Inferring schizophrenia biology from genome-wide data

A thesis submitted for the degree of Doctor of Philosophy from Cardiff University, 2013

Douglas Ruderfer

Advisors: Peter Holmans, Nick Craddock, and Shaun Purcell

Summary

Disease loci underlying complex disorders such as schizophrenia (SCZ) are likely to exist as variants of multiple types beyond the single nucleotide polymorphisms (SNPs) typically surveyed in genome-wide association studies (GWAS). Large copy number variants (CNV), small insertions and deletions (INDELS), rare compound heterozygous mutations and other classes of genetic variation are all expected to contribute to disease risk. A true understanding of disease genetics cannot be attained without careful consideration of each class of variation and how they are related. Additionally, it is often the case that no or few single variants have strong enough effect sizes to attain the level of statistical significance required to offset the large number of tests performed in a genome-wide single locus survey of any variant type. It has been shown that a cumulative burden of risk alleles is correlated with disease risk, indicating true biological signal within those variants. However, this approach merely implicates that class of variation across the genome and does not refine variants to a subset of risk regions or genes. Defining sets of genes within biological pathways and testing them as a group will allow for both the power of aggregating loci of small effect and interpretation of the underlying biology.

This dissertation assesses the role of multiple classes of variation from different technologies in risk to schizophrenia with a particular focus on brain related biological pathways.

Acknowledgements

"If I have seen further it is only by standing on the shoulders of giants"

All accomplishments are a product of motivation and inspiration, the best motivation is often within oneself but that stems from being inspired and supported by others. I recognize that this product has only been possible with the help of many. First and foremost are my parents, I am literally a product of them both biologically and otherwise. There is not a positive trait that I possess that is not evident in them. They have dedicated their lives to my success and happiness and for this I am forever grateful. I am fortunate to have a loving and supportive immediate and extended family. My grandparents Jean and Joseph who instilled in me a sense of hard work and character. My grandparents Leonard and Jean who demonstrated dedication to loved ones and the ability to stay positive despite hard times. My brother Dave who has always been there for me and keeps me grounded with his keep it simple "fruit loops" theory. I have received immense support from my Uncle Guy and my cousins Danielle and Gina who have been more like older sisters my whole life. I am continually inspired by my Aunt Jeanie who always stressed love and family and despite her passing looms large in my heart and mind.

I have an incredible collection of friends that I have made through the years, in particular those from college that are achieving success beyond what we could have ever imagined during school. Their accomplishments continue to inspire and drive me towards my own. At the most difficult times during this process it was them who provided the escape I needed.

I am also lucky to have met and worked with some of the smartest, most dedicated people I know of who have believed in me and provided opportunities that otherwise may not have been possible. My initial interest in this field was sparked from conversations with my friend Peter Chen, working with him and seeing his desire to make a difference inspired me to pursue a career in this field and to this day inspires me to learn and continue reaching for solutions to important problems. I owe my career to an unlikely opportunity provided to me by Dr. Leonid Kruglyak who allowed me to immediately work on impactful science with zero experience in biology or genetics. The two years I spent in Princeton having conversation with Leonid and colleagues such as Dr. Matt Rockman are still the basis for my knowledge in fundamental genetics.

For the last 6 years I have been fortunate to work with Drs. Shaun Purcell and Pamela Sklar. They have provided me with many opportunities to both contribute to very important science as well as explore my own ideas and projects. It is only with there support that I have been able to complete this thesis while continuing to learn and produce scientifically. I owe much to all of my colleagues at MGH, the Broad Institute and MSSM of which there are too many to name but who are indispensable sources of information and ideas.

Finally, I have had two outstanding advisors at Cardiff. Dr. Nick Craddock has always ensured my research was headed in a direction towards ultimate success and instilled a focus on preparation from day one. Dr. Peter Holmans has been an immense source of knowledge and our conversations have always led to better and more robust science. Additionally, he and Lesley Jones have been extremely welcoming and gracious, hosting me on my visits, making Cardiff an inviting place and the whole experience a pleasant one.

Chapter 1. Introduction

1.1 Brief historical perspective on human genetics and trait mapping

For nearly 150 years, going back to Mendel, individuals have sought to map the relationship between phenotype (any observable characteristic) and genotype (a heritable marker). That there was a relationship whereby heritable genetic factors contributed to observable traits was known well before a complete understanding of the foundation of genetics existed. Even the word "gene," coined by Danish botanist Wilhelm Johannsen in 1909, was used to describe units of heritability before anything was known about DNA. For nearly all of this time, trait mapping was performed in "simpler" organisms such as nematodes, yeast and fruit flies where genetic markers could more easily be identified and selective breeding experiments could be done. While genetic maps for flies have existed since 1913 (1) not until 1980 did the same exist for humans (2). This map created in families from restriction fragment length polymorphisms identified polymorphic locations across the human genome and allowed for candidate free, genome wide mapping of phenotype to inherited stretches of genome amongst related individuals. Now able to identify distinguish markers among individuals, researchers could begin to identify genomic regions responsible for Mendelian diseases by identifying segregation patterns, that is genomic stretches shared only by individuals having the disease. Unfortunately, most common disorders do not follow a simple Mendelian model of a single, fully penetrant disease causing mutation. Such "complex" disorders will likely be a

culmination of many variants of smaller effect therefore requiring denser, more complete maps of the human genome in order to have any power to detect these variants.

After some success in identifying regions of the genome responsible for certain Mendelian disorders (3) the project to sequence the entire human genome was started in 1993. The Human Genome Project, for the first time, revealed the nearly complete genome of our species and by 2001 provided a reference in which to compare to not only other species but also other individuals(4, 5). Having created a draft sequence of the human genome the next step in being able to map phenotype to genotype required identifying a dense set of polymorphic markers. In 2000, an initial large collection of the most common form of variation, single nucleotide polymorphisms (SNPs) was collected from sequencing reads during the Human Genome Project and also specifically from the SNP consortium (6, 7). The initial set of 1.42 million SNPs made it clear that most SNPs are rare but given patterns of inheritance and recombination whereby SNPs physically near each other are more likely to be inherited together as single unit known as a haplotype a more efficient map could be created. The idea behind The Haplotype Map (HapMap) Project was to create such a map whereby common SNPs that tag many other SNPs in the region could be selected in order to assess variation of a very high proportion of the genome given a relatively small number of markers (8). The HapMap project was completed in 2005 and also aimed to identify SNP frequency differences in populations around the world, selecting individuals with African, Asian and European ancestral backgrounds. These datasets have been instrumental in understanding how allele

frequency differences between individuals of differing populations could affect tests of association, otherwise known as population stratification. As technology continued to advance, it became commonplace to, on a single genotyping array, assess variation across a million SNPs identified from the Human Genome Project and HapMap project. The next planned project to expand our understanding of variation across the human genome was the 1000 Genomes Project ,which was published in 2010 (9). This project aimed to sequence the entire genomes of greater than 1,000 individuals selected from 14 populations in the largest effort yet to have a human variation map. Sequencing provides the ultimate map, in that it allows you to assess variation at every sequence-able base in the genome. The continued improvement of human variation maps has allowed for denser and more accurate assessment of relationship between phenotypes and genotypes either through linkage or association. This includes the ability to impute unseen markers by estimating the most likely genotype given a particular haplotype from a reference population usually the HapMap or 1000 genomes. Provided with these variation maps, researchers were over time more and more powered to identify variations contributing to disease.

1.1.1 Mapping disease traits

Early on it was identified that traits were shared amongst related individuals and that variants contributing to those traits would be passed down to children and both trait and variant would segregate together. This concept is known as genetic

linkage and could be directly used to identify regions of the genome "linked" to any particular trait including disease. Once a set of markers has been established and a family identified, the alleles at these markers can be assessed and recombination between any pair of markers can be predicted. The goal is to identify markers in which recombination hasn't separated it from the "disease gene" being searched for. Linkage analysis has found success in many Mendelian diseases and even a handful of more complex diseases such as Alzheimer's (10), hypertension (11) and age-related macular degeneration (12). The approach becomes more difficult as a variant becomes less penetrant, this is largely due to the small average size of human families and the low resolution given small numbers of recombination events per generation. The affect of environment and the existence of individuals with disease but absent of the expected genetic mutation (phenocopy) will also add difficulty to this approach. Even when regions can be identified they are often large (~20cM) and require further efforts to narrow down to a particular gene or locus (13, 14). To increase power to detect variants of more modest effect in complex diseases and to utilize the expansive set of markers being generated by the Human Genome Project, HapMap Project and 1000 Genomes project, researchers began to move towards genome wide association studies (GWAS) which are more powered to detect smaller effect size and require a denser map to work (15). GWAS is performed by genotyping large amounts of markers in case and control samples and testing differences in frequencies of those markers with respect to disease. GWAS has been quite successful in particular as samples have become larger, smaller effects have been identified (16). The development of genotype arrays with many markers has also

allowed for many advances in the field. With these chips, large copy number variants (CNV), that is regions of the genome that are deleted or duplicated, can be accurately called and recent methods have allowed for heritability to be estimated. All of these approaches and improvements have been extensively applied to psychiatric diseases including schizophrenia.

1.2 Genetics of schizophrenia

1.2.1 Introduction

Schizophrenia is a devastating psychiatric disorder characterized by psychosis (delusions and hallucinations), negative symptoms such as lack of motivation and social withdrawal, reduced cognitive functioning and mood symptoms (17). The disease not only affects mental health but individuals with schizophrenia live 12-15 years less than the average healthy individual and substantially reduces fecundity. Estimates of prevalence range from 0.3% to 1% depending on how narrow the phenotypic definition is and what population was used, a large national Swedish sample estimated a lifetime risk of 0.4% (18, 19). There has been exhaustive research done on estimating the amount of observable phenotypic difference in individuals that is due to genetic differences, known as heritability, of schizophrenia, including from twin studies, adoption studies and population studies. While these estimates have often varied and include large

confidence intervals they nearly all have concluded that schizophrenia is an extremely heritable disorder. A large meta-analysis of twin studies estimated the heritability to be 81% (20) and estimates from the same Swedish national sample used to estimate prevalence and used in this thesis identified a heritability of 64% (21). Both of these estimates indicate that while not complete, the contribution of genetics to schizophrenia risk is greater than other known factors including environmental factors of which several have been proposed including urbanization(22), and drug use (23). Such a strong genetic component has encouraged extensive efforts to identify particular mutations or genes contributing to schizophrenia risk. These efforts include performing GWAS on large samples of cases and controls to identify common variants of more modest effect, identifying and understanding the role of structural variation, in particular large CNVs that can be called from genotyping arrays, sequencing both case and control individuals to discover very rare mutations with theoretically larger effects and looking in families for segregation of alleles, over transmission of risk alleles and even *de novo* mutations that may be contributing to disease risk. Although *de novo* mutations would not be part of any heritability estimate, as by definition they aren't passed down from parent to offspring, they have high likelihood of being pathogenic. All of these types of variation will contribute to the full picture of the genetic causes of schizophrenia and will likely all play a role to some extent. In the rest of this introductory chapter I highlight the findings over the years and how those finding have been the background for all the analyses done in this thesis.

1.2.2 CNVs /structural variation

Before genotyping arrays and maps of human genetic markers, researchers utilized cytogenetics to assess visually whether there were large structural aberrations in an individual's genome. It was through this process that the first genomic regions and genes were implicated for schizophrenia. The first of these was 22q11 deletion syndrome or DiGeorge syndrome, this large structural defect results in many symptoms including abnormal facial features and heart defects but also substantially increases schizophrenia risk 20 to 30 fold higher than the baseline rate (24). Further cytogenetic analysis of Scottish juvenile offenders identified an individual with symptoms of a psychotic disorder carrying a translocation on chromosome 1. A follow up study was performed on his family and 34 of the 77 included members carried the same translocation, 16 of which were diagnosed with a psychiatric disease compared to 5 of the 43 without the translocation (25). This translocation occurs directly within the coding sequence of a gene that was named *DISC1* or disrupted in schizophrenia 1. A number of additional genes have been implicated by this strategy including *FEZ1* and *GRIK4* but none have demonstrated Mendelian properties of near complete penetrance and segregation within family, so further work is required to claim disease relevance (26). In fact, a recent study of *DISC1* identified no further population based genetic evidence of mutations contributing to schizophrenia and limited evidence to major depression disorder, demonstrating the difficulty of identifying true disease variants (27).

These findings implicated structural variation as playing an important role in disease risk more broadly. The first class of structural variation that was capable of

being assessed genome-wide was copy number variation (CNV). CNVs are detected in places where an individual has more or fewer copies of large stretches of the genome than the expected two. The first studies looking at CNVs across the genome utilized the array-comparative genomic hybridization (aCGH) technology comparing copy number of each sample to a normal reference individual. Initial studies identified several possible candidate regions using this technology, in particular two independent studies identified rare CNVs spanning *NRXN1* which is known to have a role in synaptic function (28, 29). Walsh et al. also noticed a significant excess of CNV in SCZ cases compared to controls but exclusively in rare CNVs, no such excess was found in more common CNVs. In general, a study by the Wellcome Trust Case Control Consortium (WTCCC) on seven disorders including SCZ found limited evidence for a role of common CNV in disease (30). Since these findings of rare CNVs, numerous studies were performed looking more closely at these events in larger populations as well as explicitly looking at *de novo* events. Advances in the ability to detect these events from high density SNP arrays greatly increased sample sizes available to study CNVs. Two pieces of software in particular, one for each of the most popular two genotyping platforms (Affymetrix, Illumina), recognized that hybridization intensity was directly correlated with copy number and that extended sequences of higher or lower than average intensity indicated stretches of higher or lower copy number than the expected number of two (31, 32). Using these new technologies two papers published at the same time presented further evidence of the role of rare CNVs in SCZ (33, 34). The paper by the International Schizophrenia Consortium (ISC) identified CNVs in a sample of over 6,000 individuals from multiple population

groups and looking explicitly at large (>100kb) and rare (<1%) CNVs identified a significantly increased rate (OR=1.15) in SCZ cases compared to matched controls. In addition, they identified three particular regions of very large deletions that were significantly associated with SCZ including 22q11.2, 15q13.3 and 1q21.1. In Stefansson et al. they utilized a large healthy population sample to identify a set of 66 *de novo* CNVs under the hypothesis that since these regions were highly mutable but very infrequent in the population they were likely undergoing negative selection. They followed up these 66 *de novo* CNVs in two separate case/control datasets and identified 3 deletions significantly associated with SCZ including 1q21.1, 15q11.2 and 15q13.3, two of which were seen in both studies. Another study involving *de novo* CNVs identified an increased rate of these events in SCZ (5% in cases compared to 2% in controls) but also looked to identify some biological function being affected by these mutations (35). They utilized a previously defined set of genes involved in the postsynaptic density (PSD)(36) and found significant enrichment of *de novo* CNVs in these genes ($P=1.72 \times 10^{-6}$) which was mostly defined by two subgroups of the PSD *N*-methyl-D-aspartate receptor (NMDAR) and neuronal activity-regulated cytoskeleton-associated protein (ARC). These results indicate that mutations in two complexes known to be important in synaptic plasticity and cognition (NMDAR and ARC) are likely involved in pathogenesis of SCZ.

All of these studies have provided clear evidence of a role for rare, large CNVs, both deletions and duplications, in SCZ. Many other CNVs have been found since these initial studies and efforts have continued to not only identify CNVs associated with SCZ but to identify biological function that is being disrupted by these

mutations. Table 1.1 presents a list of implicated structural variation events in psychiatric disease with those relevant to SCZ highlighted (37). Pathway and gene set tests such as those done in *Kirov et. al.* (38) will be instrumental in helping to understand the biological underpinnings of this disorder. The identified CNVs are also particularly large affecting many genes and being associated with many other symptoms as seen in 22q11 deletion syndrome. The question remains whether it is particular genes within these regions that result in large increases in SCZ risk or is it a combination of disrupting multiple genes that carries this increased risk. Looking at smaller CNVs for fine mapping particular genes but also for assessing their contribution to disease risk could help answer this question. While common CNVs appear to not be playing a critical role in disease risk, the size of contribution from all common variation was still an open question.

Table 1.1. Previously associated regions of structural variation to psychiatric disease (SCZ in bold)

Structural variant	Location (Mb)	Genes	Type	Disorder	Frequency in cases	Frequency in controls	Odds ratio	*P*value
1q21.1	**chr1:145.0–148.0**	**34**	**Deletion**	**SCZ**	**0.0018**	**0.0002**	**9.5**	**8×10^{-6}**
			Duplication	**SCZ**	**0.0013**	**0.0004**	**4.5**	**0.02**
2p16.3	**chr2:50.1–51.2**	***NRXN1*exons**	Deletion	ASD				0.004
			Deletion	**SCZ**	**0.0018**	**0.0002**	**7.5**	**1×10^{-6}**
3q29	**chr3:195.7–197.3**	**19**	**Deletion**	**SCZ**	**0.001**	**0**	**3.8**	**4×10^{-4}**
7q11.23	chr7:72.7–74.1	25	Duplication	ASD	0.0011			0.003
7q36.3	**chr7:158.8–158.9**	***VIPR2***	**Duplication**	**SCZ**	**0.0024**	**0.0001**	**16.4**	**4×10^{-5}**
15q11.2	chr15:23.6–28.4	70	Duplication	ASD	0.0018			4×10^{-9}
15q13.3	**chr15:30.9–33.5**	**12**	Duplication	ADHD	0.0125	0.0061	2.1	2×10^{-4}
			Duplication	ASD	0.0013			2×10^{-5}
			Deletion	**SCZ**	**0.0019**	**0.0002**	**12.1**	7×10^{-7}
16p13.11	chr16:15.4–16.3	8	Duplication	ADHD	0.0164	0.0009	13.9	8×10^{-4}
16p11.2	**chr16:29.5–30.2**	**29**	Deletion	ASD	0.0037			5×10^{-29}
			Duplication	ASD	0.0013			2×10^{-5}
			Duplication	**SCZ**	**0.0031**	**0.0003**	**9.5**	3×10^{-8}
17q12	**chr17:34.8–36.2**	**18**	Deletion	ASD	0.0017	0	6.12	9×10^{-4}
			Deletion	**SCZ**	**0.0006**	**0**	**4.49**	**3×10^{-4}**
22q11.21	**chr22:18.7–21.8**	**53**	Deletion or duplication	ASD	0.0013			0.002
			Deletion	**SCZ**	**0.0031**	**0**	**20.3**	**7×10^{-13}**

1.2.3 Linkage analysis, GWAS and heritability

With high heritability estimates and advancing genetic maps for correlating phenotype to genotype, there was strong interest in identifying variants underlying schizophrenia risk. Initial approaches involved performing linkage analysis using families. As described earlier, linkage analysis involves having a family with multiple affected and unaffected members and identifying segregation of disease with regions of the genome where recombination has not occurred. The initial studies utilized microsatellite markers. Typically, these would consist of around 400 multi-allelic markers genome wide, with 5-10cM spacing between markers. More recently, linkage analyses have been performed using SNP grids (39). Unlike association, genome-wide linkage studies do not require a very dense grid of SNPs (around 10,000 is sufficient), and care must be taken to prevent LD between markers causing false positives (40). However, such SNP maps can still give greater linkage information and power than microsatellite grids (41, 42). Linkage analyses have been performed for many years and numerous candidate genes have been identified, some of which, such as NRG1are still studied in detail today (43, 44). These candidates represented novel findings implicating new genes to schizophrenia. Many new genes were identified this way but few were replicated across studies. There are several reasons for this lack of replication including a preponderance of false positives that are randomly distributed across the genome or an underlying disease architecture of many highly penetrant, extremely rare variants specific to families that aren't seen in other families for replication. Additionally, linkage analysis of

complex traits is likely to have low power for the following reasons. Firstly, there is the issue of diagnostic uncertainty. A family often exhibits more than one type of mental illness as many of these disorders will exist making segregation more difficult to assess. Secondly, the human family is small in general but in families with high rates of psychiatric disease the number of members will be even smaller given that schizophrenia patients have a third as many children as healthy individuals. Small families, such as affected sib pairs, are less powerful than large pedigrees, due to there being fewer meioses, so very large samples are needed to achieve power (15). Thirdly, as will be discussed later, much of the genetic susceptibility to psychiatric disorders is attributable to common variants of small effect. Linkage analysis gives little power to detect such variants (15). In an attempt to perform the most powerful linkage analysis to date a group performed a meta-analysis on 32 independent family samples including 3,255 pedigrees and 7,413 affected individuals, identifying no regions reaching criteria for genome-wide significance (45, 46). In general, linkage analysis has been viewed as mostly failing to identify particular genomic regions or genes contributing to SCZ risk as nearly all of the regions that have been identified have failed to be confirmed in a replication sample. An alternative approach to variant discovery known as association forgoes the need for families and gives greater power than linkage to detect modest effect sizes (15). Association analysis requires a dataset of cases and controls with a large number of markers that are each tested for frequency differences between the two groups. A major benefit to this approach is the relative ease at which one can collect larger samples of cases and controls compared to collecting large families. Initial association studies focused on

variants within selected candidate genes with prior belief of being related to disease. As technology improved it became possible to perform hypothesis free association testing across the genome, known as a genome-wide association study (GWAS). This approach is not without its own problems as increasingly large samples are required to have power to detect decreasingly small effect sizes, markers are singly tested so genes or regions containing multiple risk variants are likely to be missed unlike in linkage where one is testing whether shared genomic regions are related to disease status. In particular, combining individuals from different populations could result in false positives known as population stratification. This is a situation where allele frequencies differ between cases and controls but due to differences in allele frequencies across populations and not because of association to phenotype. An extreme example would be to perform a GWAS with individuals of African descent as cases and individuals of European descent as controls you would identify many highly significant results. Methods for appropriately accounting for these differences are focused around having cases and controls closely matched for ancestry and then testing individuals within ancestral clusters or by including principal components from the genetic data that represent ancestry in the association model, several programs perform these analysis including EIGENSTRAT (47). Initial association analyses were performed on a limited set of genes selected as "candidates" based on previous hypotheses. The reason for this was both reducing costs but also reducing testing burden by focusing on genes with the highest prior of being related to SCZ. As variant databases and technology advanced it became possible to perform GWAS thereby testing most of the genome in an assumption free manner. GWAS has been

very successful in identifying novel candidate genes/regions for many disorders and in the last few years this has been true in psychiatric disorders as well, representing some of the strongest evidence yet for the contribution of common variants to SCZ (16). One of the first genes with significant statistical support for association was *ZNF804A* which was first identified in a sample of 479 cases and 2,937 controls and was further followed up in 16,726 additional subjects and reached genome-wide significance after combining with bipolar samples (48). In 2009, three GWAS studies were published, each pointing to the major histocompatibility complex (MHC) region being significantly associated to SCZ risk (49-51). The MHC region on chromosome six consists of many of the genes related to immune function that have undergone recent evolution in humans. It also represents the largest region of linkage disequilibrium in the human genome making fine mapping this association to a particular variant or gene extraordinarily difficult. For the first time, there was strong evidence of common variants contributing to psychiatric disease. Efforts to compile larger and larger sample sizes in the hope of identifying loci of smaller effect resulted in a paper from the Psychiatric Genetics Consortium (PGC) and a mega-analysis of SCZ identifying 5 loci meeting genome-wide significance (52). This effort has continued with the most recent PGC analysis identifying greater than 100 genome-wide significant regions (unpublished). Having this many loci will enable further study of how these regions are biologically related and help researchers understand the disease better.

Despite the dramatic increase in associated variants, the heritability explained is very small due to the very modest effects of each locus. A major

question, particularly before the most recent PGC results, was to what extent common variation was contributing to common diseases. As a major part of the second paper from the International Schizophrenia Consortium they looked to answer this question by taking a cumulative individual risk score across all the common variants and seeing how correlated this was with disease status (51). To do this they first had to define which allele was the risk increasing allele and estimate the effect of that allele on disease. They split their dataset in half and performed a GWAS on the first half to get these estimates. The second step was creating a risk score based on a weighted sum of risk alleles, in other words an individual's score is the sum of all risk alleles where each allele is weighted by the estimated effect size from the first part of the analysis. They showed that individuals with schizophrenia had a significantly higher risk score than controls, this difference persisted regardless of gender, population and even allele frequency. Based on simulation, they report an estimate of 30% of heritability being explained by common SNP variation contained on these genotyping arrays and importantly state that increasing sample sizes will improve power to identify a larger group of true disease variants. Much debate followed these finding including two particular arguments regarding other possible explanations, the first being that cryptic population stratification was biasing these results and the second that these common variants were actually tagging regions containing multiple rare variants (53). One study attempted to address the population stratification argument by performing the same risk scoring approach in family data that is absent of differences in population frequencies. In a dataset of nearly 700 trios they created pseudo cases and controls from the

transmitted chromosomes and the un-transmitted chromosomes, respectively, under the model that the child carries the disease and therefore should have been transmitted a higher proportion of risk alleles. Again, risk allele and effect sizes were estimated from a large, independent GWAS dataset and the risk scores were compared between the transmitted and un-transmitted chromosomes and again a significant difference was observed (54). Synthetic associations have also been addressed both for and against in a trio of papers published in PLoS Genetics (55-57). Cases were presented for both sides but given the paucity of very rare variant data there is no empirical evidence for this type of scenario. Despite the lack of empirical rare variant data, efforts in the original ISC paper showing polygenic signal from from variants with high minor allele frequencies are inconsistent with the associations being driven by only rare variants. Given the evidence from GWAS for common SNPs' contribution to disease risk, efforts were taken to more accurately assess the amount of heritability explained by these variants. Multiple methods have been developed to specifically calculate the contribution of common SNPs to heritability and have been applied to multiple phenotypes including height (58, 59). Initial application of these approaches to SCZ have yielded similar results to the estimates from the ISC polygenic paper with common SNPs contributing 23% of heritability (60). While much work has been dedicated to understanding the contribution of common SNPs to SCZ, recent technological improvements in sequencing have allowed for a more comprehensive survey of rare single nucleotide variants (SNV) and small insertions and deletions (indels). How much these variants contribute to disease risk is still an ongoing debate.

