Phosphorylation of FRS2-beta by activated FGFR
2.87E-02	2	24	Activated FGFR binds FRS2alpha
2.87E-02	2	24	Activated FGFR binds FRS2beta

TEZ FOTOKOPİ İZİN FORMU

ENSTİTÜ

Fen Bilimleri Enstitüsü ☐

Sosyal Bilimler Enstitüsü ☐

Uygulamalı Matematik Enstitüsü ☐

Enformatik Enstitüsü ☐

Deniz Bilimleri Enstitüsü ☐

YAZARIN

Soyadı : .....
Adı : .....
Bölümü : .....

TEZİN ADI (İngilizce) : .....
.....
.....
.....
.....

TEZİN TÜRÜ : Yüksek Lisans ☐ Doktora ☐

1. Tezimin tamamı dünya çapında erişime açılsın ve kaynak gösterilmek şartıyla tezimin bir kısmı veya tamamının fotokopisi alınsın. ☐
2. Tezimin tamamı yalnızca Orta Doğu Teknik Üniversitesi kullancılarının erişimine açılsın. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.) ☐
3. Tezim bir (1) yıl süreyle erişime kapalı olsun. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.) ☐

Yazarın imzası ..... Tarih .....