1.2.4 Sequencing / rare variants

The distinction between variants considered "rare" versus those considered "common" is mostly based on both technological and sample size limitations. Our ability to test variants is dependent on our ability to identify them and survey them in large enough numbers to statistically test them for association. For example a 1% variant in 10 people we would be unlikely to see and therefore unable to test but in 1,000,000 people we'd see it around 10,000 times and have great power to test. Therefore the word "rare" has an oft changing definition as technology improves and sample sizes get larger. It is likely that variants across the frequency spectrum will contribute to disease risk and making a distinction between "rare" and "common" is unhelpful (61). However, assessing the contribution of variants in an unexplored category of variation is important and necessary.

As with SNPs and CNVs, advancements in technology and methods were required before questions of role in diseases could be adequately addressed. Both common SNPs and rare CNVs have be shown to contribute to disease risk in SCZ however most variation is rare and large CNVs represent only a very small proportion of the overall class of rare variation. With advances in sequencing, it was possible to finally delve into the space of very rare variation in sample of increasing size. Initial uses of sequencing were focused on sequencing regions with prior evidence often from GWAS hits. The idea underlying this approach is that a common variant of modest effect is in fact linked to one or multiple rarer variants of much

larger effect. One strategy for identifying these large effect rare variants was to sequence a relatively small number of individuals to identify possible "causal" variants and then genotype those variants in a larger set of individuals thus increasing power and reducing costs. Several studies focusing on candidate genes found limited success in identifying large effect loci raising questions regarding this hypothesis (27, 62). As with the initial candidate GWAS studies, an assumption free sequencing study would be required. The first of these applied a similar strategy to the candidate gene approach in that they sequenced the whole genome or exome of 166 individuals, selected variants with the strongest association to disease and followed them up by genotyping in 2,617 cases and 1,800 controls. This approach yielded no single variant or gene demonstrating statistical significant to SCZ.

As with CNVs, one analysis approach that centers around identifying particularly rare mutations of likely functional impact is looking in families for *de novo* events, that is, variants arising in the new generation and not inherited. There has been strong evidence for the role of this class of mutation in other psychiatric diseases in particular in autism where multiple studies have shown an increased rate of disruptive *de novo* mutations above expectation and above sibling controls (63-66). Three initial papers in SCZ have identified similar trends, implicating a similar role of *de novo* single nucleotide variants (SNV). The first two utilized small samples of 14 and 53 trios and found significantly higher rates of non-synonymous mutations compared to expectation or a small set of controls. The third paper reported on a set of 231 SCZ trios comparing to 34 unaffected trios. They again identify a significant excess of nonsynonymous *de novo* SNV compared to the control trios, and in addition

they identify to 4 genes with recurrent *de novo* mutations. There are two larger, more recent studies that have yet to be published on exome sequencing in SCZ. The first one is another family study comprising of over 600 trios from Bulgaria, the second is a case/control study of more than 5,000 Swedish samples (the dataset used in much of this thesis). The results from both of these studies indicate a weaker role of very rare variants to SCZ risk compared to autism and intellectual disability. Both studies find no overall excess of disruptive mutations in cases either *de novo* or inherited, in fact only when reducing the search space to genes previously identified as being implicated in SCZ does a modest excess of very rare disruptive mutations exist. These gene sets consist of both biologically relevant sets such as post-synaptic density genes and/or genes previously associated through GWAS or CNV analyses. It has become clear that the polygenicity seen in the common variant results will remain across the spectrum of allele frequency. Thus, approaches to both maximize power both by aggregating over variants of different types and over biologically related genes will be required to provide biologically interpretable results and further our understanding of SCZ. It is with this mindset that this thesis both focuses on analyzing multiple classes of variation and testing biological hypotheses.

1.3 Pathway analysis approaches and findings in SCZ

Assigning genes to biological functions and testing them as pathways for enrichment of mutations in cases has become a common approach to increasing power and aiding understanding of disease biology. There are several strategies for

performing these types of tests and numerous methods have already been developed. A number of decision points exist in developing a pathway analysis method from determining the appropriate test statistic to deciding how to group your genetic data. Often, sets of genes are the unit of analysis but do you define a gene as only its coding regions or do you include intronic or promotor regions? How does one account for the differing sizes of genes and the relationship between gene size and number of identified variants? As described in great detail (67), there are two categories of pathway tests, competitive tests and self-contained tests. Briefly, competitive tests assume the null hypothesis of average association and look to identify pathways that are more significantly enriched than average; self-contained tests assume a null hypothesis of no association and look to identify pathways with any association. Both of these approaches have been applied to genome-wide SCZ data with mixed results. A study looking at rare CNVs in 2008 performed a pathway analysis looking for pathways overrepresented for being hit by CNVs (28). They identified several significantly overrepresented pathways including those involving neuronal activities and synaptic functions among others. Another study looking specifically at *de novo* CNVs in SCZ probands identified enrichment of CNVs hitting specific subunits of the post-synaptic density complex (38). I have utilized these findings in my selection of pathways to test in this thesis. Many methods have also been developed and applied to GWAS data (68-70), however given the many variants of small effects and the relationship between gene size, number of variants and dependence across these variants the results have been less conclusive and often do not replicate (71-73). I present only a subset of the methods and results for pathway

analysis in SCZ, this is a burgeoning area of research and one in which I delve deeper into throughout this thesis.

1.4 What's in this thesis

This thesis approaches the goal of a more complete assessment of variation contributing to SCZ risk by three principles; incorporating previously unanalyzed variation, developing methods suited to highly complex disease models and identifying and testing biologically relevant sets of genes to aid understanding of the disorder. As the chapters progress, the pathways used progress as the literature has advanced. I start by looking specifically at rare CNVs but incorporating the class of smaller 20kb-100kb events that have previously been unexplored. The second chapter introduces a method to test gene sets by explicitly estimating polygenic risk and identifying sets that do a better job discriminating risk between case and control. The third chapter introduces autozygosity as a mechanism for SCZ risk and looks for the contribution of rare disruptive recessive and compound heterozygous variants to overall risk of SCZ. Finally, I devise an analytical approach for testing variants of any length and given the combined power of rare deletions and rare disruptive mutations aim to identify a particular gene that is strongly associated with SCZ which has been a difficult task to date.

Chapter 2: Comprehensive analysis of the role of CNV in schizophrenia

2.1 Introduction

Schizophrenia (SCZ) is a chronic debilitating psychiatric illnesses that affects ~0.4% of the world's population (18, 74). Structural variation in the form of large, rare deletions and duplications has previously been shown to play a role in psychiatric disease (33, 75-78). Structural variation by definition is any large (>1kb) genetic alteration. These include inversions, translocations, deletions, duplications and insertions. Genotyping microarrays have enabled researchers to accurately identify both duplications and deletions from extended segments of increased or decreased intensities along the genome. In addition, the most recent versions of these arrays contain copy number specific probes designed to detect CNVs. Since CNVs affect genetic structure and can either duplicate or delete large regions of the genome, they are likely to have phenotypic consequences especially if affecting a gene or multiple genes. Examples of severe CNV events include trisomy 21 (duplication of chromosome 21), which results in Down's syndrome. On average, patients with schizophrenia carry a 15% higher rate of copy number variation than control individuals with no psychiatric diagnosis (33). CNVs overlapping genes, thereby deleting or duplicating part of or an entire coding region of a gene, are carried at even higher rates in SCZ cases than controls (Table 2.1a). CNV type, size and frequency all impact case enrichment with rare events being more significant

and larger deletions having a greater difference in rates of CNVs between schizophrenia patients and controls (33).

Several large CNVs have been shown to have very high penetrance with regard to disease status of SCZ (Table 2.1b). These include both deletions and duplications and have been discovered from multiple studies of differing population samples (79). Many of these CNVs have been implicated in syndromes with complex phenotypes where psychosis is just one common symptom. For example, Velo—Cardio-Facial Syndrome (VCFS) results from a 22q11 microdeletion and presents with psychosis in nearly 30% of carriers (80). CNVs of this size are bound to disrupt many genes of which only a subset represent likely SCZ candidate genes. Looking at smaller CNVs and sets of genes may help identify which genes are indeed relevant for SCZ. In this chapter, I detail CNV loci significantly enriched in individuals with SCZ in a large dataset of 4,556 SCZ and 5,969 matched controls. I address whether certain classes of CNVs are particularly enriched and whether I can identify particular genes or biologically grouped sets of genes related to the genetic etiology of SCZ.

Table 2.1: a. Previously reported burden in SCZ rates by type from International Schizophrenia Consortium

a.

ISC >100kb	SCZ DEL	CONTROLS DEL	P DEL	SCZ DUP	CONTROLS DUP	P DUP
Rate	0.43	0.40	1.1E-01	0.71	0.59	**2.0E-05**
Genic rate	1.12	0.72	**3.0E-05**	2.54	1.94	**1.0E-04**

b. Previously reported CNV risk loci in SCZ

Locus	#Genes	CNV type	Freq	OR
Replicated significant associations from case-control studies				
1q21.1	~10	Deletion	0.23-0.32	6.6-14.8
15q13.3	~10	Deletion	0.17-0.3	11.5-17.9
16p11.2	>25	Duplication	0.3	8.3-25.4
22q11.2	>25	Deletion	0.5-2.0	30
Significant association reported in a single cohort				
2p16.3	NRXN1	Deletion	0.47	9.00
15q11.2	~10	Deletion	0.55	2.73
17p12	~10	Deletion	0.13	7.80
16p13.1	~14	Duplication	0.30	3.30
1q42.2	DISC1	Balanced translocation	NA	NA

2.2 Methods

2.2.1 Sample collection / dataset creation

This study combines individual genotyping data of two unpublished datasets including 1,436 individuals from Sweden with previously published or publically available data. These data include previously published GWAS of SCZ, ISC (51), CATIE (81), WTCCC SCZ (48) and GAIN/MGS (50). All analyses were performed using Plink (82). Individuals were placed into 11 groups based on the site and date when they were collected (Table 2.2, Figure 2.1). The novel Swedish sample is comprised of 539 SCZ individuals and 905 control individuals.

Figure 2.1: First two components of multi-dimensional scaling analysis, colored by sample. There are four independent clusters representing different ancestries the top left (yellow, black) represents the UK and Irish samples, the top right (red) are the Portuguese, the bottom right (black) are the Bulgarians and the bottom left (purple, light blue) are the Swedish. The more spread samples (green, grey) represent the samples from the US with more mixed ancestry.

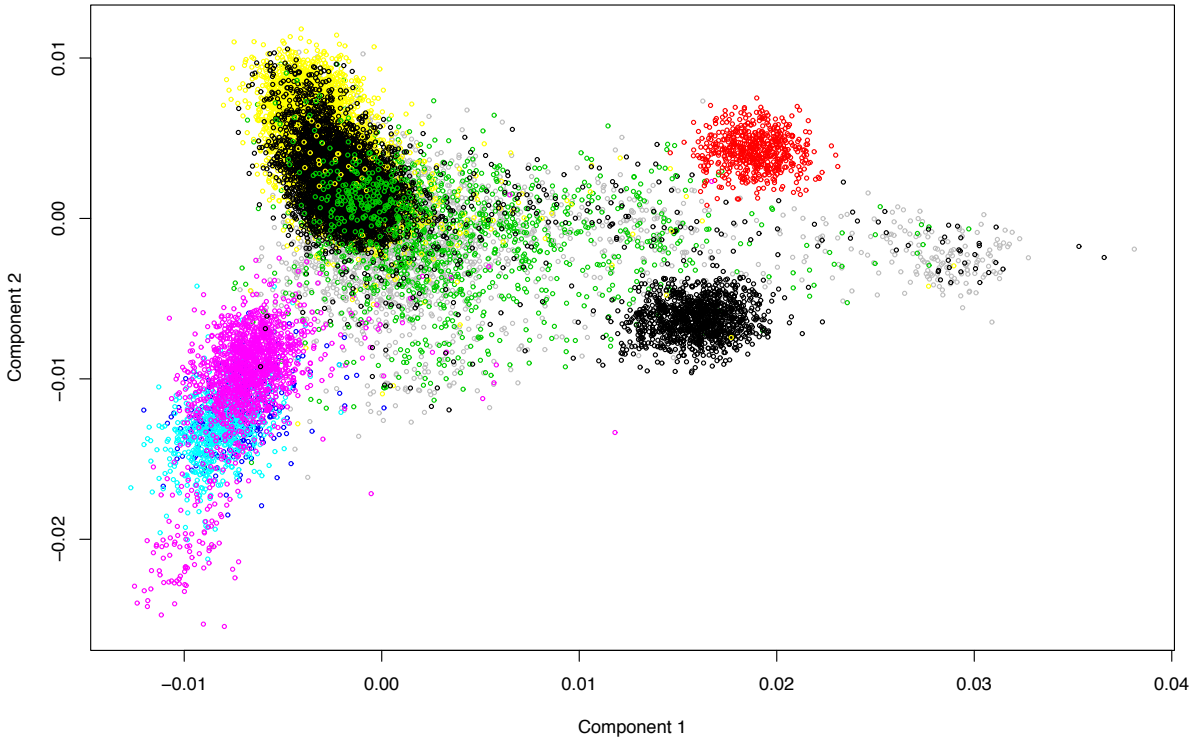
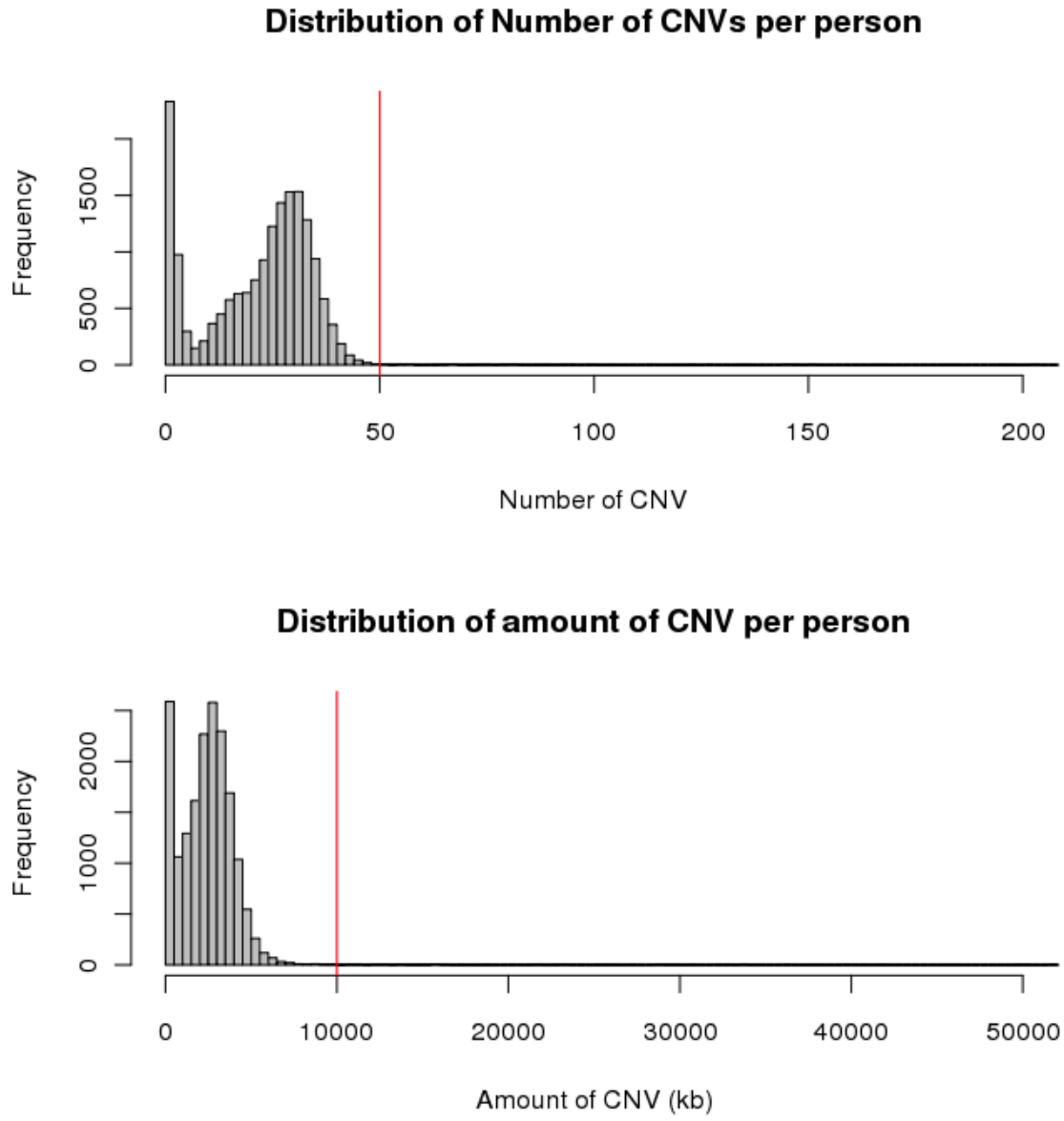


Table 2.2. SCZ samples genotyped in this study and passing CNV QC including ancestry and array used.

Ancestry	SCZ Sample	Cases	Controls	Affymetrix Array
British	University College London (UCL)	461	468	5.0
European-American	Genetic Association Information Network (GAIN) / Molecular Genetics of Schizophrenia (MGS)	1074	1312	6.0
Irish	Trinity College Dublin	254	809	6.0
Scottish	University of Aberdeen	690	663	5.0
Scottish	University of Edinburgh	350	278	6.0
Portuguese	Portuguese Island Collection	286	181	5.0
Swedish	Karolinska Institute	144	137	5.0
Swedish	Karolinska Institute	362	223	6.0
Swedish	Karolinska Institute	532	897	6.0
Bulgarian	University of Cardiff	403	558	6.0
Total		4556	5969	

Rare copy number variants (CNV) were assessed using the Birdsuite package (32) in particular Birdseye which is the program specifically designed to call rare CNVs. Birdseye works on a per individual basis by comparing hybridization intensities at each SNP to a prior for all diploid genotypes, diploid priors are estimated from the data and priors for other copy numbers are inferred. It then assesses the likelihood of there being a copy number event at any particular SNP and uses a Hidden Markov Model (HMM) to call the extent of that CNV. I performed the following QC procedure, in order to prevent outliers contributing too much to burden statistics I removed individuals at the end of the distribution with regards to number or amount of CNVs (>50 CNVs or >10MB, Figure 2.2), I reduced the set to only rare CNVs (those seen in less than 1% of the sample), I removed known HapMap CNVs and regions of known rearrangements. One major change from previous work was retaining all CNVs greater than 20kb as opposed to 100kb. In previous validation of CNVs in this sample across sites and chips we (validation done with assistance by Kim Chambert) identified a validation rate of 96.7% for CNVs greater than 100kb and 92.9% in CNVs between 20kb and 100kb. I determined this was acceptable for inclusion in analysis. Each dataset was cleaned individually to account for date processed and technology used. After filtering, there were 4,566 SCZ individuals and 5,969 controls.

Figure 2.2. CNV parameters used to determine outliers in sample, the red line indicates the selected cutoff for outliers.



2.2.2 Locus specific analysis

I looked to discover regions of the genome significantly enriched for case CNVs using two methods, both utilizing the Plink software package (82). The first method creates 100kb regions shifted every 20kb to cover the entire genome. For each region, I calculated the number of case CNVs and the number of controls CNVs overlapping any part of the region. I then permuted case/control label of individuals 10,000 times accounting for site to determine the empirical p-value, in addition I applied a multiple testing correction by empirically determining the p-value at the fifth percentile across all tests and asking how often the permuted p-value is equal to or more significant than this value. This approach allowed me to map significance across the genome and assess regions of interest beyond discrete points. It also accounts for the variability in break points called using Birdseye. One issue regarding CNV calling from SNPs is the inaccuracy of determining breakpoints, as the true breakpoint is likely to not fall exactly at the location of a SNP. While the HMM generally performs correctly in calling CNVs, the exact end points are subject to more uncertainty and often result in variable end points across samples. By taking large genomic regions for testing I minimized the role this variability will play in my results. The second method employed the same permutation procedure except instead of defining 100kb windows I used the transcription start and end positions to define the window used in the analysis. All RefSeq transcripts (hg18) were included where I defined each gene as the beginning of the first transcript to the end of the last transcript, thereby making the largest possible transcript.

2.2.3 Burden analysis

I sought to identify whether individuals with psychiatric disease carried a "burden" of either a larger number of CNVs or a larger number of genic CNVs than controls. Burden analysis was performed as described in ISC CNV manuscript (33), that is I computed summary statistics of two values: the average number of CNVs per person (RATE), the average number of genes being hit by CNVs per person (GRATE). I then took this information and compared it to 10,000 permutations of phenotype label being careful to only permute with each of the 11 sites independently.

2.2.4 Enrichment analysis

Enrichment analysis is performed by counting the number of CNVs overlapping any given set of genes or regions per individual. I then compared these counts between cases and controls after controlling for total number and amount (kb) of CNVs per person and the sample site of that individual by including these variables into a logistic regression framework (83). Permutation is performed on case/control label 10,000 times to compute p-values.

I initially selected 5 gene sets that have been previously implicated in Schizophrenia. I chose these sets to give higher probabilities of identifying a true signal given the literature. These included ***brain expressed***, defined from expression profiles of 96 tissues. 27 tissue profiles that represented brain or spinal cord were compared to the remaining 69 tissue profiles, genes obtaining $p<0.01$ as preferentially expressed with a one tailed Mann-Whitney rank-sum test were

included (83). ***Neuronal activity,*** defined by Panther (84) category "Neuronal Activities" (BP00166) and implicated in SCZ (28). ***Learning,*** defined by Ingenuity category 'Behavior--- Learning' and implicated in BP (85). ***Synapse***, defined by Gene Ontology (86) code 'Synaptic Junction' (GO: 0045202) and a set of previously identified psychiatric candidate genes (**psychiatric hits**).

2.3 Results

2.3.1 Overall rates of CNV by size

I initially wanted to compare overall rates of number of CNVs and number of CNVs overlapping genes between SCZ cases and controls (Table 2.3a). Similar analyses, albeit in a subset of these data have found significant increases in both of these measures (33). I saw significantly higher rates of genic CNVs in both deletions and duplications but only higher overall rates of CNVs in duplications that is akin to what's been previously seen (33)(Table 2.3b). I sought to better understand the contribution of smaller CNVs to this result by splitting the CNVs into 3 categories: All CNVs (>20kb), large CNVs (>100kb) and small CNVs (20kb – 100kb). The small CNV class contributes the majority of the CNVs 16,148 of 26,609 (61%) but showed no significant increase in overall CNV rate or genic CNV rate (Table 2.3b).

Table 2.3. a) Genome-wide CNV burden test values stratified by type and disease

a) Schizophrenia burden rates (All CNVs > 20kb)

Test	Site	SCZ DEL	CONTROL DEL	SCZ DUP	CONTROL DUP
N	All	5147	6284	5518	6462
RATE	All	1.33	1.29	1.43	1.33
GRATE	All	1.17	0.88	2.16	1.92
	scot6	1.35	1.50	1.68	1.54
	gain	1.40	1.19	1.47	1.29
	ucl	0.96	0.87	0.98	0.72
	dub	1.28	1.43	1.48	1.43
RATE	sw3	1.25	1.26	1.32	1.38
	sw1	0.93	0.72	0.77	0.76
	sw2	1.45	1.46	1.64	1.60
	card	1.86	1.79	1.76	1.69
	port	1.17	1.09	1.43	1.14
	scot6	0.91	0.89	2.79	1.76
	gain	1.27	0.94	2.11	1.94
	ucl	0.80	0.53	1.41	1.33
	dub	1.04	0.88	2.24	2.00
GRATE	sw3	1.42	0.91	2.02	2.03
	sw1	0.65	0.53	1.51	1.31
	sw2	1.27	1.10	2.27	2.39
	card	1.37	0.99	2.55	2.25
	port	1.13	0.75	2.64	1.59

b) Genome-wide CNV burden p-values stratified by type, disease and size

	#CNVs		SCZ Del	SCZ Dup
		Rate	0.2661	**0.0020**
>20kb	*26609*	Genic rate	**0.0001**	**0.0129**
		Rate	0.1071	**0.0020**
>100kb	*10461*	Genic rate	**0.0001**	**0.0111**
		Rate	0.6113	0.2321
20kb - 100kb	*16148*	Genic rate	0.2205	0.7708

2.3.2 Gene enrichment in brain related gene sets

I next tested whether particular gene sets, selected for having biological ties to brain related functions (see methods) were more likely to be hit by CNVs above the overall increase in genic CNVs in SCZ cases. To account for my previous result of case CNVs more likely overlapping genes, I limited the CNVs in this analysis to only those overlapping genes. I tested the full set of CNVs > 20 kb and found that the largest set of brain expressed genes showed no significant enrichment in either deletions or duplications (Table 2.4). The other four sets all showed some level of significance in one or the other with the neuronal activity set being the only one significant in both. The most significant result was enrichment of duplications in the synapse set (p-value < 0.0007, most significant gene *DOC2A* 14:1 p=0.002)

Table 2.4. P-values from genome-wide CNV enrichment tests in sets of genes previously implicated for role in SCZ (CNV > 20kb)

Test	Genes (n)	Del (n)	Dup (n)	P SCZ Del	P SCZ Dup
Brain expressed	2433	1945	2755	0.0864	0.1908
Neuronal activity	453	412	681	**0.0011**	**0.0138**
Psychiatric hits	126	349	284	**0.0439**	0.4875
Learning	141	333	279	**0.0270**	0.4798
Synapse	208	282	324	0.1023	**0.0007**

2.3.3 Locus specific analyses

I looked to identify regions of the genome with greater numbers of case CNVs compared to control CNVs. I divided the genome into 100kb windows where each successive window was shifted 20kb. I selected 100kb to allow for the average gene size (~60kb) and 20kb of either side for promoter regions. I further created

transcript specific regions for each gene in RefSeq, more specifically looking for genic disruption. I performed both approaches to ensure I covered the regions of the genome with not genes but also get a more refined assessment of genic regions.

At a nominally significant p-value < .05, there were 532 deletion regions and 450 duplications regions corresponding to 53 independent deletion regions across the genome and 63 independent duplication regions (Tables 2.5-2.6). I defined independence as being at least 100kb between nominally significant regions. The number of non-independent 100kb segments and the most significant p-value across them is presented in Tables 2.5 and 2.6. Only the previously implicated region on 22q11.2 is genome-wide significant in this analysis (p-value corrected 0.002; 19 cases and 0 controls). I further restricted my analysis to genes defined by transcription boundaries and identified 71 genes with p-value < 0.05 in deletions and 67 in duplications (Tables 2.7-2.8). Many of these genes were identified in the genomic region analysis including 43 the 71 genes within the VCFS deletion being significant after correction. While my most significant genomic region chr16:29700000-29800000 after correction was not genome-wide significant (p-value = 0.055) when I defined regions by genes several of the genes in this region are genome-wide significant further implicating genes as a driving reason for the overall CNV enrichment. However, 12 of the 63 duplication regions and 19 of the 53 deletion regions contained no genes at all (based on RefSeq list of 19,058 genes).

Table 2.5: Independent genomic regions with nominal significant enrichment of case deletions,
#regions represents the number of overlapping significant 100kb windows

Chr	Start	End	Min P	#Regions	Genes
1	4940000	5160000	0.0319	7	
1	49540000	49640000	0.0290	1	AGBL4
1	72640000	72780000	0.0051	3	
1	119820000	119940000	0.0198	2	HSD3B1
1	144960000	146160000	0.0041	54	ACP6 BCL9 CHD1L FMO5 GJA5 GJA8 GPR89B GPR89C LOC728912 NBPF11 PRKAB2
1	146300000	146460000	0.0121	4	PPIAL4
1	192480000	192680000	0.0059	6	
2	44560000	44700000	0.0234	3	C2orf34
2	54840000	54940000	0.0228	1	
2	97240000	97620000	0.0246	6	ANKRD36B
2	215120000	215340000	0.0121	7	BARD1 LOC402117
3	740000	840000	0.0377	1	
3	165620000	165820000	0.0010	6	
4	97460000	97600000	0.0179	3	
4	135520000	135640000	0.0349	2	
4	161220000	161380000	0.0381	4	
4	167420000	167640000	0.0390	7	
4	169240000	169460000	0.0376	7	ANXA10 DDX60
5	83740000	84040000	0.0448	11	
6	162620000	162860000	0.0119	7	PARK2
7	8600000	8760000	0.0226	4	NXPH1
7	14940000	15040000	0.0493	1	
7	38000000	38200000	0.0034	6	STARD3NL
7	42240000	42440000	0.0390	6	GLI3
7	64020000	64160000	0.0208	3	ZNF117 ZNF273
7	69180000	69380000	0.0221	6	AUTS2
7	110520000	110720000	0.0096	6	IMMP2L LRRN3
7	111060000	111160000	0.0447	1	DOCK4
8	13820000	14060000	0.0114	8	SGCZ
8	15620000	15760000	0.0429	3	TUSC3
8	15900000	16160000	0.0039	9	MSR1
9	6600000	6800000	0.0098	6	GLDC JMJD2C
9	41500000	41920000	0.0119	15	ZNF658B
9	68220000	68560000	0.0310	4	CBWD6 FOXD4L6
9	69240000	69440000	0.0143	6	
11	60200000	60320000	0.0209	2	MS4A10 MS4A15 MS4A8B
11	83960000	84200000	0.0067	8	DLG2
12	16220000	16380000	0.0191	3	
12	72500000	72680000	0.0108	5	
13	88520000	88720000	0.0156	6	
15	28420000	28540000	0.0092	2	CHRFAM7A
15	28880000	30740000	0.0003	76	ARHGAP11A CHRNA7 KLF13 MTMR10 MTMR15 OTUD7A SCG5 TRPM1
15	99880000	100120000	0.0183	8	TARSL2 TM2D3
16	15060000	15420000	0.0155	5	MPV17L RRN3
16	16200000	16420000	0.0074	7	ABCC6 LOC339047 NOMO3
16	17800000	18140000	0.0068	13	
16	79740000	79880000	0.0095	3	BCMO1 PKD1L2
18	56160000	56340000	0.0260	5	MC4R
19	32900000	33100000	0.0313	6	
20	14300000	14460000	0.0389	4	MACROD2
21	28020000	28360000	0.0470	13	
22	17160000	19920000	0.0001	134	AIFM3 ARVCF C22orf25 C22orf29 C22orf39 CDC45L CLDN5 CLTCL1 COMT CRKL DGCR14 DGCR2 DGCR6 DGCR6L DGCR8 GNB1L GP1BB GSC2 HIRA HTF9C KLHL22 LZTR1 MED15 MRPL40 P2RX6 PI4KA PRODH RANBP1 RIMBP3 RTN4R SCARF2 SEPT5 SERPIND1 SLC25A1 SLC7A4 SNAP29 TBX1 THAP7 TSSK2 TXNRD2 UFD1L ZDHHC8 ZNF74
22	32440000	32600000	0.0340	4	LARGE

Table 2.6: Independent genomic regions with nominal significant enrichment of case duplications
#regions represents the number of overlapping significant 100kb windows

Chr	Start	End	Min P	#Regions	Genes
1	5860000	6060000	0.0230	6	KCNAB2 NPHP4
1	17260000	17460000	0.0110	6	PADI1 PADI2 PADI3
1	120700000	120840000	0.0110	3	FCGR1B
1	143060000	143180000	0.0390	2	PPIAL4
1	143800000	143920000	0.0330	2	SEC22B
1	144740000	144960000	0.0270	4	FAM108A3 LOC728912 NBPF11
1	158800000	159000000	0.0370	6	CD48 CD84 SLAMF1 SLAMF7
1	244360000	244480000	0.0480	2	SMYD3
1	244740000	244860000	0.0300	2	C1orf71 TFB2M
1	246940000	247040000	0.0190	1	LOC646627
2	54380000	54500000	0.0470	2	ACYP2 C2orf73
2	73180000	73280000	0.0410	1	RAB11FIP5
2	86740000	87040000	0.0380	11	CD8A CD8B RGPD1 RMND5A
2	131680000	132080000	0.0150	13	A26C1A CCDC74A FAM128A LOC150786 TUBA3D
2	228320000	228460000	0.0420	3	CCL20 WDR69
3	50400000	50540000	0.0240	3	CACNA2D2
4	25200000	25340000	0.0010	3	SLC34A2
4	132580000	132680000	0.0410	1	
4	162240000	162360000	0.0210	2	
5	0	100000	0.0450	1	
5	4180000	4440000	0.0390	9	
5	37240000	37600000	0.0080	13	C5orf42 NUP155 WDR70
5	78240000	78340000	0.0290	1	ARSB DMGDH
5	110360000	110560000	0.0040	6	TSLP WDR36
6	141120000	141520000	0.0100	16	
6	141820000	142020000	0.0180	6	
6	162840000	163080000	0.0200	4	PACRG PARK2
7	32820000	33040000	0.0060	7	FKBP9 KBTBD2 NT5C3
7	69680000	69900000	0.0480	7	AUTS2
7	88020000	89060000	0.0180	45	MGC26647 ZNF804B
7	100480000	100740000	0.0010	9	AP1S1 C7orf52 CLDN15 FIS1 MOGAT3 MUC17 PLOD3 SERPINE1 TRIM56 VGF ZNHIT1
7	110880000	111220000	0.0130	13	DOCK4 IMMP2L
7	151320000	151420000	0.0190	1	GALNT11 GALNTL5
7	158040000	158740000	0.0010	27	FAM62B NCAPG2 PTPRN2 VIPR2 WDR60
8	11980000	12160000	0.0330	5	DUB3 FAM86B1 LOC728957
8	12760000	12860000	0.0370	1	C8orf79
8	42980000	43120000	0.0230	3	FLJ23356 FNTA HGSNAT HOOK3
8	43840000	44040000	0.0150	6	
8	46880000	47320000	0.0050	18	
8	53740000	54180000	0.0040	13	NPBWR1 RB1CC1
8	143200000	143380000	0.0120	5	TSNARE1
8	145060000	145180000	0.0290	2	GRINA OPLAH PARP10 PLEC1 SPATC1
9	20000	120000	0.0330	1	CBWD1 FOXD4
9	380000	480000	0.0480	1	DOCK8
9	800000	960000	0.0390	4	DMRT1
9	6880000	7000000	0.0330	2	JMJD2C
9	17480000	17580000	0.0460	1	CNTLN SH3GL2
10	14940000	15200000	0.0030	9	ACBD7 C10orf111 DCLRE1C HSPA14 MEIG1 NMT2 OLAH RPP38 SUV39H2
10	35260000	35400000	0.0460	2	CUL2
10	46960000	47520000	0.0280	24	ANXA8L2 FAM21B
10	51580000	51920000	0.0290	12	ASAH2 SGMS1
11	49980000	50420000	0.0160	12	
11	85220000	85360000	0.0280	3	CCDC83 PICALM
11	134240000	134540000	0.0060	11	
12	108080000	108240000	0.0160	4	ACACB FOXN4
14	21700000	22120000	0.0010	17	DAD1
14	93860000	94060000	0.0320	6	SERPINA11 SERPINA12 SERPINA1 SERPINA9
15	28880000	28980000	0.0370	1	
15	83500000	83600000	0.0130	1	
15	97660000	97840000	0.0370	5	LRRC28 ALDOA ASPHD1 BOLA2B BOLA2 C16orf53 C16orf54 CCDC95 CDIPT CORO1A DOC2A FAM57B FLJ25404 GDPD3 GIYD1 GIYD2 HIRIP3 KCTD13 MAPK3 MAZ MVP PPP4C PRRT2 QPRT SEZ6L2 SPN SULT1A3 SULT1A3 SULT1A4 SULT1A4
16	29380000	30180000	0.0010	36	TAOK2 TBX6 TMEM219 YPEL3
16	87400000	87540000	0.0200	3	APRT CBFA2T3 CDT1 GALNS LOC390748 TRAPPC2L
17	10020000	10140000	0.0480	2	GAS7

Table 2.7: Genes with significant enrichment of deletions in SCZ and if they are in any of the tested genesets

CHR	GENE	P	P corrected	Gene set
1	PRKAB2	0.00819918	0.305569	
1	FMO5	0.00819918	0.305569	
1	CHD1L	0.00819918	0.305569	
1	BCL9	0.0472953	0.936506	
1	ACP6	0.019698	0.554345	
1	GJA5	0.019698	0.554345	
1	GJA8	0.029697	0.728027	
1	NBPF11	0.039796	0.936506	
1	LOC728912	0.039796	0.936506	
2	C2orf34	0.00659934	0.962904	
4	KCNIP4	0.0475952	1	brain
4	ANXA10	0.0391961	0.9977	
4	GALNT17	0.0455954	1	
6	PARK2	0.0233977	0.9998	learn,psych
7	NXPH1	0.0252975	1	brain,neuro
7	HDAC9	0.0308969	0.872513	
8	MSR1	0.00749925	0.9977	
15	CHRFAM7A	0.0424958	1	
15	MTMR15	0.00239976	0.554345	
15	MTMR10	0.00239976	0.554345	brain
15	TRPM1	0.00159984	0.305569	brain
15	KLF13	0.00239976	0.554345	
15	OTUD7A	0.00239976	0.554345	brain
15	CHRNA7	0.00969903	0.9988	learn,psych,synap
15	TM2D3	0.019898	0.917908	neuro
15	TARSL2	0.0341966	0.962904	brain
16	NOMO3	0.00889911	0.554345	brain
16	LOC339047	0.00889911	0.554345	
22	DGCR6	1.00E-04	0.00079992	
22	PRODH	1.00E-04	0.00079992	brain
22	DGCR2	1.00E-04	0.00029997	
22	DGCR14	1.00E-04	0.00029997	
22	TSSK2	1.00E-04	0.00029997	
22	GSC2	1.00E-04	0.00029997	
22	SLC25A1	1.00E-04	0.00029997	
22	CLTCL1	1.00E-04	0.00019998	neuro
22	HIRA	1.00E-04	0.00029997	
22	MRPL40	1.00E-04	0.00029997	
22	C22orf39	1.00E-04	0.00029997	
22	UFD1L	1.00E-04	0.00029997	
22	CDC45L	1.00E-04	0.00029997	
22	CLDN5	1.00E-04	0.00029997	
22	5-Sep	1.00E-04	0.00029997	
22	GP1BB	1.00E-04	0.00029997	
22	TBX1	1.00E-04	0.00029997	
22	GNB1L	1.00E-04	0.00029997	
22	C22orf29	1.00E-04	0.00029997	
22	TXNRD2	1.00E-04	0.00029997	
22	COMT	1.00E-04	0.00029997	neuro,learn,psych
22	ARVCF	1.00E-04	0.00029997	
22	C22orf25	1.00E-04	0.0164984	
22	DGCR8	1.00E-04	0.00019998	
22	HTF9C	1.00E-04	0.00019998	
22	RANBP1	1.00E-04	0.00019998	
22	ZDHHC8	1.00E-04	0.00029997	
22	RTN4R	1.00E-04	0.00079992	brain
22	DGCR6L	1.00E-04	0.00029997	
22	RIMBP3	1.00E-04	0.00429957	
22	ZNF74	1.00E-04	0.00339966	
22	SCARF2	1.00E-04	0.00339966	
22	KLHL22	1.00E-04	0.00339966	
22	MED15	1.00E-04	0.00119988	
22	PI4KA	1.00E-04	0.00339966	brain
22	SERPIND1	1.00E-04	0.00339966	
22	SNAP29	1.00E-04	0.00339966	synap
22	CRKL	1.00E-04	0.00339966	
22	AIFM3	1.00E-04	0.00339966	brain
22	LZTR1	1.00E-04	0.00339966	
22	THAP7	1.00E-04	0.00339966	
22	P2RX6	1.00E-04	0.00339966	neuro
22	SLC7A4	1.00E-04	0.00339966	

Table 2.8: Genes with significant enrichment of duplications in SCZ and if they are in any of the tested genesets

CHR	GENE	P	P corrected	Gene set
1	NPHP4	0.0224	1.0000	
1	SEC22B	0.0355	1.0000	
1	KIF26B	0.0370	1.0000	
1	SMYD3	0.0283	0.9982	brain
1	SH3BP5L	0.0488	0.8540	
1	ZNF672	0.0488	0.8540	
1	PGBD2	0.0458	0.8540	
2	FAM110C	0.0279	0.8540	
2	RMND5A	0.0345	1.0000	
2	CD8A	0.0345	1.0000	
2	CD8B	0.0345	1.0000	
2	TUBA3D	0.0229	0.9871	brain
2	FAM128A	0.0229	0.9871	
5	WDR36	0.0041	0.5862	
5	TRIM7	0.0416	0.9871	
7	ZNF804B	0.0289	0.9935	
7	MGC26647	0.0435	0.9982	
7	SERPINE1	0.0047	0.4145	
7	AP1S1	0.0047	0.4145	brain
7	VGF	0.0047	0.4145	brain
7	C7orf52	0.0047	0.4145	
7	MOGAT3	0.0047	0.4145	
7	PLOD3	0.0047	0.4145	
7	ZNHIT1	0.0047	0.4145	
7	CLDN15	0.0047	0.4145	
7	FIS1	0.0047	0.4145	
7	IMMP2L	0.0087	0.9871	
7	GALNT11	0.0309	0.9871	
7	MLL3	0.0422	1.0000	
7	NCAPG2	0.0166	0.7630	
7	FAM62B	0.0035	0.2145	
7	WDR60	0.0012	0.1001	
8	FLJ23356	0.0237	1.0000	
8	HGSNAT	0.0428	1.0000	
8	TSNARE1	0.0203	0.9871	
9	DMRT1	0.0438	1.0000	
9	NOTCH1	0.0344	1.0000	
10	DCLRE1C	0.0061	0.8566	
10	MEIG1	0.0061	0.8566	brain
10	ANXA8L2	0.0367	0.9982	
10	FAM21B	0.0367	0.9982	
10	SGMS1	0.0430	1.0000	
16	SPN	0.0005	0.1033	
16	QPRT	0.0005	0.1033	
16	C16orf54	0.0003	0.0243	
16	MAZ	0.0003	0.0094	
16	PRRT2	0.0003	0.0094	
16	C16orf53	0.0003	0.0094	
16	MVP	0.0003	0.0094	brain
16	CDIPT	0.0003	0.0094	
16	SEZ6L2	0.0003	0.0094	brain
16	ASPHD1	0.0003	0.0094	brain
16	KCTD13	0.0003	0.0094	brain
16	TMEM219	0.0004	0.0233	
16	TAOK2	0.0004	0.0233	
16	HIRIP3	0.0004	0.0233	
16	CCDC95	0.0004	0.0233	
16	DOC2A	0.0004	0.0233	brain,neuro,synap
16	FLJ25404	0.0004	0.0233	
16	FAM57B	0.0004	0.0233	brain
16	ALDOA	0.0004	0.0233	
16	PPP4C	0.0004	0.0233	
16	TBX6	0.0001	0.0034	
16	YPEL3	0.0001	0.0034	
16	GDPD3	0.0002	0.0078	
16	MAPK3	0.0002	0.0078	brain
16	HYDIN	0.0453	1.0000	

2.4 Discussion

In this chapter I looked to further assess the role of CNVs in SCZ, I verified a previously seen result of increased rates of CNVs in SCZ cases compared to controls and an even greater increased rate of CNVs overlapping genes in SCZ cases compared to controls. I divided CNVs into different sized bins including all CNVs > 20kb, CNVs used in previously analyses (>100kb) and a set of smaller CNVs (20kb – 100kb) and discovered that the larger CNVs were driving the result. I further looked to refine this signal beyond all genes by testing sets of genes identified for their relationship to brain function. Of the five sets that I tested, I identified several that were enriched even after accounting for the genic enrichment, pointing to likely set of disease relevant genes.

Previous literature has identified rare CNVs as being a strong contributor to risk of SCZ. As technology gets better, the resolution at which CNVs can be called will improve. An open question is whether smaller CNVs (<100kb) are contributing as strongly to SCZ risk as larger CNVs. All of the microarrays used in this study carry nearly one million probes across the genome and have demonstrated high accuracy in calling CNVs down to 20kb. Most studies to date have focused on larger CNVs (100kb) but there is an exponentially increasing number of events as CNV size gets lower. In this sample, 61% of the CNVs are between 20kb and 100kb and I saw no higher rates of CNVs or genic CNVs in cases in that size range. There exists a number of possible reasons for this result: 1) despite better resolution and more probes it is still easier to detect larger CNVs and therefore there will be much more noise in

terms of false calls in the smaller CNVs, 2) larger CNVs are more likely to hit multiple genes and it is the combination of many genes being affected that actually increases the risk to SCZ, 3) there could be a subset of genes that if affected by a CNV will increase risk to disease and these genes are more likely to be hit by larger CNVs. In the last model we might expect to be able to "fine map" which genes are relevant by looking for smaller CNVs piling up in a particular gene within a significant larger CNV.

I tested for significant overlap of CNVs in two ways, regionally across the genome and specific to particular genes. I identified only a single genomic region reaching genome-wide significance, a deletion known to cause VCFS of which 30% of carriers have psychosis. I looked for a refined signal of smaller CNVs pointing to a specific gene and found only one additional deletion in the region increasing case overlap but even that deletion contained 4 genes. My results cannot be refined to any single gene relevant to SCZ.

The absence of gene specific pileup of CNVs does not preclude identifying a set of genes with similar function or within the same "pathway" as being relevant to SCZ. In this scenario, we would expect to see an enrichment of CNVs in cases compared to controls in aggregate across a set of genes. By combining genes with similar biological function we both increase our power to detect higher CNV rates in cases but also imply mechanism of disease. Here, I tested 5 brain related sets of genes. Because I was most interested in set specific results I looked only for enrichment above the genic CNV enrichment seen previously. I identified the most significant enrichment in duplications in genes related to synaptic junction.

Mutations that affect this mechanism are possible candidates for SCZ as synaptic pruning is known to continue further and last longer in individuals with SCZ. Additionally, I see significantly higher rates of CNVs in genes related to neuronal activity and learning. While I selected these sets for having been previously implicated in SCZ, expanding this approach to a broader set of gene sets will aid in the biological understanding of SCZ.

Chapter 3: Identifying disease related pathways from common GWAS data

3.1 Introduction

3.1.1 Polygenic signal in Schizophrenia

Disease loci contributing to the genetic risk of complex disorders such as schizophrenia (SCZ) come in various forms from single base pair changes to large genomic alterations (33, 52, 79). These loci will range in effect size but many will remain undetected by single locus analyses. Recent literature has shown that there exists an additive "polygenic" signal captured by the common variants in GWAS data (51). More specifically, risk-increasing alleles can be identified and effect sizes can be estimated from a GWAS analysis. These GWAS results can be taken into an independent genotyping dataset and a risk score can be calculated for every individual by summing up the number of risk alleles weighted by their effect size. This scoring procedure when applied to SCZ data yields significantly higher risk scores in cases compared to controls implicating a true additive polygenic component to SCZ. Alternative explanations for this polygenic result remain. Some have argued that this effect and many of the GWAS results are due to difficult to assess frequency differences across populations (53). The reasoning is that because common variants have survived many generations of selection they must not be related to diseases that affect fitness. Efforts to address this issue have applied the same scoring procedure to trio data removing the element of population

stratification but looking only within a family. They found increased risk scores in chromosomes transmitted to a proband with SCZ compared to the untransmitted chromosomes (87), thus demonstrating this polygenic risk exists in the absence of population stratification.

Identifying a polygenic component correlated with disease risk does not implicate any particular gene or biological function. It will take a deeper understanding or refinement of this signal to learn disease biology. Assessing the contributions of particular genes or sets of genes to this signal could contribute to understanding the genetic etiology of SCZ without pointing to any particular variant of strong effect.

3.1.2 Pathway tests using GWAS data

In the absence of large effect loci pointing to a particular variant or gene, a more cumulative approach is required. Under a model of many small effects and additive liability, summing effects across variants within a gene and across genes will result in increased power to detect the combined effect of all variants in those genes. One logical approach to determining how to define these sets of genes is to focus on shared biological function. Here, the hypothesis is that affecting the function of any gene of a given set with similar biological role could have the same consequence in terms of disease risk. Another approach is to take a set of genes previously implicated in the disorder from an independent study or in a different class of variation and test as a set. This approach is more akin to replicating enrichment in a

geneset across datasets or across variant classes. Given a defined set of genes the question is then to assess whether variants in these genes are at increased frequencies in cases compared to controls. There are a number of ways to approach this "pathway" or "geneset" question as seen by multiple methods that currently exist as well as those described in more comprehensive reviews (67-70). Defining genes, as in whether to include introns or promoter regions, dealing with non independence across markers (linkage disequilibrium), deciding which variants to include, appropriately weighting variants and identifying the appropriate test statistic are just a few of the issues in designing a method for testing pathway. These issues aside, a well-powered pathway test could provide novel insights into the underlying disease biology.

Here, I present a new method of testing for enrichment of genetic signal in pathways that leverages the polygenic risk component already identified in SCZ. I assess the statistical properties of method, test a set of pathways specifically selected for having previous literature implicating them to SCZ risk, interpret the results and finally attempt replication in a larger set of pathways from the Gene Ontology Project (GO).

3.2 Methods

3.2.1 Sample collection / dataset creation

This chapter utilizes the same datasets discussed previously in chapter 2 (Table 2.4) with two additions. Here, we include the WTCCC SCZ (48) and CATIE (81) samples for which we did not have CNV calls. Individuals were placed into 11 groups based on the date and location they were collected (Table 3.1). I created a single Scottish sample by combining samples from the University of Aberdeen and the University of Edinburgh. This set was combined to bring the ratio of cases and controls closer to one which increases power to detect associated variants. I cleaned each of the eleven samples independently by removing SNPs with: MAF < 1%; genotyping rate below 95%; significantly different genotype frequencies than expected under Hardy-Weinberg equilibrium (P<1e-6); and significant differential missingness (P<0.05) between cases and controls. Individuals were removed for: having genotyping rates below 90%; high relatedness to any other individual or low relatedness to many other individuals; substantially increased or decreased heterozygosity across the genome; and being population outliers based on 10 MDS components (Figure 2.1). The differences in case/control counts between Table 2.4 and Table 3.1 represent the different QC procedures required for CNV and SNP data.

After cleaning, there were 14,778 total samples including 5,670 SCZ cases and 9,108 controls.

Table 3.1. Number of SCZ and control samples genotyped in this study and passing SNP QC

Ancestry	SCZ Sample	SCZ cases	Controls	Total
British	University College London (UCL)	511	498	1505
European-American	Genetic Association Information Network (GAIN) / Molecular Genetics of Schizophrenia (MGS)	1101	1323	3395
Irish	Trinity College Dublin	267	844	1261
European-American	Clinical Antipsychotic Trials in Intervention Effectiveness (CATIE)	339	446	1716
Scottish	University of Edinburgh / University of Aberdeen	1073	990	2340
Portuguese	Portuguese Island Collection	329	213	628
British	Wellcome Trust Case-Control Consortium (WTCCC)	461	2901	5152
Swedish	Karolinska Institute I	162	168	330
Swedish	Karolinska Institute II	380	229	609
Swedish	Karolinska Institute III	534	902	1436
Bulgarian	University of Cardiff	513	594	1107
Total		*5670*	*9108*	*14778*

3.2.3 Imputation / Association

Independent SNPs were obtained by pruning within windows of a hundred SNPs and randomly removing a SNP from any pair with $r^2 > 0.2$, I then performed multi-dimensional scaling (MDS) on this set of SNPs. MDS is a form of principal component analysis that uses similarities instead of distances. I selected the first 10 components to use as covariates in addition to site in which an individual belonged. Components were selected for being correlated with case/control status. Imputation was performed on the full genotyping sample using the 60 HapMap Phase2 CEU samples and BEAGLE (88, 89). HapMap Phase2 was selected for being the densest reference panel at the time with nearly 2.5 million high quality SNPs, more SNPs will give better resolution and increase the number of genes with association results.

Individuals were separated by group (Table 3.1) and then split into random sets of around 300 to reduce computation time during phasing and imputation. I performed logistic regression on the imputed dosages from BEAGLE using the both the site and the 10 MDS covariates.

3.2.4 Gene sets

I selected 4 of the same five gene sets that were previously used in chapter 2 (**brain expressed, synapse, learning and neuronal activity**). Additionally, I included three new sets that have since become implicated in schizophrenia. ***ARC*** neuronal activity regulated cytoskeleton associated protein complexes and ***NMDAR*** N-Methyl-D-Aspartate receptor genes (36) and implicated by a set of 34 validated ***de novo*** CNVs from a Bulgarian SCZ trio dataset (38).

3.2.5 Polygenic risk score regression method

A risk score for a given individual can be calculated by adding the number of times the individual carries the risk allele at each variant multiplied by the size of the effect (log of the odds ratio) for that variant as described in (51). In this paper, they demonstrate that common SNPs used in GWAS and aggregated in this fashion provide significantly higher risk scores for SCZ individuals then for controls. This approach, however does not narrow down the loci that are meaningfully contributing to the risk score. It detects a genomic enrichment of risk but does not

point to specific regions or genes. To address this issue I have developed a new method to test for enrichment of risk loci within a subset of SNPs defined by a pathway, geneset or SNPs of interest. My approach works by calculating two scores per individual, the cumulative risk score across all SNPs as described above and a score based only on a designated subset of SNPs. The question then becomes: does the risk score from the subset of SNPs predict phenotype over and above than the risk score using all SNPs. In other words, does the subset of SNPs contain a higher proportion of true risk loci compared to all SNPs across the genome? This test is performed using logistic regression with phenotype as the response variable, the subset risk score as an independent variable and the all SNP risk score as a covariate. The method has the benefit of allowing for the inclusion of any individual level covariates including genotyping missing-ness, ancestry components, gender, etc.

I looked to assess the type 1 error of my test on a set of null data with varying numbers of SNPs. First, I simulated 100,000 independent markers across a dataset of 5,000 cases and 5,000 controls with no true disease loci representing genome-wide data. I randomly selected 1,000 sets of each of the following numbers of SNPs, 1, 100, 1000, 10000 and 50000. I found no significantly increased type 1 error with all sets being below 5% at a p-value of 0.05 and flat p-value distributions across the board (Figure 3.1, Table 3.2). If anything I found a slightly reduced type 1 error for the largest set of 50,000 SNPs with less than 4% of sets having a p-value < 0.05 (Figure 3.1). I next performed the same assessment in the HapMap Phase 2 CEU dataset in which the SNPs are no longer independent. As before, I selected sets of SNPs of varying sizes at random. Again, there was no true association as I randomly split the

60 individuals into a case and control set. Here, I saw a very similar result to the simulated data with expected levels of type 1 error across all sets.

Figure 3.1. Type I error rates from simulated genotype data; $p < 0.01$ (blue), $p < 0.05$ (red), $p < 0.1$ (green).

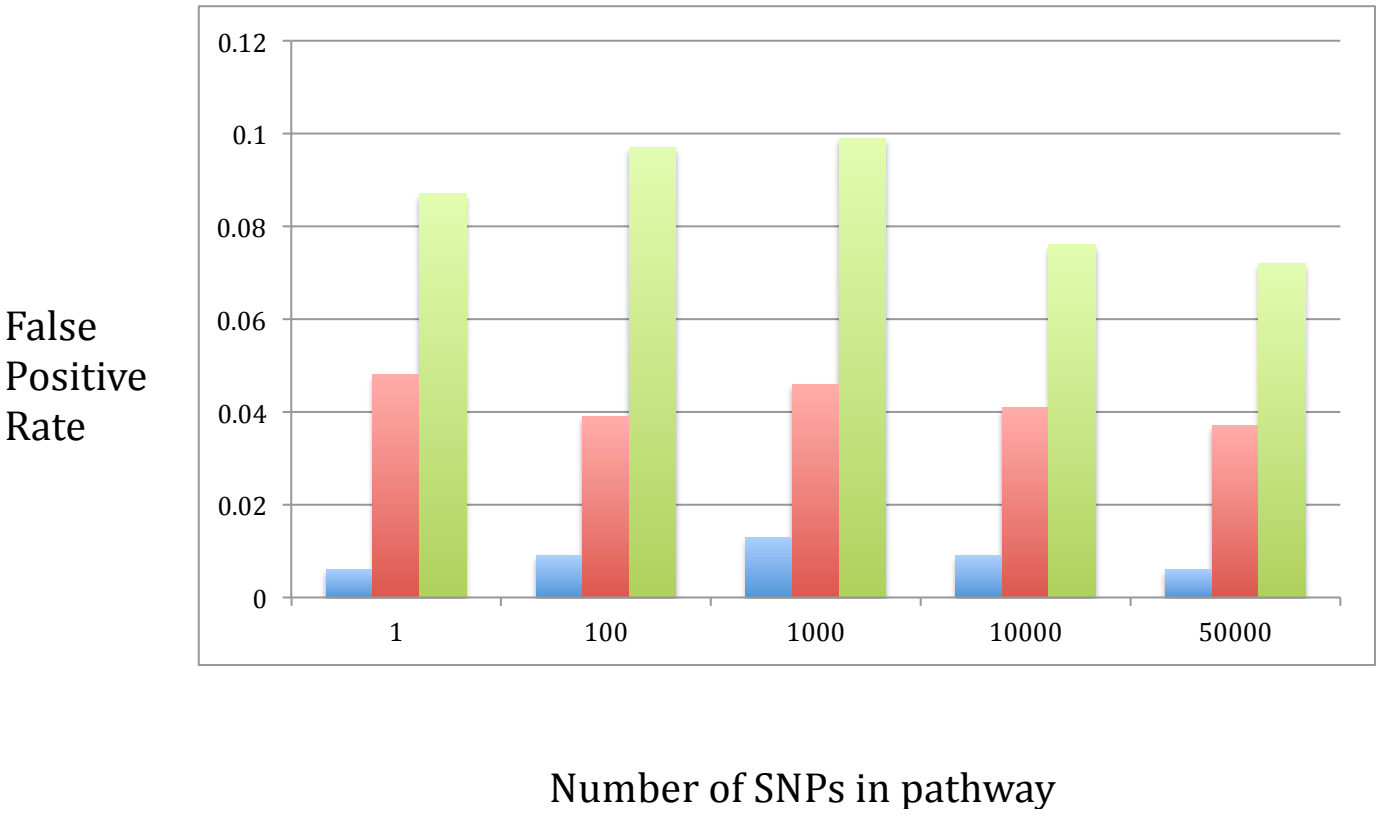
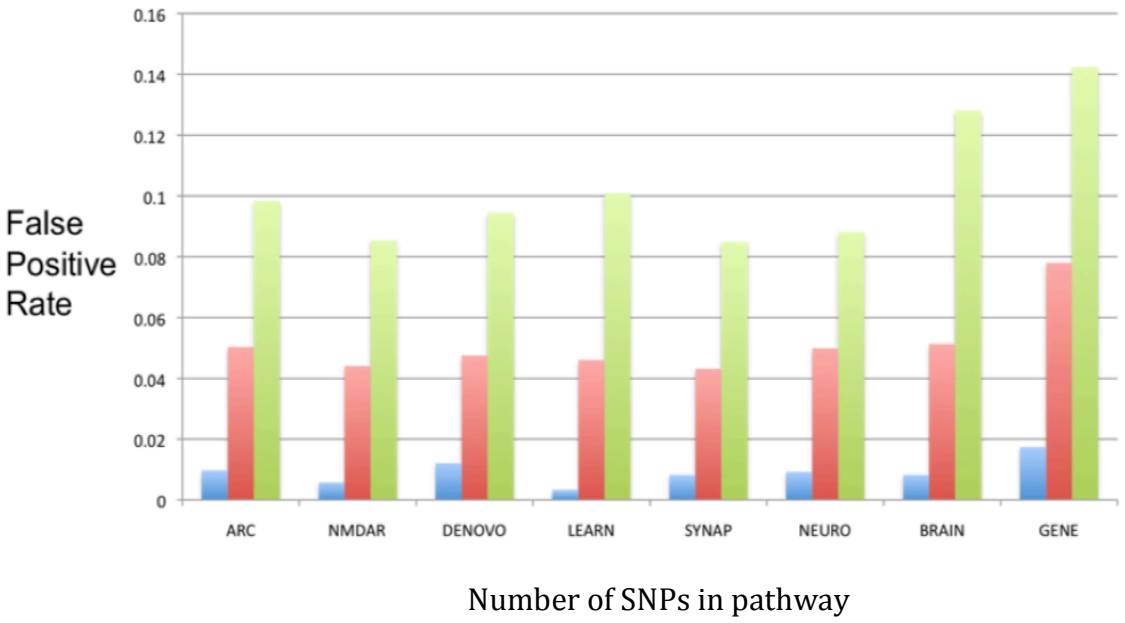


Table 3.2. Type I error in simulated genotype datasets

p-value	1	100	1000	10000	50000
0.01	0.006	0.009	0.013	0.009	0.006
0.05	0.048	0.039	0.046	0.041	0.037
0.1	0.087	0.097	0.099	0.076	0.072

I next tested the data on my true disease case/control samples. I started testing my previously defined gene sets that had sizes similar to the randomly selected sets described above but that have been previously identified in the literature as being relevant to psychiatric disease. In order to assess the behavior of my test in my observed data I permuted case/control label and tested each of the previously defined gene sets at each permutation. I looked for type 1 error rate at p-values of 0.01, 0.05 and 0.1 across 828 permutations and identified a trend of higher false positive rate with increasing number of SNPs (Figure 3.2). Given that I didn't' see any type 1 inflation in the null datasets, I concluded that I was likely improperly handling LD and needed to address the issue.

Figure 3.2. Type I error rates from permuted SCZ genotype data; $p < 0.01$ (blue), $p < 0.05$ (red), $p < 0.1$ (green).



3.2.6 Adapting method to account for LD

When initially introduced the risk profiling strategy handled linkage disequilibrium in a straight forward manner by reducing the set of SNPs to a set of quasi-independent markers by randomly removing a SNP from any pair in high LD ($r^2 > .25$) and within 200 SNPs. Since the question revolved around producing a risk score from the genome as a whole there was no concern with how to select markers other than removing as much LD as possible. Since my goal is to assess the polygenic signal within subsets of genes this approach is not ideal. If I define genic SNPs as those within the transcriptional boundaries of the gene then most SNPs will not be genic and those that are will likely be in LD with each other. Randomly removing SNPs in LD with each other would indiscriminately and overwhelmingly remove genic SNPs needed to assess the contribution of that gene to the risk score. I aimed to utilize as many SNPs as possible to be able to assess risk across all genes containing variants in the genome. The issue is that regions of higher LD and significance such as the MHC would unfairly inflate the risk scores by adding the non-independent risk contribution of many SNPs in high LD with each other together. A possible solution to this problem is to include all SNPs but to produce association results where the variant effect size incorporates LD structure. What this means is that the most significant variant will add the weight of the full effect to the score and variants in LD with it will only add additional weight corresponding to the amount of independence, for example a variant in completely LD would add no weight and a variant not in LD at all will add the full weight of it's effect size. This should eliminate the unequal contribution of regions of the genome with high LD. There are a series of

methods for handling highly correlated variables, one in particular is ridge regression. Ridge regression is a linear regression that incorporates a penalty on the parameter estimates to keep them from becoming inaccurately large (90). The penalty term or ridge parameter can be estimated by cross-validation, that is by splitting the data into equal parts, estimating the parameter on one and testing on the other to obtain the parameter that minimizes the difference between the predicted values and the true values or the prediction error. For the results of this test, I utilized this cross-validation procedure.

3.2.7 Polygenic risk score regression method using ridge regression to estimate variant effects

I applied the ridge regression method described above to account for the multi-collinearity across SNPs that occurs due to LD. That is, SNPs near each other genomically are more likely to be inherited together and a dependency will exist between their frequencies that would be counter to the assumptions of independence across variables in the general case of linear regression. The application of ridge regression allows us to keep all markers and therefore all genes that have at least one marker, avoiding any loss of information. Unfortunately, this method is computationally intensive and infeasible to compute over the whole genome or even an entire chromosome. To maximize biologically interpretable results I defined each gene as a dependent unit and performed ridge regression over all SNPs within that gene. This provided us with SNP based association results where the LD among all SNPs in the gene was accounted for in the effects estimates

returned from the regression. I then created a genic score for each individual as before adding risk alleles weighted by the new effect estimates. I also created a pathway score by simply summing up the genic scores for all genes in a given pathway. In addition, I calculated a genomic risk score for every individual summing up every RefSeq gene to be used as the covariate in my test of each geneset. After calculating the geneset score, I performed the initially proposed regression analysis using the new pathway score and including the all genes score from the ridge regression as a covariate. As described in the methods, I used a cross validation procedure, that is I split the data into ten equal sized chunks for training on all but one chunk and testing on the left out data this repeats until each chunk is tested, to take the penalty term with the smallest prediction error and use this parameter in my regression.

3.3 Results

3.3.1 Polygenic enrichment in *de novo* CNV regions without LD correction

I tested for an enrichment of common SNP association in the pathways defined above which included three pathways of particular interest that were implicated in a family study of *de novo* CNVs performed by collaborators (16). These pathways include 34 validated *de novo* CNVs and two biologically interesting post-synaptic density genesets, neuronal activity-regulated cytoskeleton-associated protein complexes (ARC) and N-Methyl-D-Aspartate receptor (NMDAR) that were

shown to be enriched for overlap with *de novo* CNVs (35). For the *de novo* CNVs, I defined the regions in which to test for enrichment as the start and end of the CNVs themselves as opposed to using gene boundaries. My hypothesis was that these regions likely contained disease relevant genes in which there could also be an enrichment of common risk variants. Of the 4 genesets I tested not related to *de novo* CNVs, I identified significance in the brain expressed genes and learning category of genes but I saw no enrichment in either the synaptic junction set or the neuronal activity set. While there was no significant enrichment of common risk signal in the gene sets enriched for *de novo* CNV mutations (ARC, NMDAR), I did see an enrichment of risk loci within the *de novo* CNVs themselves (Table 3.3).

Table 3.3. Enrichment of associated SNPs in functional genesets and regions of de novo CNVs (bold indicates significant at P<.05)

Set	#genes	SCZ (P)
DENOVO	157	**4.49E-04**
ARC	28	0.7
NMDAR	61	0.75
BRAIN-EXPRESSED	2433	**4.08E-05**
LEARNING	126	**4.70E-04**
SYNAPTIC JUNCTION	208	0.06
NEURONAL ACTIVITY	453	0.06

3.3.2 Polygenic risk score regression results after permutation

To assess the statistical properties of the proposed risk score enrichment regression method we've proposed, I performed permutation in my dataset and tested for enrichment in the same seven pathways described in the methods as well as set containing all genes. This latter set was chosen in order to investigate the

performance of the method on extremely large gene sets. For computational reasons I was limited to 828 permutations, I randomized case/control label, performed association and then tested the 8 pathways for significance in the same dataset. I found an increasing false positive rate (FPR) as the geneset size increased (Figure 3.2) with the set containing all genes, consisting of 471,000 SNPs, having an FPR of nearly 0.08 when 0.05 was expected and over 0.14 when 0.1 was expected. This inflation raised concerns regarding my initial results. Therefore, I chose to get a deeper understanding as to what was driving my most interesting result, that is, the significant enrichment of polygenic risk in the *de novo* CNV set. I first asked if the signal was driven by SNPs within genes, I also asked if given a non-significant result in the PSD genes whether the CNV enrichment was being driving by the genes that aren't overlapping with those sets. The signal was stronger in both cases, either when limited to SNPs in genes and after the PSD genes were removed (Figure 3.3). I then split the *de novo* CNV regions by chromosome to determine if any particular locus was driving the signal. It turns out that the region on chromosome 7 had the strongest effect (Figure 3.3). I next asked if this enrichment was stronger than the overall *de novo* CNV enrichment, I did this by testing the chr7 region while using the *de novo* CNV set as a covariate. I still saw significant enrichment from this analysis indicating that this region has a higher proportion of true risk loci than the *de novo* CNVs. Identifying a single significant region could potentially be pointing to relevant biology or it could be pointing to a technical issue with the method. A big concern is the effect of variable LD across different regions. A locus with a large number of SNPs in high LD with each other and significant p-values could be contributing more to the

risk score than a significant locus with lower SNP density and/or lower LD. I looked at the LD pattern for this region on chromosome 7 (Figure 3.4) and found a large region of LD around the most significant SNP (rs735956). To test whether the significant result from the chr7 locus was driven by the SNPs having high LD with this particularly associated variant, I re-ran the individual SNP association tests conditioning on rs735956 and then proceeded to re-score all individuals using the estimates of effects from the conditional association. While previously I found a highly significant result (p=0.000137), after conditioning on the most significant SNP, the result was no longer significant (p=0.07). I took this as direct evidence that LD is unevenly contributing to my results and needs to be explicitly handled.

Figure 3.3. Log scaled significance of *de novo* genesets and by chromosome at multiple p-value thresholds ($p < 0.001$, $p < 0.01$, $p < 0.1$, $p < 0.2$, $p < 0.3$, $p < 0.4$, $p < 0.5$, $p < 1$)

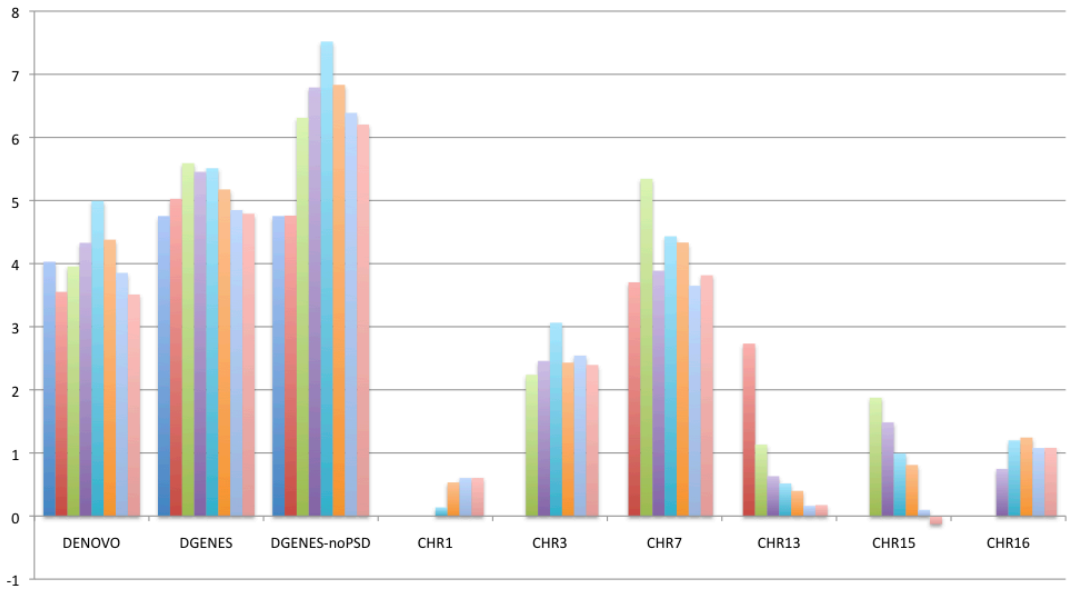
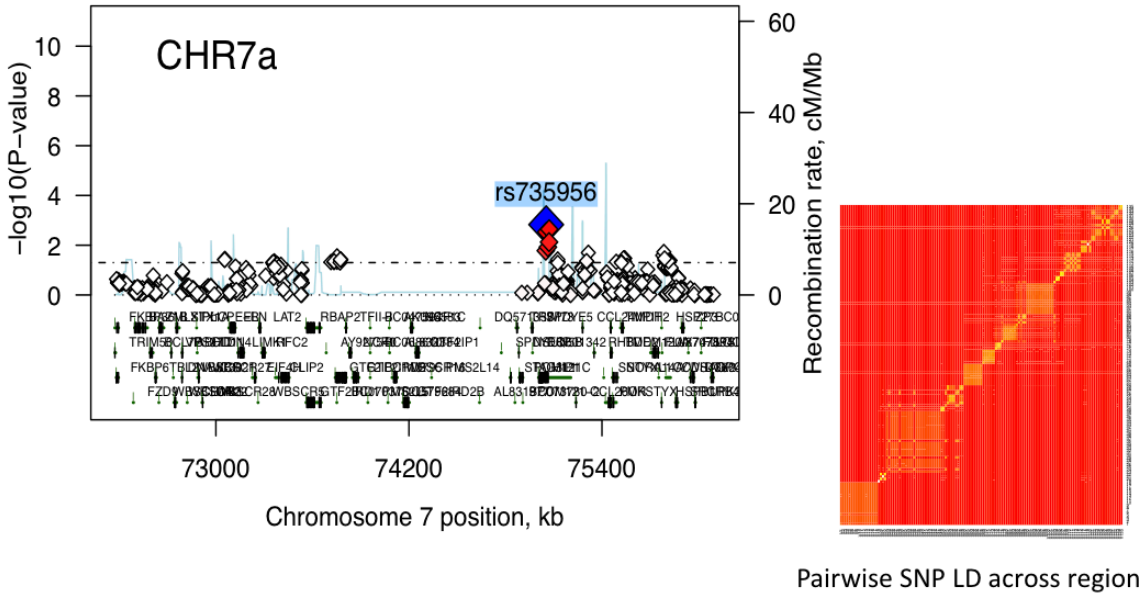


Figure 3.4. Regional association and LD plot for significant chromosome 7 locus



3.3.3 Ridge Regression

I implemented a ridge regression approach to account for correlation across markers. I then proceeded to test the same pathways tested in my initial assessment including the specific chromosomal regions that helped identify the incomplete handling of LD in the initially proposed method. Because of the computationally intensive nature of ridge regression I only included SNPs within genes (described further in methods). I found significant enrichment for all genesets except the PSD sets (ARC, NMDAR) (Table 3.4). More importantly I did not see an effect in the CHR7 region. Although I specifically implemented ridge regression to correct the LD issue there still appears to be a correlation between number of SNPs and gene set p-value resulting in larger pathways being more significant. Despite this issue I proceeded to

test 1,454 GO pathways defined at an arbitrarily selected level of the hierarchy to produce pathways of reasonable size. I found the correlation of number of SNPs and –log10(p-value) to be significantly correlated with r=0.27 (Figure 3.5). In an attempt to correct these p-values I performed 100 permutations and of the 330 nominally significant pathways ($p < 0.05$), 48 remained nominally significant after permutation (Table 3.5, full results Appendix A1). Permutation was performed by randomly selecting genes with a similar number of SNPs as those in the gene set being tested.

Table 3.4. Ridge regression results

Set	# genes	P
BRAIN	2433	**5.7E-48**
NEURO	453	**1.5E-15**
SYNAPSE	208	**1.9E-09**
DGENES	157	**3.1E-05**
DGENES-noPSD	149	**7.7E-04**
PSYCH	141	**5.2E-03**
LEARNING	126	**8.5E-03**
CHR16	22	**0.03**
CHR15	16	**0.03**
CHR1	16	0.12
NMDAR	61	0.32
CHR7	46	0.43
ARC	28	0.51
CHR13	7	0.64
CHR3	29	0.95

Figure 3.5. Scatter plot of number of SNPs and log scaled significance of all GO sets after ridge regression based polygenic pathway test (r=0.3).

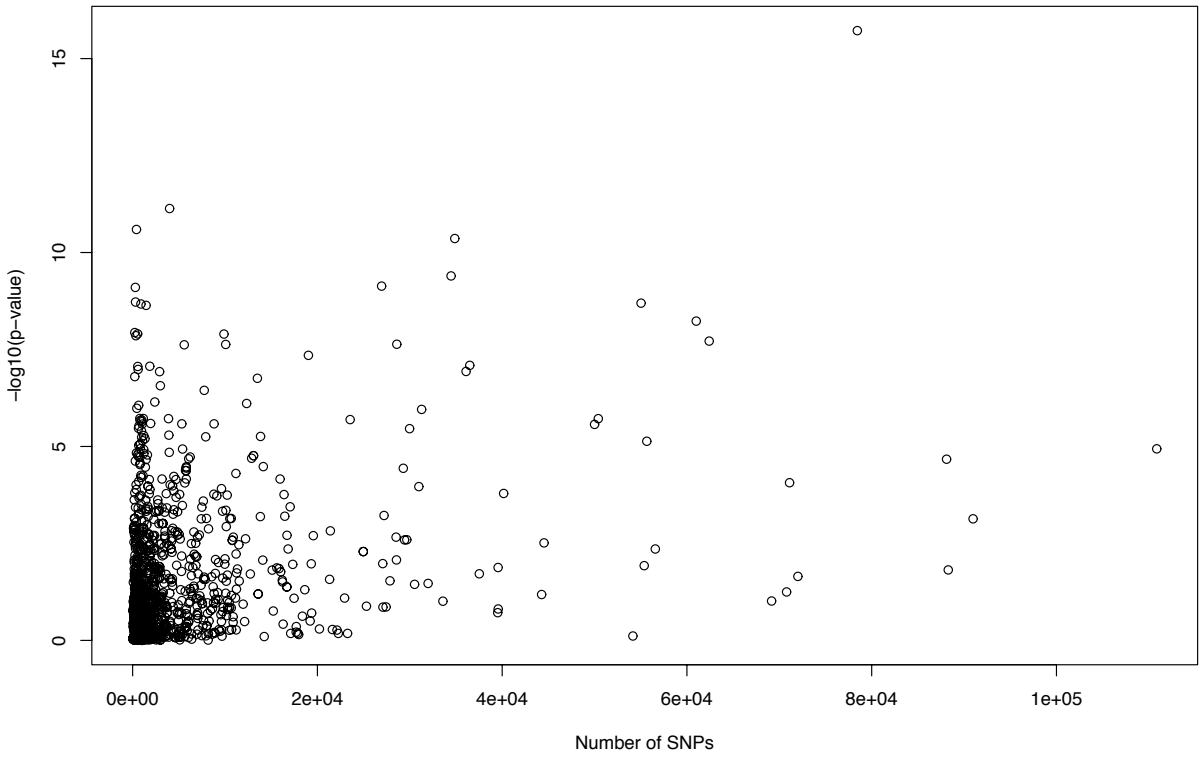


Table 3.5. Nominally significant permuted GO ridge regression results with permutation and replication p-values from independent Swedish sample

GO SET	#GENES	#SNPS	Ridge Stat	Ridge P	Perm P	Sweden Stat	Sweden P
SIGNAL_TRANSDUCTION	1634	78444	8.23	**1.91E-16**	0	-0.58	0.56
NEUROLOGICAL_SYSTEM_PROCESS	379	26955	6.16	**7.35E-10**	0	0.13	0.89
SYSTEM_DEVELOPMENT	856	55028	6.00	**2.02E-09**	0	0.43	0.67
ANATOMICAL_STRUCTURE_DEVELOPMENT	1008	61013	5.82	**5.88E-09**	0	0.83	0.41
SENSORY_PERCEPTION	190	9885	5.69	**1.27E-08**	0	-1.33	0.19
INTRACELLULAR_SIGNALING_CASCADE	667	28606	5.59	**2.31E-08**	0	0.67	0.50
G_PROTEIN_COUPLED_RECEPTOR_ACTIVITY	189	10076	5.59	**2.33E-08**	0	0.58	0.56
BIOPOLYMER_MODIFICATION	649	36492	5.36	**8.10E-08**	0	-0.09	0.93
PROTEIN_MODIFICATION_PROCESS	630	36085	5.30	**1.17E-07**	0	-0.06	0.96
CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS	209	12333	4.94	**7.77E-07**	0	-0.02	0.98
CELL_CELL_SIGNALING	405	23543	4.75	**2.02E-06**	0	0.94	0.35
CELL_SURFACE_RECEPTOR_LINKED_SIGNAL_TRANSDUCTION_GO_0007166	641	34865	6.59	**4.36E-11**	0.01	0.82	0.42
SYSTEM_PROCESS	562	34478	6.25	**4.01E-10**	0.01	0.95	0.34
MULTICELLULAR_ORGANISMAL_DEVELOPMENT	1042	62415	5.62	**1.93E-08**	0.01	0.34	0.73
REGULATION_OF_BIOLOGICAL_QUALITY	418	19025	5.47	**4.46E-08**	0.01	0.82	0.41
CENTRAL_NERVOUS_SYSTEM_DEVELOPMENT	122	13511	5.22	**1.75E-07**	0.01	0.70	0.48
PHOTORECEPTOR_CELL_MAINTENANCE	10	3008	5.14	**2.72E-07**	0.01	-0.32	0.75
NERVOUS_SYSTEM_DEVELOPMENT	382	31283	4.87	**1.11E-06**	0.01	0.04	0.97
RHODOPSIN_LIKE_RECEPTOR_ACTIVITY	134	3878	4.76	**1.92E-06**	0.01	0.67	0.50
ACTIN_CYTOSKELETON	129	7916	4.54	**5.67E-06**	0.01	-1.59	0.11
PROTEIN_METABOLIC_PROCESS	1228	55652	4.48	**7.33E-06**	0.01	0.66	0.51
PHOSPHORIC_DIESTER_HYDROLASE_ACTIVITY	39	3965	4.34	**1.42E-05**	0.01	-0.80	0.43
G_PROTEIN_COUPLED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY	342	12885	4.26	**2.01E-05**	0.01	1.63	0.10
CELLULAR_MACROMOLECULE_METABOLIC_PROCESS	1128	50410	4.76	**1.93E-06**	0.02	0.46	0.65
ENZYME_LINKED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY	140	13845	4.55	**5.49E-06**	0.02	0.34	0.74
MEMBRANE	1985	110848	4.39	**1.15E-05**	0.02	-0.37	0.71
BASEMENT_MEMBRANE	37	4868	3.49	**4.81E-04**	0.02	-1.26	0.21
TRANSMEMBRANE_RECEPTOR_PROTEIN_TYROSINE_KINASE_SIGNALING_PATHWAY	83	8809	4.70	**2.60E-06**	0.03	0.86	0.39
PLASMA_MEMBRANE	1422	88107	4.25	**2.14E-05**	0.03	0.52	0.61
VOLTAGE_GATED_CATION_CHANNEL_ACTIVITY	66	5716	4.10	**4.16E-05**	0.03	1.06	0.29
CELL_DEVELOPMENT	577	30987	3.87	**1.09E-04**	0.03	-0.18	0.86
RECEPTOR_ACTIVITY	579	40165	3.77	**1.63E-04**	0.03	-0.85	0.40
CYTOSKELETAL_PART	235	9737	3.50	**4.69E-04**	0.03	-2.07	0.04
ACTIN_FILAMENT_BUNDLE_FORMATION	13	653	4.92	**8.70E-07**	0.04	-1.32	0.19
CELLULAR_PROTEIN_METABOLIC_PROCESS	1114	50004	4.69	**2.69E-06**	0.04	0.44	0.66
ORGAN_DEVELOPMENT	568	29985	4.64	**3.46E-06**	0.04	1.56	0.12
SMALL_GTPASE_MEDIATED_SIGNAL_TRANSDUCTION	90	6224	4.28	**1.86E-05**	0.04	1.36	0.17
SYNAPSE	27	6069	4.26	**2.08E-05**	0.04	0.62	0.53
PHOSPHORIC_ESTER_HYDROLASE_ACTIVITY	151	14140	4.15	**3.32E-05**	0.04	-0.76	0.45
POST_TRANSLATIONAL_PROTEIN_MODIFICATION	475	29297	4.13	**3.64E-05**	0.04	-0.20	0.84
BEHAVIOR	155	5810	4.09	**4.23E-05**	0.04	1.73	0.08
MEMBRANE_ORGANIZATION_AND_BIOGENESIS	135	11178	4.06	**4.97E-05**	0.04	1.69	0.09
ACTIN_FILAMENT_ORGANIZATION	24	909	4.00	**6.27E-05**	0.04	-0.43	0.66
BRAIN_DEVELOPMENT	51	4685	3.97	**7.14E-05**	0.04	2.09	0.04
REGULATION_OF_TRANSLATIONAL_INITIATION	31	658	3.95	**7.95E-05**	0.04	-1.19	0.23
PLASMA_MEMBRANE_PART	1155	71117	3.93	**8.62E-05**	0.04	0.96	0.34
TRANSFERASE_ACTIVITY__TRANSFERRING_GLYCOSYL_GROUPS	111	5084	3.76	**1.72E-04**	0.04	0.56	0.58
PEPTIDE_RECEPTOR_ACTIVITY	53	1667	3.53	**4.15E-04**	0.04	0.55	0.58

3.3.4 Results using independent dataset

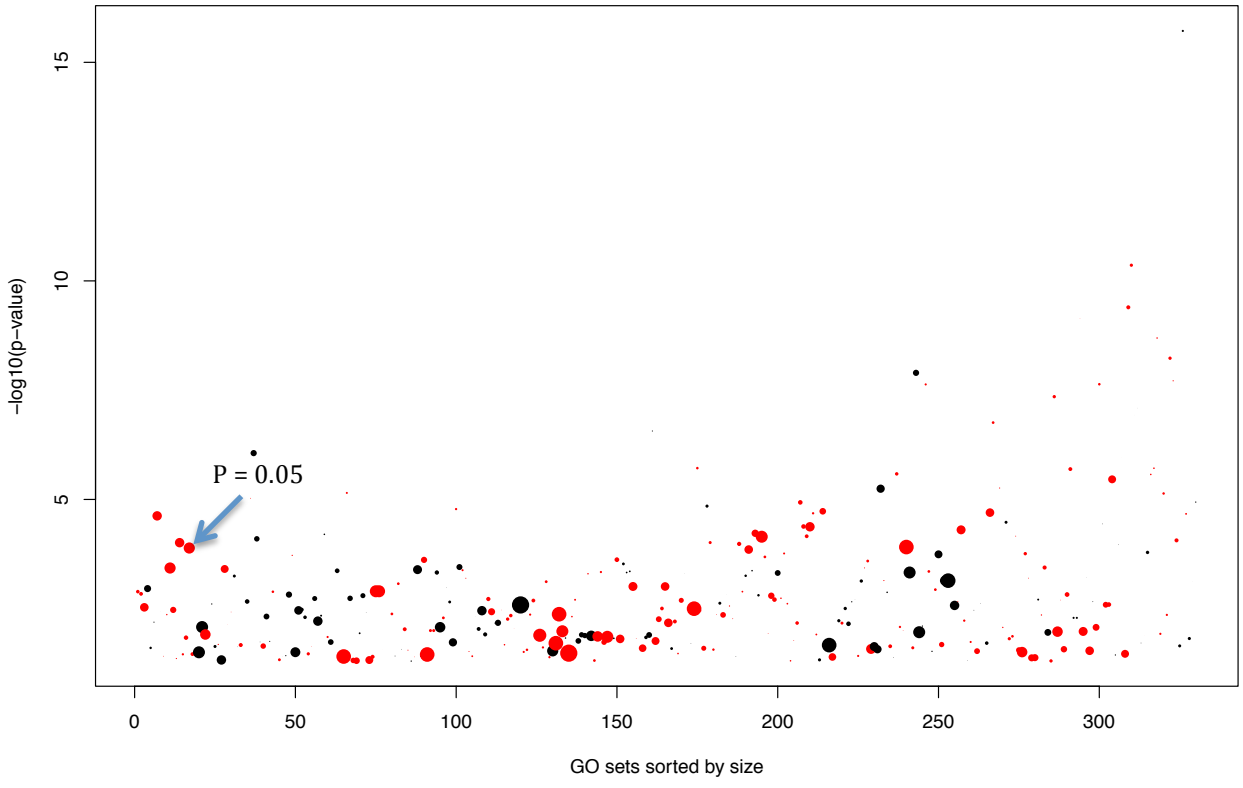
Despite remaining issues with correlation of significance and pathway size I posited that I should be able to identify true significant pathways by looking for consistency in an independent dataset. For this I used a dataset of 1,154 schizophrenia patients and 1,339 controls that were from the same Swedish collection. As with the discovery dataset, I imputed into a HapMap2 reference

dataset. I took effect size estimates from the full sample used previously and created a score for every Swedish individual for each gene and tested the 7 functional gene sets and all 1,454 GO gene sets for differences in risk scores between cases and controls. I identified 330 sets that reached my nominal significance threshold of 0.05, none of those were in the initially selected functional sets (Table 3.6). Of the ones that were nominally significant in the discovery dataset, there were no clear sets that showed significance above what's expected given the size of the pathway itself (Figure 3.6). Figure 3.6 shows each GO set plotted along the x-axis sorted by number of SNPs, the y-axis is the significance of that pathway and the size of the point represents the significance in the replication data with larger points being more significant. Red dots indicate pathways where SCZ individuals have higher scores than controls and black points indicate where control individuals have higher risk scores. Two things to note about this figure are the clear correlation between number of SNPs in the pathway and significant result in the original dataset and that the pathways with the most significant result in the initial dataset are devoid of significance in the independent dataset. In fact, only a single GO pathway (Brain Development) is both significant after permutation and replicates at $p < 0.05$ (Table 3.5).

Table 3.6. Results of functional genesets in replication data

Set	# genes	P
DENOVO	157	0.13
ARC	28	0.08
NMDAR	61	0.08
BRAIN-EXPRESSED	2433	0.60
LEARNING	126	0.07
SYNAPTIC JUNCTION	208	0.39
NEURONAL ACTIVITY	453	0.77

Figure 3.6. Replication results for 330 nominally significantly ($p < 0.05$) enriched GO genesets. Y-axis represents significance in original sample, size of dot indicates significance in replication data and red indicates positive direction of enrichment.



3.4 Discussion

In this chapter, I developed a new method to test for enrichment of associated markers across sets of genes using common GWAS SNPs. I tested this method using both simulated genotype datasets and real genotype data absent of any phenotypic effects. I found expected type I error rates in both analyses. I then proceeded to test seven disease relevant sets of genes and identify significance in multiple sets including a set of *de novo* CNVs from SCZ probands. Unfortunately, I also noted an inflation of my test statistic with increasing number of SNPs in my genesets. I determined that dependence across markers is the likely cause for this inflation and implemented an adapted method to account for this inflation. I presented a modified form of my polygenic regression method that uses ridge regression up front to estimate effects of each variant accounting for LD in that gene. This new approach did not correct the size bias as geneset significance is still correlated with number of SNPs. I attempted replication in an independent dataset and identified only a single geneset that was nominally significant. Of note, I saw no similar geneset size bias in my replication results.

My initial results on simulated and null datasets yielded appropriate error statistics but this did not remain true upon using disease GWAS data. A likely reason for this is the absence of LD all together in simulated data and the absence of true association in the null data. True association peaks will have extended LD with the most significantly associated variant and if not accounted for will unfairly sum all the effects of those variants despite not being independent. This is clear when looking

more carefully at the specific loci driving the *de novo* CNV result. I identified a significant enrichment in the chr7 locus that is driven by LD to a single variant, as performing conditional association on this variant and then testing polygenic regression results in a region that is no longer significant. I selected one of several options to attempt and correct this correlation while not having to reduce the number of variants, hoping to keep as much information as possible. Ridge regression allowed us to perform association on all SNPs in a gene accounting for LD. The result of this method was to correct the artificial inflation of the chr7 locus, however I still saw increased significance with increasing number of SNPs across many GO sets pointing to an obvious limitation of this method.

Despite this limitation, my results suggest a possible overlap between the genes underlying common variants of small effect and the genes underlying the very rare class of *de novo* CNVs. The hypothesis here is that *de novo* CNVs are more likely to be deleterious and that a proportion of the genes underlying these CNVs will be relevant to disease. Common variants or other inherited forms of variation may aid in refining which genes are indeed relevant to disease. My result suggests that genetic variation of all types could be pointing to similar sets of genes and power to detect these genes can be improved by combining all variation into a single test. Additionally, I found enrichment of associated SNPs within several of the other sets including genes expressed in the brain and genes related to learning. I also tested all sets created by the Gene Ontology Project that are assembled based on biological function. I tested these pathways in my sample and in a replication sample, despite many significant results in each dataset I found only one pathway that replicated at a

nominally significant p-value ($p<0.05$). Two possible reasons for this are; 1) the bias in my method creates false significance where none exists and therefore replication fails (e.g. in large pathways), 2) there are power issues in both detecting these signals and replicating them (in smaller samples). Replication has been difficult across all pathway analyses to date and will be required before the biological interpretations can be taken forward into further research. The pathway that did replicate is of particular interest however and represents genes involved in the progression of the brain overtime, from its formation to the mature structure. This pathway is particularly small, containing only 51 genes and follow up will be required to assess the true contribution of these genes to SCZ risk.

There is one additional assumption made in this method that could also be contributing to the pathway size bias. In the standard polygenic testing approach, two datasets are required, a discovery dataset in which GWAS is performed and a test dataset in which scores are created and then compared between cases and controls. Here, I believed this two-dataset approach was unnecessary given that I was using the same effect estimates for my gene set score and my all gene score. The observation that significance and gene set size are not correlated in my replication sample raises the question of whether this was a valid assumption to begin with.

Finally, I aimed to develop a fast, flexible framework to test particular sets of genes for greater enrichment of risk loci than the genome overall. I believed that there are many variants contributing to disease but that these variants are not randomly scattered across the genome, instead they will be distributed in greater volume to genes of particular functions that are involved in biological processes

relevant to the underlying mechanisms of disease. Under this model, if I can identify these sets of genes I should be able to significantly distinguish them from the baseline rate of risk. In testing this method, I identified an issue that affects most aggregation methods, in that variants across the genome are not independent. Summing over larger numbers of variants will invariably increase the number that are true risk loci because I expect there to be many and those that are true risk loci will be correlated with many other markers further increasing the estimated effect. My sophisticated approach to account for this approach did not completely remove this effect, furthermore it turned a very simple, fast, scalable approach into a more complicated, computationally intensive approach. This exercise has identified some issues in using polygenic risk scores as a means to perform pathway analysis but has implicated some possibly interesting biology and sets the stage for an alternative solution to the LD issue that may produce a less biased test.

Chapter 4: Analysis of recessive and multi locus variants in schizophrenia exome sequencing.

4.1 Introduction

The previous chapters and much literature have supported schizophrenia as a highly polygenic disorder with many variants of small effect contributing to an individual's risk for the disease (51). This highly polygenic, complex genetic architecture requires studying all forms of variation and all types of disease models to identify variants related to risk. Focusing on mutations likely to have deleterious effects on protein function will both reduce the number of variants tested and increase the size of effect thus improving power to identify disease variants. One particular class of variation with a strong prior for relevance to disease is putative loss of function (LoF) mutations, those with high probability of disrupting the protein, these include a premature creation of a stop codon (nonsense), the insertion or deletion of sequence that alters the size of the protein from being divisible by three (frameshift) and a mutation effecting the most conserved part of the splicing junction, two base pairs into the intron from a splice junction (essential splice sites). While these mutations are likely to have a functional impact, often the redundancy of a diploid genome mitigates the loss of one copy. However the loss of both copies would exclude the making of that protein all-together and theoretically have a major effect on any process that required that gene product. LoF mutations that occur on both chromosomes, either at the same site (homozygous) or at different sites in the same gene (compound heterozygous), will be a reduced set of high impact mutations

that are likely to be related to disease. Here, I analyzed the role of this "two-hit" model in exome sequencing of schizophrenia patients and controls.

4.1.1 Two-hit hypothesis as a model for schizophrenia

Autosomal recessive mutations are known causes of numerous diseases such as Tay Sachs, Sickle Cell and Cystic Fibrosis. Recessive mutations occur more often in autozygous stretches of the genome where an individual carries both chromosomes from the same ancestor. These regions will therefore have increased likelihood of carrying two copies of a deleterious mutation. We can identify autozygous regions by searching the genome for large continuous stretches of homozygosity, called runs of homozygosity (ROH). A recent paper (91) looked at the contribution of autozygosity to schizophrenia risk in a large, diverse sample of 9,388 cases and 12,456 controls. They defined a metric (Froh) as the proportion of an individual's genome that is made up of ROH. As Froh increased so did the likelihood of having schizophrenia, such that with every 1% increase of ROH there is a 17% increase in disease risk. They did not however identify any particular ROH significantly associated with SCZ. By sequencing the genes in these homozygous regions we may identify rare, risk increasing homozygous events that are driving this correlation between autozygosity and SCZ.

4.1.2 Exome sequencing in psychiatric disease

Next generation sequencing has allowed for the survey of all variation across the coding part of the genome in a rapid and cost effective manner. With this information it becomes feasible to assess the contributions of nearly all mutations with functional impact on translation. While not complete (representing only 1% of the genome) protein coding regions contain all the genes and are currently the most interpretable parts of the genome for biological understanding and further follow up. However, there have yet to be many significant findings of rare, large effect variants published from the early sequencing studies of numerous common disorders. This includes a study of 166 SCZ cases that followed up the most schizophrenia specific variants in 2,617 cases and 1,800 controls with no significant hits (92). Although small, these initial sample sizes were powered to detect variants of high effect at frequencies not tested previously. Other recent sequencing papers have demonstrated that looking at small sets of likely deleterious mutations can point to genes and pathways of biological relevance for disease. One such approach is the exome sequencing of trios to detect *de novo* mutations (63-66, 93, 94). These studies have had success in finding disease related mutations in diseases such as autism and intellectual disability and provide hope that similar approaches can be used to further understand the biology of other psychiatric diseases like SCZ. Exome sequencing data will also allow us to identify recessive coding mutations including those underlying ROHs. An analysis of both recessive and compound heterozygous LoF variants in autism identified a two-fold enrichment of rare two hit mutations in

cases compared to controls as well a significant increase in LoF mutations on the X chromosome in male cases (95).

4.1.3 Loss of Function mutations on the X chromosome

Despite containing nearly 10% of all genes, the X chromosome receives little attention when assessing genetic variation contributing to disease risk. While evidence exists demonstrating a 1.4 fold increase in male risk to schizophrenia (96) there has been less work focused on identifying any particular variants contributing to this difference on the X chromosome. This persists despite a large body of literature on both X-linked intellectual disability and cognitive impairment in individuals with schizophrenia (97). Here, focusing only on X chromosome in males I ask similar questions to the autosomes, that is: are LoF variants contributing to disease risk and/or are any genes on the X chromosome enriched for such mutations in cases.

In this chapter, I first identify ROH in a large GWAS sample and ask if there exists any enrichment in cases or if any particular genomic region is associated to SCZ. I then identify two hit LoF variants (homozygous and compound heterozygous) in exome sequencing of 2,536 SCZ cases and 2,543 controls. I assess the contribution of these variants to disease risk and test whether any gene is significantly burdened with case events. Finally, I ask if these mutations are contained within previously called ROH pointing to a proposed mechanism for disease risk.

4.2 Methods

4.2.1 Sample

The subjects included in this study were all born in Sweden and the cases were identified as schizophrenia (SCZ) through the Swedish Hospital Discharge Register which includes all individuals that have been hospitalized in Sweden since 1973. These samples include the Swedish samples presented in chapters 1 and 2. Cases were required to have had at least two hospitalizations with a SCZ diagnosis. This form of diagnosis has been shown to agree strongly with diagnoses that were established by the attending physician and confirmed in a subset of subjects by medical record review (98, 99). Control subjects, also selected through registers, were group-matched by age, sex and county of residence, and must not have been hospitalized with a psychiatric diagnosis. All subjects were at least 18 years old and gave written informed consent to participate. The study was approved by the Ethical Committee at Karolinska Institutet.

4.2.2 Genotyping, QC and imputation

The full Swedish sample consists of 5,000 SCZ cases and 6,242 controls that were genotyped on a combination of Affymetrix and Illumina chips and have been published previously (*Ripke et. al. 2013, in press*). In order to have a consistent set of SNPs across all individuals I first converted HapMap Phase 3 imputed dosages into genotypes. I retained only SNPs where the posterior probability of the most likely

genotype was greater than 99% for every individual. I further removed SNPs failing HWE ($p<1e-5$), having genotyping rate < 99% and having MAF < 2% and pruned to remove pairs of SNPs in LD ($r^2 > 0.05$) with each other. My resulting dataset contained 54,867 SNPs which I called ROHs from. For accurate calling of ROH independent SNPs are required, otherwise regions of strong LD will disproportionately affect ROH calls.

4.2.2 Identifying and testing runs of homozygosity (ROH)

I called ROHs using the Plink software package (82), a window based approach was taken where for every 100 SNPs a determination was made regarding whether that region was "homozygous" based on having no more than 1 heterozygous genotype. An ROH would be called if it spanned at least 500kb of sequence, had SNP density no less than 1 SNP every 100kb and every SNP in that region was considered homozygous in at least 5 percent of all windows it was in.

4.2.3 Illumina exome sequencing and QC

A subset of the Swedish sample was selected for exome sequencing. Exons were enriched for sequencing using Agilent SureSelect hybrid capture arrays platform covering ~32Mb and sequenced on Illumina GAII and HiSeqs with 76 base pair reads. Alignment and variant calling were performed using the Picard/BWA/GATK pipeline (100-102). Each sequenced DNA fragments is called a

read and the number of DNA fragments aligned to a particular base represents the read depth. The greater the read depth the more opportunities exist to see reads with both reference and non-reference alleles which will increase power to detect variants. Mean depth was ~90 reads per base across the targeted exons with 93% of targeted bases having 10 reads or more and 81% having 30 reads or more. 2,536 SCZ cases and 2,543 controls remained after removing individuals for having too small of a proportion of variants in dbSNP (< 95%) (exome sequencing QC was performed by Shaun Purcell and myself). In calling double-hit variants, I restricted to variants less than 5% minor allele frequency in the population, allele balance (the proportion of reads containing non-reference alleles) for heterozygotes between 30% and 70% and for homozygotes greater than 90%, and genotype quality (scaled confidence that called genotype is true) than 30.

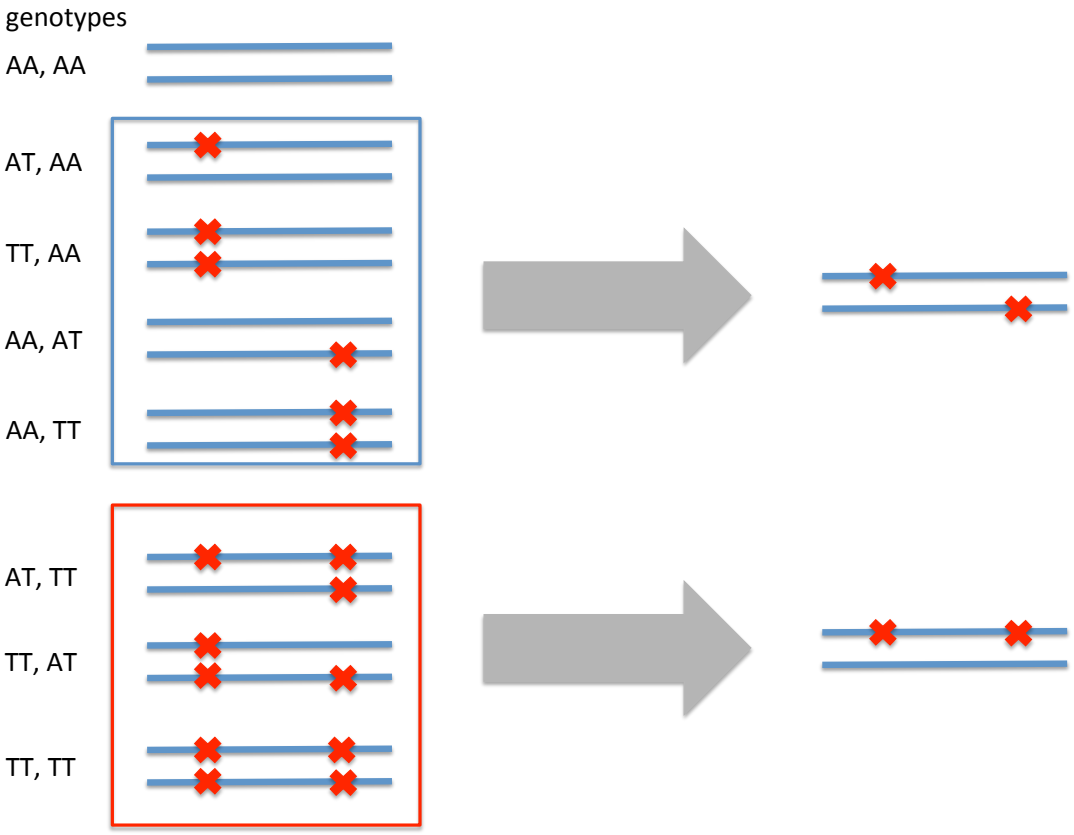
4.2.4 Plink/Seq two-hit test

In order to evaluate the role of two hit variants in SCZ I designed and implemented a method to both identify and test these variants within the framework of the statistical package Plink/Seq. The method fits into the gene based testing framework and analyzes each transcript independently, a crucial detail when thinking about multiple LoF events or "double knockouts," as it ensures both functional mutations are affecting the same protein. Only mutations occurring in the same read can confidently be asserted to be from the same chromosome (in phase).

Since most pairs of mutations will not be within the read length of 76 base pairs, I employed a previously used statistical phasing approach to identify compound heterozygous events (95). I defined compound heterozygosity as an individual carrying a heterozygous mutation on each chromosome at different locations within the same gene.

When trying to identify compound heterozygous mutations one can use the haplotypes seen in a large population to infer the likelihood that these variants were inherited together. I looked at all possible combinations of two bi-allelic variants seen in the population and categorized each individual into one of two groups (Figure 4.1). The first group is situations where there is no evidence that the variants are inherited together, this is only when an individual carries one or two copies of non-reference allele in one variant but not the other. The second group is when under a diploid scenario both variants must have been inherited together. This occurs when an individual carries non-reference alleles at both sites and is homozygous for at least one. In order for us to call a compound heterozygous event I must see at least one individual carrying one or two non-reference alleles of variant 1 with only reference alleles at variant 2 and vice versa (group 1), additionally I must never see an individual with the alleles inherited together (group 2). Upon identifying all two hit events, variants are written to a file with positional and annotation information and then tested for association.

Figure 4.1. Visual representation of statistical phasing algorithm, blue lines represent chromosomes, red x's represent mutations at two different variants. Two variant haplotypes in the blue box assume mutations do not occur on the same chromosome (in phase), haplotypes in the red box represent evidence for variants being in phase and not compound heterozygous. In order to be called a compound heterozygous event, you must see both variants independently of each other (blue box) and never in phase (red box).



4.2.5 Permutation procedures

I wanted to implement a testing procedure that accounted not only for the difference in counts between cases and controls but also the likelihood of seeing the observed number of events given the allele frequency. For homozygous variants I employed a permutation process that unlinked both phenotype and alleles. I created

a set of N x 2 chromosomes, where N is the number of individuals, with z chromosomes having the alternative allele (number of heterozygous individuals + twice the number of homozygous individuals). I then randomly selected a pair of chromosomes for every individual and count the number of two hit events in cases and controls. I computed an odds ratio from these numbers and compared to the observed odds ratio to determine the p-value. Up to 100,000 permutations were performed for each variant with the exact number being determined by significance level after each order of magnitude. For example, a variant where more than 20 permutations have greater than or equal odds ratios to the observed data will only undergo 100 permutations. For compound heterozygous mutations I performed a similar strategy except for each individual I randomly selected one chromosome from each of the two variants. To test gene sets with multiple variants I assumed independence across sites and selected multiple chromosomes from either a single variant (homozygous) or two variants compound heterozygous) for each two-hit mutation in the set.

4.2.6 Previously implicated gene sets

In a similar fashion to those presented in chapters 2 and 3 I have selected a the most current and relevant genes that have been implicated in schizophrenia. This includes two sets: s**ynaptic genes** are a set of synaptic protein complexes previously implicated in SCZ *de novo* CNVs (38) and ***de novo* genes** which are genes carrying at least one *de novo* mutation in recent AUT, ID and SCZ manuscripts (63-66, 93, 94).

4.3 Results

4.3.1 ROH mapping

I first identified ROH from the larger GWAS sample (5,000 cases, 6,242 controls). I called 4,116 ROH across all individuals with higher rates of ROH seen in cases compared to controls (case rate=0.40, control rate=0.34, p=0.004, Table 4.1). I next asked if this result was driven by ancestry, as reported (103), I tested again using logistic regression but included the first MDS component as a covariate and the difference was no longer significant (p=0.825). Given this result, I decided to focus on a more homogenous subset of individuals shown in red in Figure 4.2 Within this sample, I also didn't see any increase in case rate of ROH compared to controls. Given the small effect size seen the *Keller et al.* study (OR=1.17), my significantly reduced sample size and differences in sample populations it is not surprising that I didn't replicate this signal. I then asked if any particular region carried a significant increase in case ROH compared to controls. I found a few regions of nominal significance (P<0.05) but none remained significant after multiple test correction (Table 4.2). I specifically tested each gene and those that reached nominal significance were genomically distributed into five regions. Of these, the most significant region occurred on chr2:206-208Mb and at this significance level spanned 17 genes and 1.5 Mb with the most significant section having p-value 0.002 (7 cases, 0 controls). Of note, I also saw a number of ROH in cases in the 10p14 region that has been previously implicated in individuals presenting with DiGeorge and VCFS

syndromes, which includes intellectual disability, and not carrying the typical chr22q11 deletion known to cause this disorder (104). Despite not seeing an overall enrichment of ROH in cases I hoped to utilize this finding to possibly identify a particular homozygous variant underlying these regions of ROH that could be significantly contributing to disease. I first needed to identify these variants from exome sequencing data and summarize their overall contribution to SCZ.

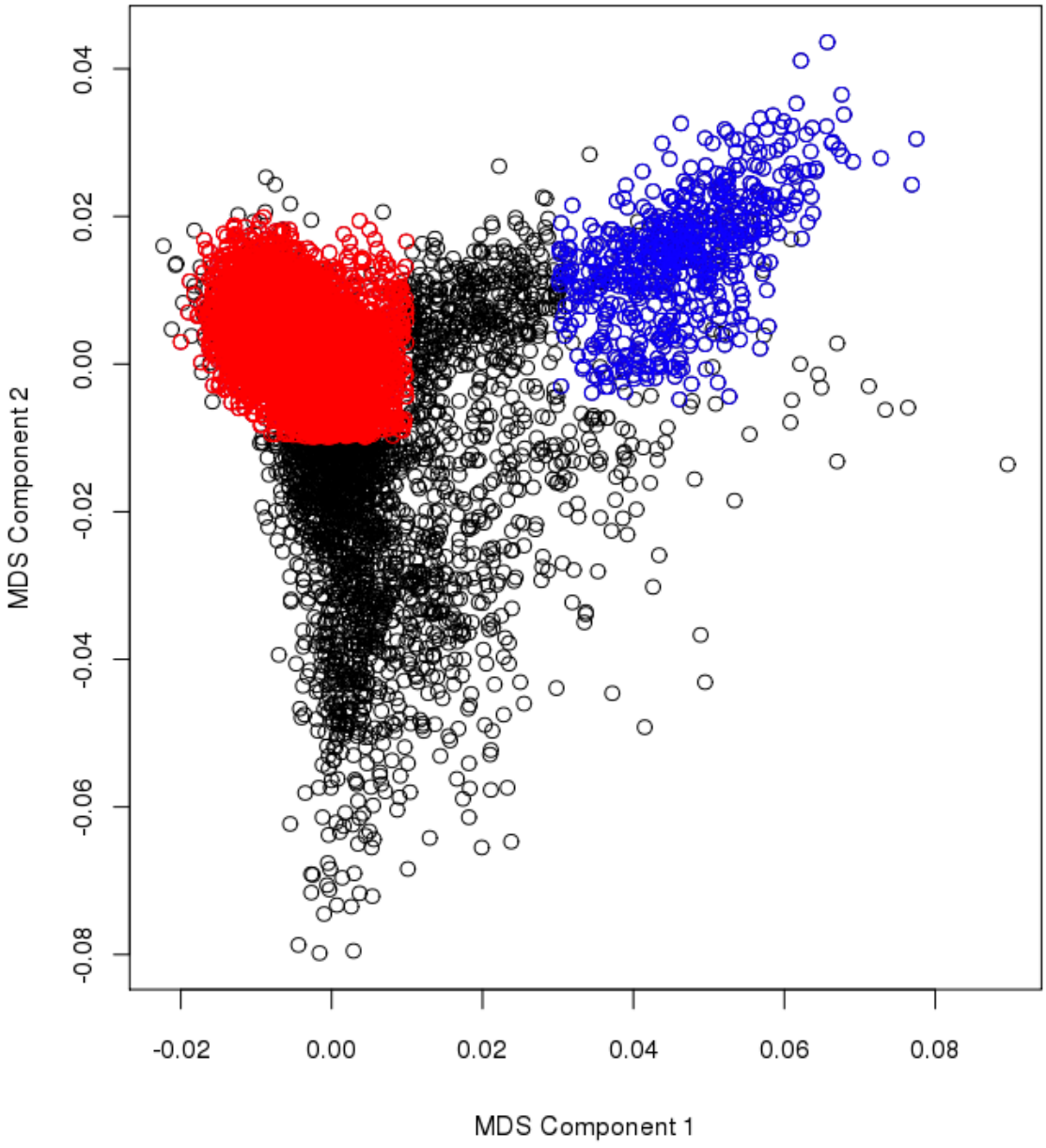
Table 4.1. Summary of ROH

Individuals	N cases	N controls	N ROH	Case Rate	Control Rate	P logistic	P (MDS1 covar)
Full sample	5000	6242	4116	0.40	0.34	0.004	0.825
Homogenous Swedish sample	3386	4933	1708	0.21	0.20	0.637	0.695
Finnish subset	414	205	547	0.87	0.90	0.817	0.076

Table 4.2. All genes with p < 0.01 for ROH enrichment analysis

	CHR	GENE	START	END	CASE	CONTROL	P	Pcorr
	2	LIMS1	109150811	109303702	9	2	0.007	0.824
	2	RANBP2	109335937	109402267	9	2	0.007	0.824
	2	CCDC138	109403219	109492847	9	2	0.007	0.824
	2	EDAR	109510927	109605828	9	2	0.007	0.824
	2	SH3RF3-AS1	109743784	109746575	9	2	0.007	0.824
	2	SH3RF3	109745997	110262207	9	2	0.007	0.824
	2	MIR4265	109757946	109758044	9	2	0.007	0.824
	2	MIR4266	109930027	109930081	9	2	0.007	0.824
	2	SEPT10	110300374	110371783	9	2	0.007	0.824
	2	SOWAHC	110371911	110376564	9	2	0.007	0.824
	2	MIR4267	110827538	110827619	9	2	0.007	0.824
Region 1	2	MALL	110841447	110874143	9	2	0.007	0.824
	2	NPHP1	110880914	110962639	9	2	0.007	0.824
	2	LINC00116	110969106	110980517	9	2	0.007	0.824
	2	LOC100507334	111003215	111024135	9	2	0.007	0.824
	2	LOC151009	111132686	111142113	9	2	0.007	0.824
	2	BUB1	111395409	111435684	9	2	0.007	0.824
	2	ACOXL	111490150	111875799	9	2	0.007	0.824
	2	BCL2L11	111878491	111926022	9	2	0.007	0.824
	2	DPP10	115199899	116602326	6	0	0.005	0.732
	2	LOC389023	115901625	115918920	6	0	0.005	0.732
	2	DDX18	118572255	118589953	6	0	0.005	0.732
	2	CCDC93	118673054	118771739	6	0	0.005	0.732
	2	NRP2	206547224	206662857	6	0	0.004	0.732
	2	INO80D	206858445	206950906	7	0	0.002	0.421
	2	GCSHP3	206980297	206981296	7	0	0.002	0.421
	2	NDUFS1	206987803	207024243	7	0	0.002	0.421
	2	EEF1B2	207024318	207027653	7	0	0.002	0.421
	2	SNORD51	207026605	207026674	7	0	0.002	0.421
	2	SNORA41	207026952	207027083	7	0	0.002	0.421
	2	GPR1	207040042	207082771	7	0	0.002	0.421
Region 2	2	ZDBF2	207139523	207179148	7	0	0.002	0.421
	2	ADAM23	207308368	207482679	7	0	0.002	0.421
	2	LOC200726	207507142	207514173	7	0	0.002	0.421
	2	DYTN	207516345	207583120	7	1	0.009	0.970
	2	MDH1B	207602489	207630050	7	1	0.009	0.970
	2	FASTKD2	207630112	207660911	7	1	0.009	0.970
	2	MIR3130-1	207647958	207648032	7	1	0.009	0.970
	2	CPO	207804278	207834198	7	1	0.009	0.970
	2	KLF7	207938862	208031970	7	1	0.009	0.970
Region 3	5	LOC340094	5034472	5070115	7	1	0.009	0.970
	5	ADAMTS16	5140443	5320412	7	1	0.009	0.970
Region 4	6	RNU6-48	12392559	12392586	7	1	0.010	0.970
	6	PHACTR1	12717037	13288075	7	1	0.010	0.970
	6	LOC100130357	13279527	13295818	7	1	0.010	0.970
	6	TBC1D7	13305184	13328787	7	1	0.010	0.970
Region 5	10	CAMK1D	12391583	12871733	7	1	0.009	0.970
	10	MIR4480	12620752	12620822	6	0	0.004	0.732

Figure 4.2. Population components for all individuals and those selected subsets (Red: homogenous Swedish sample, Blue: Finnish sample).



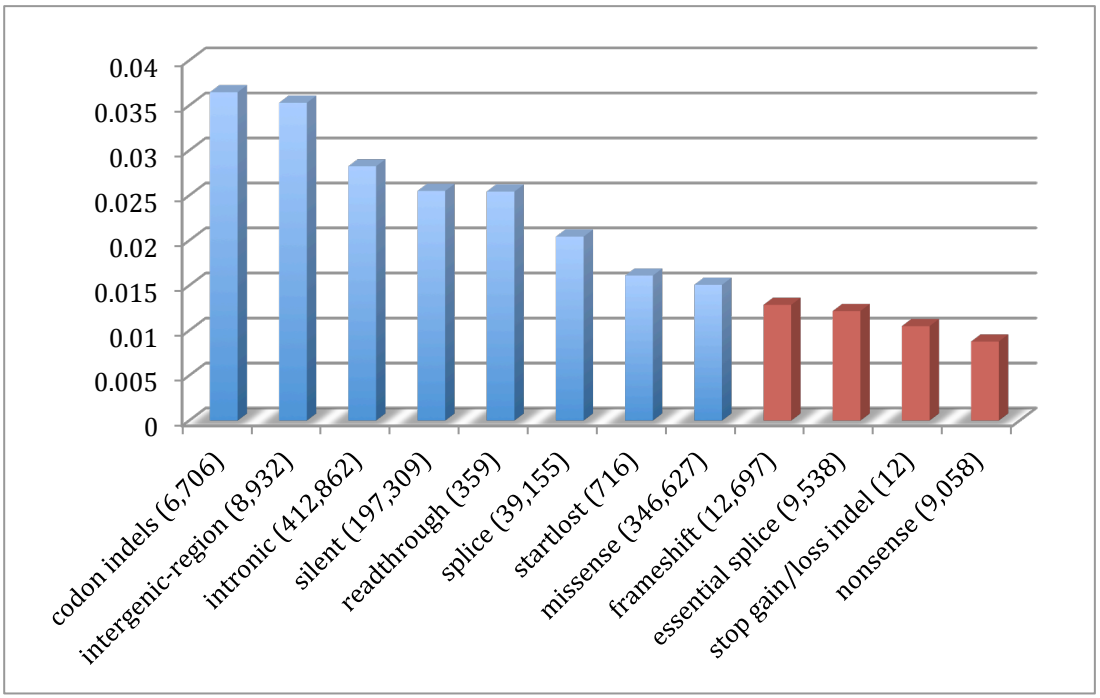
4.3.2 Rare two hit loss of function burden

I defined loss of function (LoF) as variants with high probability of disrupting the protein, these include nonsense, frameshift and essential splice sites (2 base pairs into the intron from a splice junction that alters the conserved sequence). I posited that these will be the most deleterious and therefore most likely to be contributing to disease risk. If purifying selection is actively removing deleterious alleles from the population than these variants should have lower frequencies than non-deleterious mutations. Figure 4.3 shows the average minor allele frequency stratified by functional annotation, in red are the annotations considered LoF in this analysis. Here, reduced frequency is observed in these classes compared to annotations less likely to be functional such as silent and intergenic mutations. I also excluded variants with MAF greater than 5% under the logic that they are less likely to be deleterious. This reduced my set of included variants to less than 3% of the total. I identified 496 two hit LoF mutations across my sample, a rate of 0.1 per person which is 9 times less common than coding *de novo* mutations (Table 4.3). I found no enrichment of two hit LoF mutations in cases compared to controls (242 case two hit LoF; 254 control two hit LoF, p=0.73) and no significant difference in number of variants per individual with the same 0.1 rate in both cases and controls (Table 4.3).

Table 4.3. Summary of LoF recessive variants (MAF < 5%)

Individuals	N Cases	N Controls	Cases (Var)	Controls (Var)	Case Rate	Control Rate	OR	P Fisher
Full sample	2536	2543	229 (242)	241 (254)	0.10	0.10	0.95	0.73
Homogenous Swedish subset	1534	1692	120 (129)	146 (151)	0.08	0.09	0.90	0.81
Finnish subset	274	139	40 (41)	15 (15)	0.15	0.11	1.41	0.18

Figure 4.3. Average minor allele frequency for functional annotations from exome sequencing.



Given the variation seen in amount of homozygosity and ROH across differing populations, I also compared rates of two hit LoF in a homogenous Swedish dataset and a Finnish subset. In both cases, I saw no significantly higher rate in cases compared to controls, although similarly to ROH (Table 4.1) the Finnish sample has several times more two hit events (Table 4.3).

4.3.3 Two hit LoF association

I looked to assess if any particular two hit LoF variant was associated to SCZ in my sample. As described in the methods, I implemented a permutation procedure

to test every two hit variant. Of the 194 variants in the full sample, 35 have nominally significant p-values below 0.05 but none surpass the Bonferroni corrected threshold of 0.0002 (0.05/194) (Tables 4.4-4.6, See Appendix A2-4 for full results). I performed the same test within the homogenous dataset and the Finnish dataset with the similar results.

Table 4.4. Nominally significant autosomal recessive variants from all individuals

Var	Gene	Case	Control	MAF	Exp Case	Exp Control	P perm
chr17:26684704..26684706	TMEM199	1	0	3.49E-07	0.0	0.0	0.0008
chr15:102211693..102211695	TARSL2	1	0	6.20E-07	0.0	0.0	0.0011
chr1:19166501..19166515	TAS1R2	1	0	4.75E-07	0.0	0.0	0.0013
chr6:109768320..109768322	MICAL1	1	0	4.75E-07	0.0	0.0	0.0013
chr1:16528294..16528295	ARHGEF19	1	0	7.85E-07	0.0	0.0	0.0014
chr19:36351305	KIRREL2	2	0	2.52E-05	0.1	0.1	0.0017
chr10:73082628..73082629	SLC29A3	1	0	7.85E-07	0.0	0.0	0.0018
chr9:5164266	INSL6	1	0	1.17E-06	0.0	0.0	0.002
chr15:75656531,chr15:75658862	MAN2C1	1	0	7.75E-07	0.0	0.0	0.0025
chr7:5920569	OCM	2	0	3.85E-05	0.2	0.2	0.0036
chr2:26644264	CCDC164	1	0	1.90E-06	0.0	0.0	0.0045
chr21:35094909..35094910	ITSN1	1	0	2.18E-06	0.0	0.0	0.005
chr15:50593417	GABPB1	1	0	3.14E-06	0.0	0.0	0.0054
chr6:83938651,chr6:84108085	ME1	1	0	2.21E-06	0.0	0.0	0.0061
chr1:228469903	OBSCN	2	0	6.36E-05	0.3	0.3	0.0063
chr20:3687141	SIGLEC1	3	0	0.00016886	0.9	0.9	0.0073
chr1:156521765	IQGAP3	1	0	3.88E-06	0.0	0.0	0.0084
chr14:24679928	CHMP4A	1	0	3.14E-06	0.0	0.0	0.0084
chr6:160969693,chr6:161006077	LPA	3	0	0.00024885	1.3	1.3	0.0141
chr1:97915614,chr1:97981498	DPYD	1	0	5.99E-06	0.0	0.0	0.0156
chr3:9859364,chr3:9874914	TTLL3	1	0	7.02E-06	0.0	0.0	0.0159
chr1:40947434	ZFP69	1	0	7.60E-06	0.0	0.0	0.017
chr1:36181910	C1orf216	1	0	8.15E-06	0.0	0.0	0.0193
chr2:231988421	HTR2B	1	0	8.15E-06	0.0	0.0	0.0223
chr3:195452000..195452001	MUC20	1	0	9.92E-06	0.1	0.1	0.025
chr5:148989259	ARHGEF37	2	0	0.00010889	0.6	0.6	0.0265
chr12:31815202	METTL20	1	0	1.26E-05	0.1	0.1	0.0293
chr16:48130781,chr16:48174765	ABCC12	2	0	0.00012799	0.6	0.7	0.0312
chr17:80369330	OGFOD3	1	0	1.40E-05	0.1	0.1	0.0321
chr4:89912303	FAM13A	1	0	1.40E-05	0.1	0.1	0.0352
chr14:94931042	SERPINA9	1	0	1.55E-05	0.1	0.1	0.0374
chr2:108863758,chr2:108881338	SULT1C3	1	0	1.55E-05	0.1	0.1	0.0398
chr9:125239501	OR1J1	2	0	0.00016378	0.8	0.8	0.0418
chr17:37034365	LASP1	1	0	1.96E-05	0.1	0.1	0.0481
chr16:48130781,chr16:48141231	ABCC12	2	1	2.18E-05	0.1	0.1	0.0482

Table 4.5. Nominally significant autosomal recessive variants from homgenous Swedish subset of individuals

Var	Gene	Case	Control	MAF	Exp Case	Exp Control	P perm
chr11:118851208..118851209	FOXR1	1	0	6.01E-07	0.0	0.0	0.0007
chr10:73082628..73082629	SLC29A3	1	0	1.18E-06	0.0	0.0	0.0019
chr15:102211693..102211695	TARSL2	1	0	1.18E-06	0.0	0.0	0.0019
chr15:75656531,chr15:75658862	MAN2C1	1	0	1.08E-06	0.0	0.0	0.0020
chr21:35094909..35094910	ITSN1	1	0	2.40E-06	0.0	0.0	0.0032
chr14:57947421	C14orf105	6	0	0.00147746	4.5	5.0	0.0037
chr2:26644264	CCDC164	1	0	3.46E-06	0.0	0.0	0.0042
chr6:83938651,chr6:84108085	ME1	1	0	3.56E-06	0.0	0.0	0.0051
chr1:156521765	IQGAP3	1	0	4.71E-06	0.0	0.0	0.0064
chr14:24679928	CHMP4A	1	0	6.94E-06	0.0	0.0	0.0100
chr1:97915614,chr1:97981498	DPYD	1	0	7.78E-06	0.0	0.0	0.0121
chr12:31815202	METTL20	1	0	1.27E-05	0.0	0.0	0.0200
chr2:209302328,chr2:209309610	PTH2R	1	0	1.47E-05	0.0	0.0	0.0202
chr16:48130781,chr16:48174765	ABCC12	2	0	1.63E-04	0.5	0.6	0.0220
chr19:35718891,chr19:35719490	FAM187B	1	0	1.51E-05	0.0	0.1	0.0220
chr9:21481483	IFNE	1	0	2.02E-05	0.1	0.1	0.0292
chr14:94931042	SERPINA9	1	0	2.31E-05	0.1	0.1	0.0325
chr17:80369330	OGFOD3	1	0	2.16E-05	0.1	0.1	0.0327
chr2:108863758,chr2:108881338	SULT1C3	1	0	2.51E-05	0.1	0.1	0.0347
chr3:44943164	TGM4	1	0	2.78E-05	0.1	0.1	0.0373
chr1:3544083	TPRG1L	1	0	3.11E-05	0.1	0.1	0.0433
chr6:99885246,chr6:99930627	USP45	5	4	3.19E-05	0.1	0.1	0.0437
chr7:143048771	CLCN1	1	0	3.47E-05	0.1	0.1	0.0479

Table 4.6. Nominally significant autosomal recessive variants from subset of Finnish individuals

Var	Gene	Case	Control	MAF	Exp Case	Exp Control	P perm
chr11:123813863..123813864	OR6T1	1	0	5.86E-06	0.0	0.0	0.0006
chr9:95076685..95076686	NOL8	1	0	5.86E-06	0.0	0.0	0.0006
chr14:68256161..68256162	ZFYVE26	1	0	5.86E-06	0.0	0.0	0.0008
chr7:130038730..130038732	CEP41	1	0	5.86E-06	0.0	0.0	0.0008
chr14:23523813..23523814	CDH24	1	0	5.86E-06	0.0	0.0	0.0008
chr1:152191442..152191462	HRNR	1	0	5.86E-06	0.0	0.0	0.0009
chr11:4566459..4566460	OR52M1	1	0	5.86E-06	0.0	0.0	0.0009
chr3:121341076..121341093	FBXO40	1	0	5.86E-06	0.0	0.0	0.0009
chr9:5164266	INSL6	1	0	5.86E-06	0.0	0.0	0.0012
chr19:57723399..57723400	ZNF264	1	0	5.86E-06	0.0	0.0	0.0014
chr19:47425241..47425243	ARHGAP35	1	0	1.32E-05	0.0	0.0	0.0022
chr16:48130781,chr16:48141231	ABCC12	1	0	5.28E-05	0.0	0.0	0.0152
chr1:228469903	OBSCN	2	0	0.00091605	0.5	0.3	0.0176
chr17:45447802,chr17:45468858	EFCAB13	1	0	7.62E-05	0.0	0.0	0.0203
chr6:160969693,chr6:161006077	LPA	2	0	0.00109633	0.6	0.3	0.0288
chr17:72588438	C17orf77	1	0	0.00014657	0.1	0.0	0.0351

4.3.4 LoF variants on the X chromosome

I next considered the contribution of LoF mutations occurring on the X chromosome. I identified zero two hit LoF variants in females and focused on males. As males carry only a single X chromosome, I reduced the population frequency required for inclusion to 0.0025 which is equivalent to the expected frequency of homozygous events with allele frequency of 5%. In addition, because males are hemizygous and will be inaccurately called homozygous events I removed the HWE restriction. I identified 111 LoF mutations on the X chromosome in males, of which 61 were in cases and 50 were controls, a non-significant difference (one sided Fisher's exact test p=0.46). No variant achieved significance with the most associated variants being present in six cases and one control in the genes *CTAG2* and *FMR1* (Table 4.7).

Table 4.7. Male chrX hemizygous LoF mutations with MAF < 0.0025 (1,520 cases, 1,291 controls)

Var	Gene	Cases	Controls	MAF Female	P
chrX:153880695..153880696	CTAG2	6	1	0.000	0.087
chrX:147019617	FMR1	6	1	0.003	0.097
chrX:142967383	UBE2NL	3	0	0.002	0.152
chrX:135427684..135427685	GPR112	2	0	0.000	0.25
chrX:74514456	UPRT	3	1	0.027	0.44
chrX:2772157	GYG2	1	0	0.000	0.46
chrX:153044080	PLXNB3	1	0	0.000	0.49
chrX:13337978	ATXN3L	4	2	0.002	0.51
chrX:133922863	FAM122B	1	0	0.000	0.51
chrX:35820600..35820601	MAGEB16	1	0	0.000	0.51
chrX:13780462	OFD1	2	1	0.000	0.52
chrX:8763157..8763159	FAM9A	1	0	0.000	0.53
chrX:132092607..132092608	HS6ST2	1	0	0.002	0.53
chrX:134994961	SAGE1	1	0	0.000	0.53
chrX:135593752..135593754	HTATSF1	1	0	0.000	0.54
chrX:70524455	ITGB1BP2	1	0	0.000	0.54
chrX:118221676	KIAA1210	1	0	0.000	0.54
chrX:99933386	SYTL4	1	0	0.000	0.54
chrX:120181950	GLUD2	1	0	0.001	0.57
chrX:13337125	LOC1000936	1	0	0.000	0.57
chrX:30237351..30237352	MAGEB2	1	0	0.000	0.57
chrX:49035692	PRICKLE3	1	0	0.000	0.57
chrX:71349860	RGAG4	1	0	0.001	0.57
chrX:100088442	CSTF2	1	0	0.000	0.58
chrX:19443740	MAP3K15	1	0	0.000	0.58
chrX:24552072..24552073	PDK3	1	0	0.000	0.58
chrX:31089629	FTHL17	2	1	0.001	0.59
chrX:5821139	NLGN4X	1	0	0.000	0.59
chrX:54784856	ITIH6	1	0	0.000	0.63
chrX:64722532..64722536	ZC3H12B	1	1	0.000	0.76
chrX:140785773	SPANXD,SPA	1	1	0.000	0.8
chrX:41586494..41586498	GPR82	1	1	0.000	0.82
chrX:69479172	P2RY4	2	3	0.001	0.83
chrX:19389113	MAP3K15	3	4	0.003	0.84
chrX:135312997	MAP7D3	1	1	0.001	0.85
chrX:57475022	FAAH2	1	2	0.001	0.91
chrX:15333546	ASB11	2	4	0.001	0.92
chrX:110951511	ALG13	0	1	0.000	1
chrX:47430409	ARAF	0	1	0.000	1
chrX:100243424	ARL13A	0	2	0.001	1
chrX:39930346..39930347	BCOR	0	1	0.000	1
chrX:134947910	CT45A5	0	1	0.000	1
chrX:45050995	CXorf36	0	1	0.000	1
chrX:8763164..8763166	FAM9A	0	1	0.000	1
chrX:48631801	GLOD5	0	1	0.000	1
chrX:114541092..114541096	LUZP4	0	3	0.000	1
chrX:19410128..19410130	MAP3K15	0	1	0.000	1
chrX:20062644	MAP7D2	0	1	0.000	1
chrX:100117666	NOX1	0	2	0.001	1
chrX:100118566	NOX1	0	1	0.000	1
chrX:77224385..77224401	PGAM4	0	1	0.000	1
chrX:152226619	PNMA3	0	2	0.001	1
chrX:37312605..37312606	PRRG1	0	2	0.000	1
chrX:150869254	PRRG3	0	1	0.000	1
chrX:16669120	S100G	0	1	0.000	1
chrX:65242332	VSIG4	0	1	0.000	1
chrX:47919461	ZNF630	0	1	0.000	1
Total		**61**	**50**		**0.46**

4.3.5 Pathway tests

While schizophrenia cases do not carry more two hit LoF than controls it does not preclude the contribution of specific mutations affecting particular functional pathways that increase risk to disease. Therefore, I selected 2 gene sets with prior implications to psychiatric disease. These include synaptic, and *de novo* sets of genes (see methods for descriptions). In total, I tested 10 gene sets that had at least one two-hit event across two categories (Table 4.8). The only significant result is from a set of LoF *de novo* mutations from multiple trio sequencing studies in SCZ (p=0.014).

Table 4.8. Aggregate tests of autosomal two-hit variants in selected gene sets

Category	Gene set	#genes	Case	Control	P	Individual genes
	SCZ-LoF	**136**	**4**	**0**	**0.014**	CLCN1 (1:0), ITSN1 (1:0), DPYD (1:0), STAP2 (1:0)
	SIB_CONTROL-NS	566	16	12	0.24	
De novo sets	SCZ-NS	988	10	7	0.4	
	AUT-LoF	186	1	5	0.7	
	AUT-NS	1088	7	13	0.9	
	Pre-synaptic_active_zone	79	4	2	0.13	
	PSD_(human_core)	493	1	0	0.4	
Synaptic genes	Pre-synapse	295	4	3	0.4	
	PSD	1058	7	6	0.4	
	Synaptic_vesicle	211	0	1	1	

4.3.6 Recessive variants underlying ROH

Lastly, I identified homozygous LoF variants within stretches of previously identified ROH to ask whether case ROH were pointing to variants related to schizophrenia. In general, I saw increased rates of ROH in individuals carrying two hit LoF mutations (Table 4.9). I took all 4,116 ROH and asked if they contained a homozygous LoF in the same individual (Table 4.10). I found 16 situations where this occurred. Of those 16, 6 occurred in cases and 10 occurred in controls. Only a

single gene had more than one individual with *ATP2C2* being seen in both a case and a control. Of note, none of these events occurred in the region with increased rates of case ROH discussed previously.

Table 4.9. Rates of ROH stratified by number of two hit LoF mutations carried

	Full Sample				Homogenous Swedish				Finnish			
	Case		Control		Case		Control		Case		Control	
# two-hit	N	Rate	N	Rate	N	Rate	N	Rate	N	Rate	N	Rate
0	4779	0.40	6003	0.33	3234	0.20	4760	0.20	388	0.86	193	0.92
1	208	0.50	227	0.45	142	0.37	163	0.34	25	1.08	12	0.67
2	13	0.23	11	1.55	10	0.10	9	0.89	1	1.00	0	N/A
3	0	N/A	1	0.00	0	N/A	1	0	N/A	N/A	N/A	N/A

Table 4.10. Two hit LoF variants within ROH in same individual

Gene	Var	ID	Status	Chr	Start	End
ABCC12	chr16:48174765	PT-L1FP	control	16	46848028	67358816
ATP2C2	chr16:84495318	PT-8VQW	case	16	81417366	86070195
ATP2C2	chr16:84495318	PT-8W95	control	16	82671409	85523857
CC2D2A	chr4:15482360	PT-ERVH	control	4	9889448	32635777
CLCN1	chr7:143048771	PT-ESQR	case	7	138429851	144720781
CSRP2BP	chr20:18142462	PT-9ZIK	control	20	17883827	26239964
CYP2A13	chr19:41594954	PT-FG37	control	19	38444986	46860058
FCN3	chr1:27700455	PT-CDG1	control	1	1130727	35368760
GABPB1	chr15:50593417	PT-8U75	case	15	44387383	58156552
HRCT1	chr9:35906438	PT-BRM1	control	9	33635164	38766174
OBSCN	chr1:228469903	PT-L1SZ	case	1	226683251	234541801
RAB25	chr1:156035856	PT-IT7L	control	1	153294653	159910337
RNF32	chr7:156468559	PT-IRW6	control	7	151643909	157930785
TNK1	chr17:7290695	PT-BSB6	control	17	4488884	9268502
TRIM31	chr6:30071363	PT-BRFP	case	6	25106783	43405525
UGT2A1	chr4:70512787	PT-IT7J	case	4	52864476	73195704

4.4 Discussion

I looked to assess the role of two-hit LoF variants in risk of schizophrenia. I utilized a large, mostly homogenous sample that has been both heavily genotyped for common SNPs and of which a large proportion of samples have been exome sequenced. Previous literature has pointed to autozygous regions as being implicated in SCZ pointing to a mechanism for contribution of recessive variants. I attempted to replicate this result in my sample but was unable to likely due to smaller sample size. I then looked for particular overlap of any region predominantly in cases and identified a few nominally significant regions but none withstanding genome-wide correction. I next shifted attention to specific variants from exome sequencing and found no case enrichment of two hit LoF mutations, nor did I see any associated variant after correcting for multiple testing. Finally, I looked to understand the relationship between large regions of autozygosity from recent inbreeding by looking at recessive mutations underlying ROH in the same individual but only 16 out of 496 variants were in this category and most of them were in controls. In general, I saw at most a limited role of two-hit LoF mutations in risk to SCZ.

Inbreeding results in increased rates of autozygosity which is a known cause of numerous disorders. In my sample of over 11,000 individuals I identified no discernible difference in number or amount of ROH between cases and controls after accounting for ancestry. I further noted as has been seen before that the Finnish subset of my sample is enriched for ROH indicating more recent reduced diversity of

gene pool due to a population bottleneck. I also didn't identify any particular regions of the genome enriched for ROH in cases that withstand Bonferroni correction. There are however several regions with nominal significance that could be followed up on in larger samples.

A closer look at specific homozygous variants and compound heterozygous events can be done using sequencing data. Since my goal was to select a small pool of likely deleterious variants I focused on those variants that were both putative LoF and rare (MAF < 5%). I identified only 194 unique two hit variants seen 496 times across 5,079 individuals, a rate of less than 0.1 per individual. For comparison, the expected rate of *de novo* mutations within the exome is just under 1 per individual or almost 9 times more frequent than rare double hit LoF variants. I saw no significantly increased rates of these events in cases across the full dataset, homogenous Swedish dataset or the Finnish dataset. Although, as with the ROH, I saw several fold more double hit events in the Finnish sample. It was also the only sample with an increased rate of these variants in cases compared to controls (OR=1.4). Studying populations that have undergone recent bottlenecks or have higher rates of recent inbreeding will increase the total number of recessive events and therefore increase power to detect case/control differences. After testing each variant using permutation, I identified a handful of variants that are nominally significant but none withstood multiple testing correction. Outside of the Finnish sample, I saw lower than expected numbers of two hit mutations. Whether this is a product of inaccurate allele frequencies at small numbers, under-calling of

homozygous mutations or a reduction of recessive mutations for LoF variants remains to be tested.

Differences in prevalence and severity of disease between genders have been shown (105) and a possible mechanism for these could be sex linked mutations. I therefore identified all LoF mutations with MAF < 0.0025 on the X chromosomes in males. Again, I identified no significant enrichment of LoF mutations in cases compared to controls. Of the 57 variants, only a few approached significance, one of which was *FMR1* (6 cases and 1 control, p=0.097). A trinucleotide expansion in *FMR1* (Homo sapiens fragile X mental retardation 1) is a known cause of Fragile X Syndrome (106), and mutations in this gene have also been linked to autism (107).

Two hit LoF mutations in genes such as *FMR1* and other nominally significant genes could be pointing to a set of very rare syndromic events in which the symptoms of schizophrenia are just one possible outcome to that mutation. As in large, rare CNVs such as 22q11.2, where only 30% of individuals with that deletion have psychosis while most suffer from a range of diseases including physical deformation and heart problems. This genetic heterogeneity also fits into the highly polygenicity of SCZ whereby instead of the aggregate of many loci of small effect hitting many genes, I have a single locus of strong effect hitting many genes. However, the lack of increased rates of these events presents a limited role at most for these mutations and perhaps it is only in the context of the individual's genetic background that these mutations will have an effect on disease risk. I did however identify a certain set of genes with increased rates of two hit LoF mutations in cases including previously identified genes with *de novo* LoF from SCZ trios. This result

implicates that a small subset of these variants are likely relevant for risk of SCZ but larger numbers will be required to assert this statistically, perhaps by expanding to deleterious non-synonymous variants.

Finally, the results presented here are in line with previous findings of rare, deleterious variants in SCZ such as *de novo* variants where contribution of the majority of variants will be minimal. This differs substantially with what has been seen in autism and intellectual disability indicating a likely difference in genetic architecture of these disorders.

Chapter 5. Aggregation across variant classes to identify specific genes related to SCZ

5.1 Introduction

Over the last several years it has been demonstrated convincingly that schizophrenia is a highly polygenic disorder consisting of thousands or possibly tens of thousands of risk variants contributing to disease (51). Polygenicity has been shown not only in common SNPs but also in CNVs and in very rare mutations that affect translation (disruptive) mutations (33). Recent efforts both in this dissertation and in the literature have been aimed at reducing the vast search space of variants to a small set of functionally related genes. This has been shown both in common SNPs as in Chapter 3 and CNVs both in Chapter 2 and elsewhere (38). The goal of these analyses is to both increase power to detect enrichment of variants in cases by incorporating larger numbers of variants and to aid in understanding the etiology of the disease by testing biologically related genes. While productive and beneficial to understanding the disorder, these approaches may not point to any particular genes that could be studied further to possibly identify therapeutic targets. As more and more data are collected more variants will be identified for any given gene and testing individual genes will be better powered. One analytical strategy that can be employed immediately to increase the power to test specific genes is the aggregation of all variant data regardless of class and technology used. For example, it is possible that rare CNVs and rare disruptive mutations are occurring in the same genes and by testing those sets of variants together we are more powered to identify a particular

gene contributing to SCZ risk. Here, I develop a log likelihood case/control test that surveys each gene for whether there are more or fewer mutations of various variant classes than expected after incorporating both gene length and variant length.

Complete understanding of the role of genetics in disease will require analyzing all forms of variation in concert. Just as identifying recessive variants could contribute to identifying genes related to disease, an individual carrying a large deletion and a disruptive mutation in the same gene can only be appropriately assessed in the context of analyzing both the CNV and the SNV data together. While this specific case is likely to be rare at present the expected increase in sequencing will drastically increase these scenarios. Incorporating all variants within the same gene, one can fully test genetic variation as opposed to limiting a test to common or rare or structural variation. The degree at which these variants occur in the same genes is as yet unclear. An initial look into the 11 previously identified large effect CNVs (37) found no significant overlapping GWAS results with p-value < 10^{-7} in any psychiatric disease based on the GWAS catalog (16). This result could be representing limited power due to high polygenicity or it could be pointing to different mechanisms such as relationship between gene function, disease risk and tolerance to mutation. Here, I look to increase power to detect specific gene associations by aggregating multiple classes of variation. I specifically focus on the most deleterious variants I have, including the rarest class of disruptive mutations from samples that have been exome sequenced and large, rare deletions called from genotyping arrays. I limit my analyses to these putative "loss of function" variants as they are likely to have similar deleterious effects on gene function and translation.

5.2 Methods

5.2.1 Data quality control

I combined data from CNV calling and exome sequencing for the same 5,079 Swedish samples described in chapter 4 which overlap those from chapters 2 and 3. We performed CNV QC as described in chapter 2, that is, we selected only CNVs with length greater than 20kb and having LOD scores greater than ten and removed CNVs overlapping known common variants seen at greater than 1% in the Database of Genomic Variants (108). We removed individuals for having an excessive number or amount of CNV (>30 CNVs or > 10mb of CNV) and retained only CNVs seen in less than 1% of the sample (QC of this sample was performed by Colm O'Dushlaine). To test a specific hypothesis regarding disruptive or "loss of function" mutations I also limited my analysis to deletions only. After QC, there were 3,047 rare deletions remaining.

QC of the exome sequencing data was performed in the same manner as described in chapter 3 with a few minor differences. I restricted the set of variants to those with high genotype qualities scores (GQ > 30) and that were seen only a single time in the dataset. GQ represents the scaled likelihood that an alternative allele is true, every 10 units of GQ is an order of magnitude of likelihood (GQ of 30 = 10^{-3}). I additionally reduced my set of variants to the rarest possible class of disruptive

mutations as those are the most likely to have deleterious effects. This QC procedure resulted in 12,386 singletons, disruptive SNVs.

5.2.2 Likelihood ratio test for variable length variants

I sought to define a method that accounts for both the variable length of genes as well as variable length of variants, in order to model the likelihood of seeing *x* variants with lengths $l_1, l_2 \ldots l_x$ within a gene of length *y*. As I aimed to test individual genes, I restricted my analysis to only coding sequence and defined the length of variants as the number of coding bases they overlap. Therefore a 10,000 base pair deletion that completely knocks out two exons of length 300 bp and 400 bp would be ascribed a variant length of 700 bp. I define *g* to be the span of coding sequence at 32mb, *p* is the length of the coding sequence of a gene, *q* is then the remainder of all coding sequence not consisting of *p (q=g-p),* l represents the length of the variant as described above. *P* is the probability of a variant of length *l* hitting a gene of length *p* by taking number of possible locations that the variant could overlap the gene *(p+l-1)* and dividing by the number of possible locations of the variant across the exome *(g-l+1).* I define a parameter λ as the effect size where $\lambda > 1$ signifies more than expected mutation within the gene.

$$x = number\ of\ variants\ hitting\ a\ gene, \qquad N = total\ number\ of\ variants$$

$$G = total\ amount\ of\ coding\ sequence, \qquad l = variant\ length, \qquad p = gene\ length,$$

$$\lambda = effect\ size$$

Amount of coding sequence NOT comprised of a gene with length p: $q = G - p$,

The proportion of coding sequence start sites for which a variant of length l will overlap a gene of length p:

$$P = \frac{p + l - 1}{G - l + 1}$$

The proportion ofcoding sequence start sites for which a variant of length l will NOT overlap a gene of length p:

$$Q = \frac{q - 2(l - 1)}{G - l + 1}$$

The probability of variant with length l overlapping gene with length p where λ *is the relative enrichment for start sites where the variant overlaps the gene:*

$$r = \frac{\lambda P}{\lambda P + Q}$$

Likelihood if all variants have length 1: $L(\hat{\lambda}) = r^x(1 - r)^{N-x}$

Likelihood for variants of different lengths: $L(\hat{\lambda}) = \prod_{i=1}^{x} r_i * \prod_{j=x+1}^{N} 1 - r_j$

Solving for mle of λ*, we take the derivative of the log*

$$lnL(\hat{\lambda}) = \sum_{i=1}^{x} \ln(r_i) + \sum_{j=x+1}^{N} \ln(1 - r_j) = 0$$

$$\frac{d}{dr} = \sum_{i=1}^{x} 1/r_i - \sum_{j=x+1}^{N} 1/(1 - r_j) = 0$$

$$\sum_{i=1}^{x} (\lambda P_i + Q_i)/\lambda P_i - \sum_{j=x+1}^{N} (\lambda P_j + Q_j)/Q_j = 0$$

results in a quadratic in λ *with the following solution*

$$a = -1 * \prod_{i=1}^{x} P_i * \sum_{j=x+1}^{N} (P_j * \prod_{k=x+1, k \neq j}^{N} Q_k)$$

$$b = (2x - N) * \prod_{i=1}^{x} P_i * \prod_{k=x+1}^{N} Q_k$$

$$c = \prod_{i=x+1}^{N} Q_i * \sum_{j=1}^{x} (Q_j * \prod_{k=1,k\neq j}^{x} P_k)$$

$$\hat{\lambda}_{mle} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

We calculate the mle of λ for cases, controls and the full sample and determine likelihoods for each

$$L(\hat{\lambda}_{case}) = \prod_{i=1}^{x_{case}} (\lambda_{case} P_i / (\lambda_{case} P_i + Q_i)) \prod_{j=1}^{N_{case} - x_{case}} (Q_j / (\lambda_{case} P_j + Q_j))$$

$$L(\hat{\lambda}_{control}) = \prod_{i=1}^{x_{control}} (\lambda_{control} P_i / (\lambda_{control} P_i + Q_i)) \prod_{j=1}^{N_{control} - x_{control}} (Q_j / (\lambda_{control} P_j + Q_j))$$

$$L(\hat{\lambda}_{all}) = \prod_{i=1}^{x_{all}} (\lambda_{all} P_i / (\lambda_{all} P_i + Q_i)) \prod_{j=1}^{N_{all} - x_{all}} (Q_j / (\lambda_{all} P_j + Q_j))$$

We then calculate a chi squared with 1 degree of freedom using the log likelihood test below

$$\chi^2 = 2lnL(\lambda_{case}) + 2lnL(\lambda_{control}) - 2ln\mathrm{L}(\lambda_{all})$$

5.2.3 Permutation

In an effort to assess the statistical robustness of my test, I proceeded to test 100 iterations with case and control label randomly assigned, thereby breaking the

correlation between disease status and variant. Each iteration, I permuted phenotype and tested each of the 868 genes that carried at least 1 CNV and 1 SNV (86,800 total tests). I assessed the behavior of my test at both the nominally significant p-value of 0.05 and a more significant value of 0.01. My permutation results demonstrated that my test is robust if not conservative at lower significant levels with almost 6% of all genes having an asymptotic p-value < 0.05 and only 0.005 percent of genes having an asymptotic p-value below 0.01 (Figure 5.1). I also directly compared the asymptotic p-values to those empirically derived from the permutations and identified a correlation of 0.84 (Figure 5.2). Clusters of genes represent those genes with limited numbers of variants and minimal power to accurately estimate significance. Additionally, I identified no significant correlation between gene length and –log10(p-value). Finally, I present an empirical qq-plot, that is, I compared my asymptotic p-values to the mean value across all permutations after sorting the p-values for each permutation independently (Figure 5.3). In this manner, I directly compared my observed p-values to those expected across all permutations. I saw no significant inflation that would indicate issues with my test. With these results, I deemed it appropriate to use the asymptotic p-values.

Figure 5.1. Observed proportion of significant genes in SNV, deletions and combined compared to permutation results.

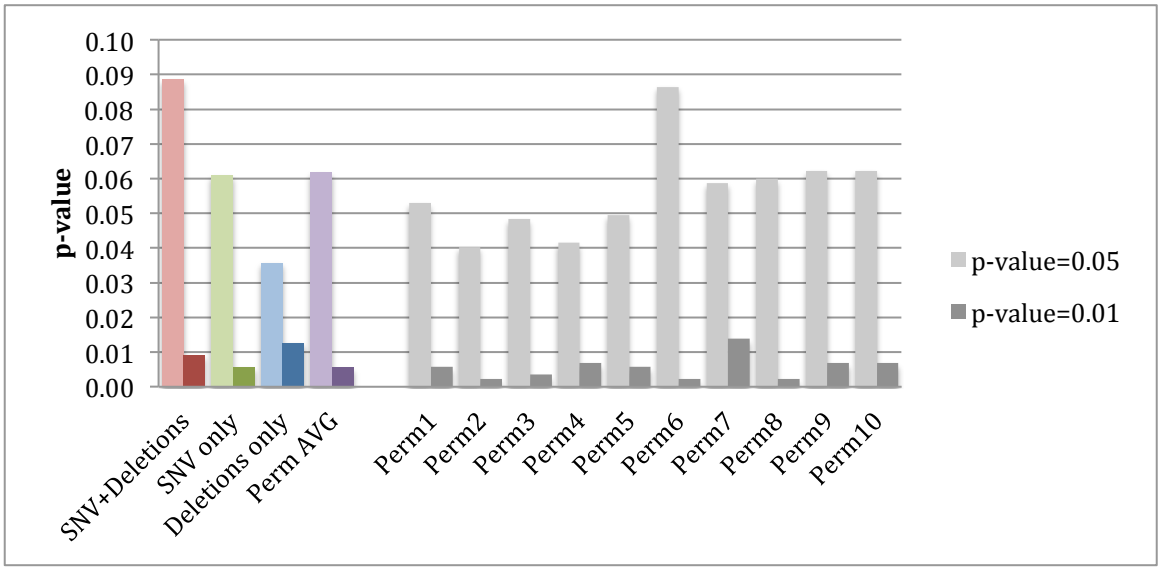


Figure 5.2. Asymptotic p-value compared to empirical p-value based on 100 permutations (r=0.84).

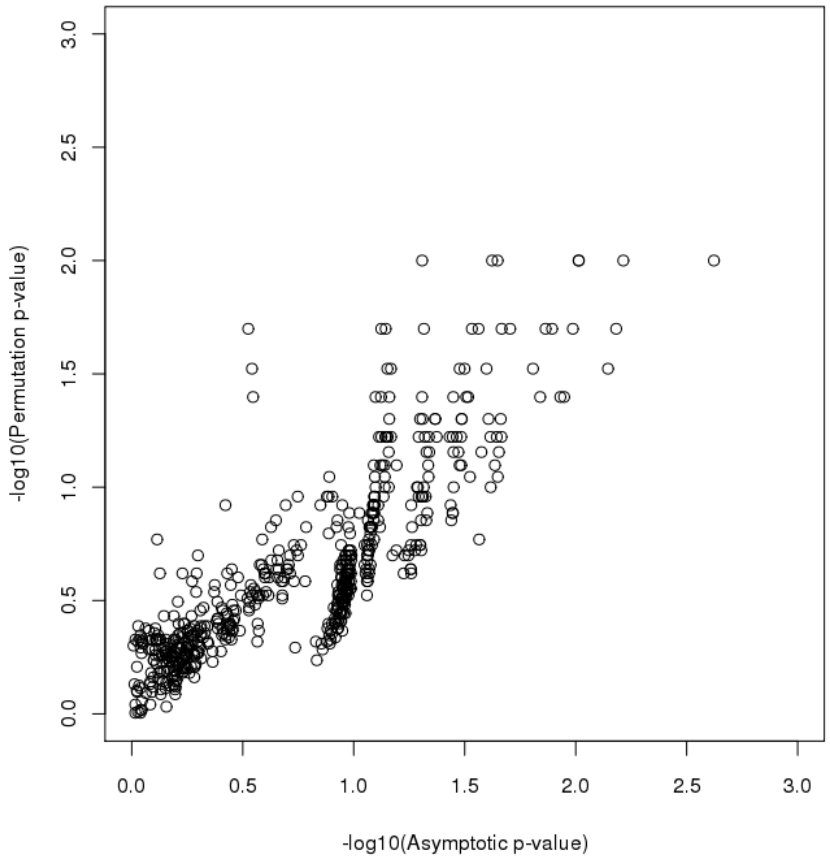
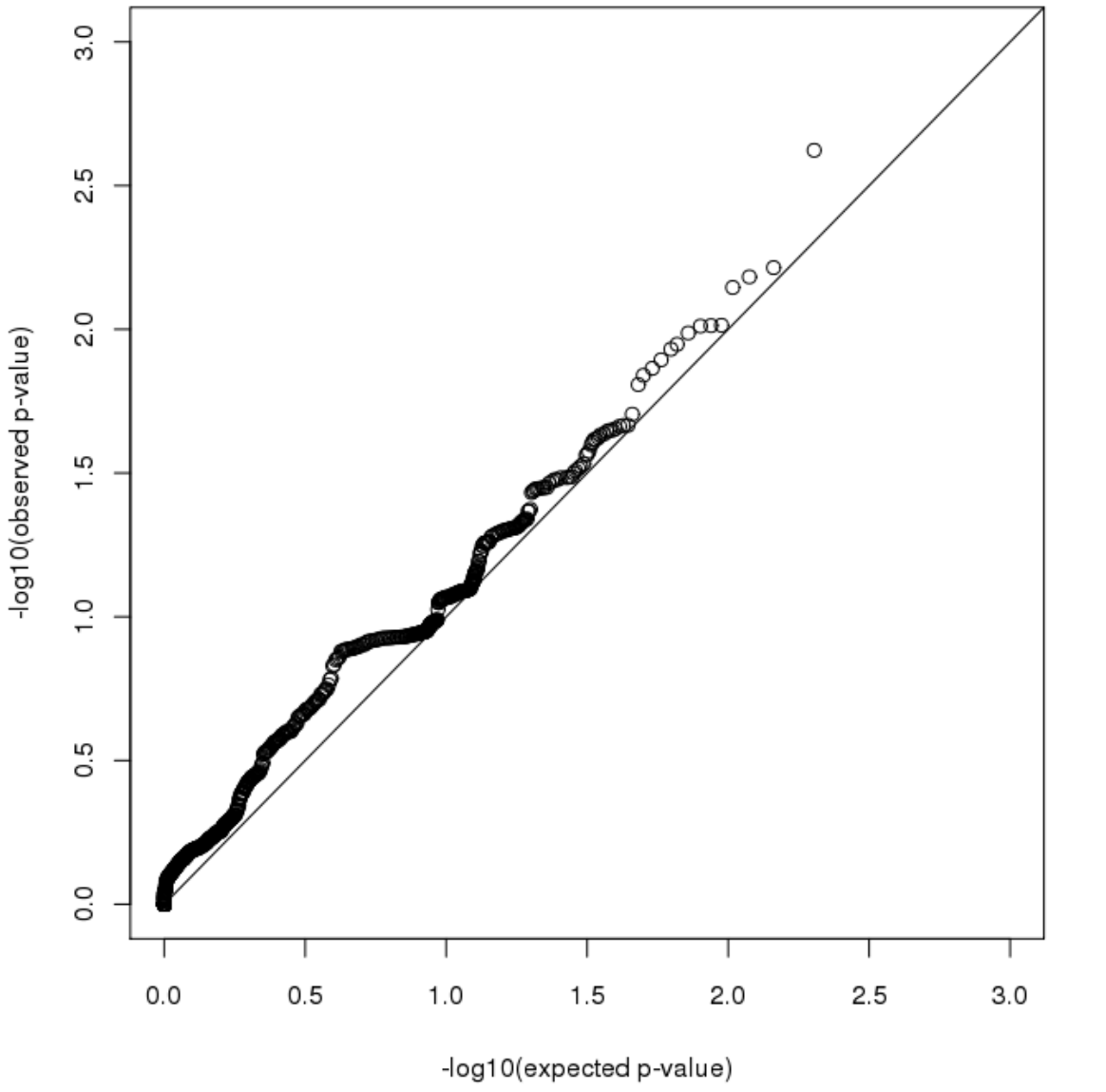


Figure 5.3. Observed genic p-values compared to expected based on results of 100 permutations.



5.3 Results

5.3.1 Overlap of genes containing rare deletions and rare LoF variants

I first wanted to take the 3,047 deletions and 12,386 singleton disruptive mutations and understand how often these variants fall within the same genes. There were 2,259 genes hit by at least one deletion, 7,852 genes hit by at least one disruptive singleton mutation and 868 genes hit with at least one of each. I next hypothesized that genes carrying only case singleton disruptive mutations were more likely to have case only deletions compared to controls. I define N:0 to refer to scenarios where there is greater than 0 case mutations and 0 control mutations and 0:N to be greater than 0 control mutations and 0 case mutations. There were 3,111 genes that are N:0 with respect to singleton disruptive mutations and 2,923 that were 0:N (Table 5.1). Of those with at least one deletion, there were 312 N:0 and 307 0:N, not significantly different from expected (Fisher's exact test one sided p=0.742). Of the 312 N:0 genes, 172 had only case deletions and 82 had only control deletions, this result was significant when comparing to genes with both a singleton disruptive and a deletion (440 case specific with > 0 SNV, 267 control specific with > 0 SNV, Fisher's exact test one sided p=0.015) but not when comparing to all case specific and control specific genes with at least one deletion (1185 case specific with > 0

Table 5.1. Overlap between rare deletions and singleton disruptive mutations.

		>0 CNV	N:0 CNV	0:N CNV	fisher p-value genes with SNV	fisher p-value all deletion genes
		2259	1185	614		
>0 SNV	7852	868	440	267		0.9961
N:0 SNV	3111	312	172	82	**0.01467**	0.276
0:N SNV	2923	307	153	99	0.7588	0.9726
fisher p-value genes with CNV			0.1079	0.9573		
fisher p-value all SNV genes		0.7418	0.3268	0.9629		

deletion, 614 control specific with > 0 deletion, p=0.276) (Table 5.1). I saw no control specific significance from any comparison.

5.3.2 Combined rare CNV + LoF variants case/control tests

With some indication that rare deletions and rare disruptive mutations might be occurring in some of the same genes, I sought to specifically test the likelihood of seeing X mutations, whether deletion or SNV, in gene Y. I focused only on genes carrying both classes of mutation as I assumed the individual variant classes were tested appropriately and I was looking for novel findings. As described in the methods, I adapted a log likelihood framework used to test the likelihood of seeing X mutations in a gene of length Y to incorporate both variable length mutations (e.g. deletions) and directly assess differences between cases and controls. With this test I can now ask the question, what is the likelihood of seeing X variants of variable lengths $L_1, L_2 \ldots L_x$, in a gene of length Y in both cases and controls and test whether it's significantly different from expected. I tested all 868 genes carrying both a singleton disruptive mutation and a rare deletion (genes with p-value < 0.01 Table 5.2, see Appendix A5 for full results). I identified 76 genes reaching a nominally significant level of $p < 0.05$, this represented almost 9% of all genes which was not significantly more then expected from the permutations, as 11 of 100 permutations had >= 76 genes with nominal significance (p-value = 0.11). I also identified 8 genes at $p < 0.01$ which was twice as many as expected but was also not significant (p=0.13). Of these eight genes, the most significant gene was *MS4A10* (7 case deletions, 2 case disruptives, 0 control events, asymptotic p=0.0006). *MS4A10* is a

member of the membrane-spanning 4-domains, subfamily A (MS4A) gene cluster. This gene cluster has shown genome-wide association to Alzheimers disease (109) but has otherwise been poorly characterized. Only three of these eight genes carried more variants in cases compared to controls (*MS4A10*, *SIGLEC12* and *DGCR14*), the other 6 carried more variants in controls. *DGCR14* is a gene located within the Velo-Cardio-Facial Syndrome (VCFS) and DiGeorge Syndrome critical region on chromosome 22 which is characterized by developmental disorders, heart issues and facial anomalies. This syndrome also increases risk for schizophrenia and large deletions have been associated with disease (110). In addition, a family study in 2006 specifically implicated this gene using transmission data from a Han Chinese population (111). While several CNVs have been implicated as likely increasing risk to SCZ, they are nearly always large, carrying many genes, making it difficult to identify if a single gene is contributing to disease risk. This approach could help fine map these regions by assessing the total contribution of variation across the gene to disease where variants like disruptive mutations could contribute. *DGCR14* represents a possible example but additional genes could be identified in this manner.

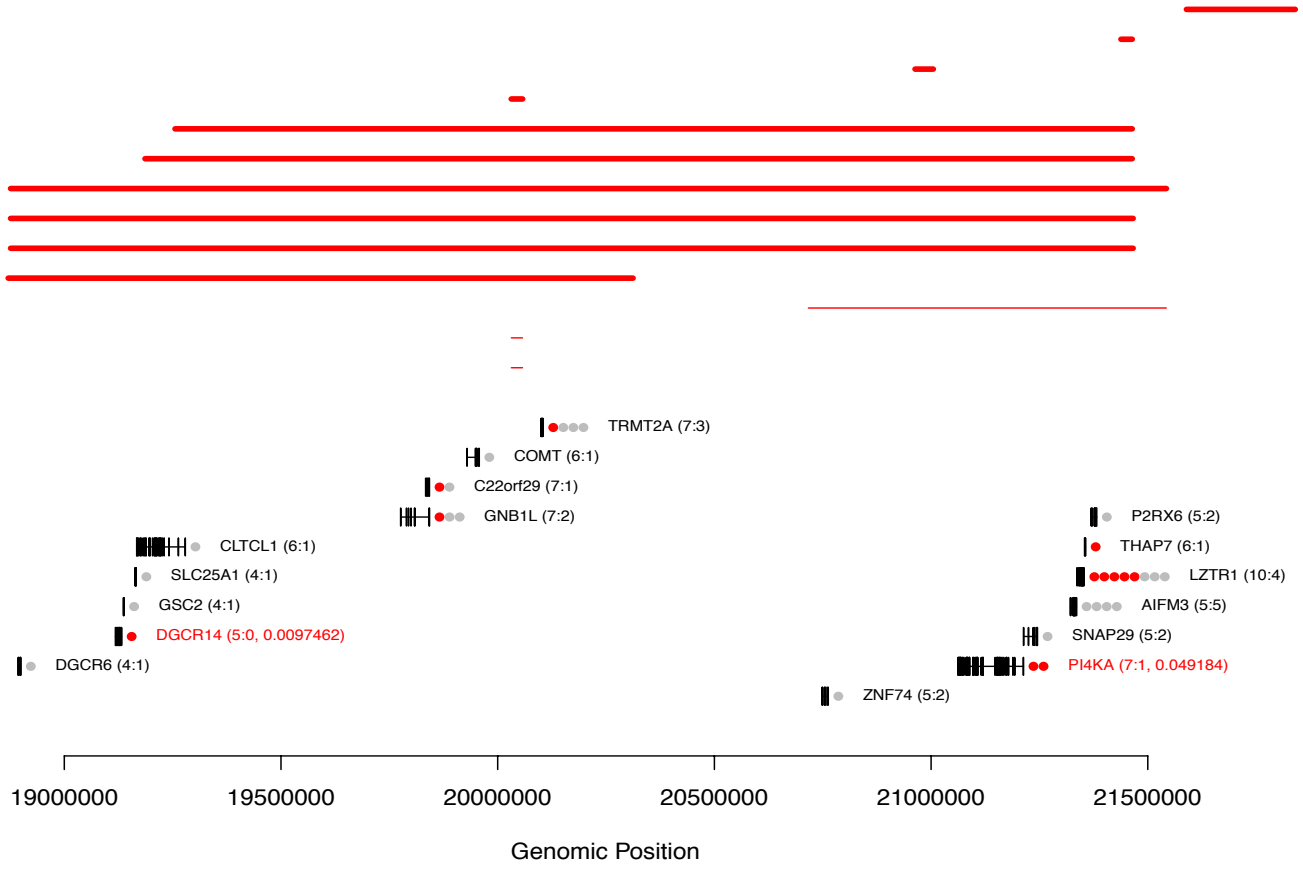
Table 5.2. Significant genic results (p<0.01) for disruptive singletons and rare deletions combined

									SNV only			CNV only		
Gene	**chr**	**start**	**end**	**gene length**	**case**	**control**	**all**	**p-value**	**case**	**control**	**p-value**	**case**	**control**	**p-value**
MS4A10	11	60552820	60568778	2296	9:7458	0:7189	9:14647	0.0006	2:6266	0:6120	0.0987	7:1192	0:1069	0.0050
PFKP	10	3110818	3178997	2781	0:7458	6:7189	6:14647	0.0024	0:6266	3:6120	0.0397	0:1192	3:1069	0.0128
POLRMT	19	617222	633568	3800	0:7458	5:7189	5:14647	0.0061	0:6266	4:6120	0.0175	0:1192	1:1069	0.1985
SIGLEC12	19	51994612	52005043	2126	6:7458	0:7189	6:14647	0.0066	4:6266	0:6120	0.0195	2:1192	0:1069	0.1456
DNAH2	17	7623038	7737058	13505	0:7458	5:7189	5:14647	0.0072	0:6266	4:6120	0.0175	0:1192	1:1069	0.2120
UGT1A7	2	234590583	234681945	2333	1:7458	8:7189	9:14647	0.0097	0:6266	8:6120	0.0008	1:1192	0:1069	0.3053
UGT1A9	2	234580543	234681951	2376	1:7458	8:7189	9:14647	0.0097	0:6266	8:6120	0.0008	1:1192	0:1069	0.3050
DGCR14	22	19117791	19132190	5385	5:7458	0:7189	5:14647	0.0097	1:6266	0:6120	0.2430	4:1192	0:1069	0.0240

5.3.4 Fine mapping known psychiatric CNVs

I selected six previously identified deletions implicated in SCZ to follow up further (37) (see Appendix A5 for full results). These included 1q21.1, *NRXN1*, 3q29, 15q13.3, 17q12 and 22q11.2. Of these, four had at least one gene where at least one individual carried a deletion and at least one individual carried a disruptive singleton mutation. Two of these deletion regions contained a gene that reached nominal significance ($p<0.05$), these being *DGCR14* at 22q11.2 and several genes at 3q29. There were 16 genes tested in the 22q11.2 (DiGeorge, VCFS) region, all but *DGCR14* had at least one control singleton disruptive mutation and only *DGCR14* was significant (4 case deletions, 1 case singleton disruptive mutation, 0 control events, $p=0.0097$, Figure 5.4). The 3q29 region includes eleven genes that were tested of which four were nominally significant (*RNF168*, *PAK2*, *PIGZ*, *DLG1*). Six of the other seven carry at least one control mutation, the gene that did not is just over my definition of nominally significant (*SLC51A* $p=0.0504$). Each of the four significant genes in this region were hit by the same 2 large deletions, three of them also carried a single SNV while the fourth and most significant gene *RNF168* carried two SNVs. *RNF168* is a known cause of Riddle Syndrome which among other symptoms often includes learning difficulties. A second gene (*DLG1*) in this region has significant biological implications for psychiatric disease playing a role in post synaptic density and being implicated by *de novo* CNVs as contributing to SCZ (36, 38). Both these genes represented ideal candidates for follow up in larger cohorts.

Figure 5.4. Fine mapped 22q11.2 region. Red lines indicate deletions; think lines are case deletions thin lines are control deletions. Each gene in the region is below, vertical lines distinguish exons, dots next to the gene represent each singleton disruptive mutation; red dots are case mutations; grey dots are control mutations. Total number of case and control events are listed for every gene, genes with nominally significant p-values are in red and include p-value.



5.4 Discussion

In this chapter, I looked to aggregate data across multiple variant classes including large, rare deletions and disruptive singleton mutations to increase power to identify specific genes contributing to SCZ disease risk. I identified 868 genes with at least one variant of both classes and showed that there is some significant overlap

between genes with only case disruptive mutations and genes with only case deletions and no respective overlap for control only genes. I developed an analytical framework to test each gene for significant enrichment of case or control mutations compared to expectation after accounting for both gene length and variant lengths. I demonstrated that this test provides robust asymptotic p-values and showed a non-significant increase in nominally significant genes in my sample of 5,000 individuals compared to expectation from permutation. I identified 76 genes reaching nominal significance ($p<0.05$) of which no gene reached significance after Bonferroni correction ($0.05/848$, $p<5.8e-4$) or an empirically determined correction, that is the 5% most significant p-value seen across all 100 permutations ($p<0.0002$). In addition, I took previously identified CNVs demonstrated to have significant association with high penetrance in psychiatric disease and identified two regions containing nominally significant genes at 3q29 and 22q11.2 indicating two genes with higher likelihood of disease relevance.

The question of how strongly variants of different classes and different frequency spectrums implicate the same genes is still fundamentally unknown. My preliminary analysis comparing overlap of singleton disruptive mutations and rare deletions only presents a limited case for this overlap. The most likely explanation for this is the highly polygenic nature of SCZ genetic susceptibility, such that while there exists burden across these variants overall it will be scattered across many genes contributing to disease risk. Any overlap will be further diluted by most of the variants not being disease relevant. Only by increasing the number of disease variants being tested within a gene will it be possible to identify associated genes. In

my case, I am attempting to increase power by expanding my set of variants. When looking at my combined CNV and disruptive SNV tests, I did not see significantly more genes at a nominal p-value than I would expect by chance, revealing that the sample is still underpowered.

I implemented a framework to test genes for both the likelihood of being hit by mutations of variable length and for difference in likelihood between cases and controls. This flexible framework allows us to incorporate variants of any length including, SNVs, CNVs, indels, inversions, transitions, SNPs, etc. Here, I focused on only rare disruptive singleton mutations and rare deletions with the goal of identifying specific genes significantly associated with SCZ. The most significant gene by an order of magnitude was *MS4A10* which is part of a cluster of genes near a locus reported as genome-wide significant in late-onset Alzheimers disease (109). In addition, this gene contained no variants in a public repository for structural variation (Database of Genomic Variants) that overlap coding regions indicating that deleterious mutations in this gene are rare. This gene is poorly characterized but could be a candidate for follow up in larger studies and also functional work to assess its biological role.

Combining CNVs and SNVs also provided an opportunity to fine map regions of previous association. In this case I selected a set of six deletions that have strong implications for SCZ and looked to identify if there are any particular genes in these regions with stronger or weaker evidence. My results indicate that in at least two CNV regions of large effect in SCZ I can prioritize certain genes. The variant with the largest known effect for SCZ is a deletion on 22q11.2, a known cause of VCFS and/or

DiGeorge Syndrome, of which 30% of individuals carrying this deletion have SCZ. This dataset contained four individuals with this deletion, but only a single gene in this region was more than nominally significant (*DGCR14*, 4 case deletions, 1 case SNV, p=0.0097). All other genes carried at least one disruptive singleton mutation in controls. I believe this result implicates *DGCR14* for further study but is neither conclusive nor excludes the idea that it is the deletion of multiple genes in this region that is contributing to SCZ risk. Expanding this approach to study pairs or combinations of genes with mutations could help answer this question.

In this chapter I have developed a flexible framework for assessing the association of any gene to disease incorporating variants of any type or length. I made several assumptions that could be directly assessed in future work. I assumed a fixed exome size for all variants based on the targets of the sequencing study despite the fact that CNV calling is not constrained to those regions. This assumption will increase likelihoods of a CNV hitting a gene but given that the increase will be the same between cases and controls will have limited effect on my results. A possible solution to this would be to actually restrict the variants from all sources to the minimum exome size for consistency, in my case I would redefine deletion length to only include the coding sequence represented on the hybrid selection array in the exome sequencing. It's also possible that additional information could be garnered from the number of genes affected by the variant or amount of a gene affected by variant which while correlated to overall length is not completely dependent. A large CNV hitting a single large gene could in that case be less functionally relevant than a smaller CNV hitting multiple small genes. Such information could be ultimately

incorporated in my model along with other possible information such as sequence context or ancestry. As more rare variant data becomes available methods such as ours will become more powered to identify particular genes and will aid in the fundamental understanding of the underlying biology for diseases like SCZ.

Chapter 6. Discussion

6.1 Motivation

In this thesis I looked to better understand the underlying biology of schizophrenia by assessing the role of many classes of variation to disease risk. I aimed to both increase power to identify biological mechanisms by testing sets of genes with similar functions for disease association and by aggregating different forms of variation to implicate specific genes. I utilized multiple technologies to survey different classes of variation from genotyping arrays to assess common SNPs and rare CNVs to exome sequencing to test disruptive heterozygous, recessive and compound heterozygous mutations. I additionally aimed to develop methods that allowed for the testing, in aggregate, of these variants to identify genesets or single genes associated to schizophrenia. It is with these approaches, along with increasing sample sizes and a more fundamental incorporation of other biological data such as expression, histone modification, etc. that we will ultimately identify the causes of this terrible illness. While sample sizes have begun to reach the levels of other disorders, such as diabetes, just recently in GWAS, they remain low for rare variants and approaches to improve power are necessary to identify signal scattered across the genome and interspersed with lots of noise.

6.2 Results

The first three research chapters of this thesis presented both original results and original methods with a particular variant class in focus within each chapter. The

final chapter presented a novel method to incorporate the different classes of variation described in detail in the previous chapters. The overarching goal was to understand the underlying biology of schizophrenia by identifying genes associated to disease across all variation classes. I first presented a large CNV dataset of SCZ and matched controls, I reaffirmed increased rates of CNVs and genes hit by CNVs in SCZ cases compared to controls. I went further by specifically assessing the role played by smaller CNVs (<20kb) that are still accurately called from these genotyping arrays. I concluded that this class of CNVs, while making up ~60% of all CNVs in the dataset are not contributing to SCZ risk as strongly as the larger CNVs. I identified a number of genes with nominally significant association worthy of follow up and began to ascertain functionally similar genes that are more significantly enriched for case CNVs than the genome over all. The gene sets tested represented those that had previously been implicated in psychiatric disease up until that point. This included sets of brain expressed genes and those related to learning, synaptic function and neuronal activity. These results confirmed the role CNVs play in risk of SCZ and also demonstrated that these CNVs were not randomly distributed but were disproportionately occurring in particular sets of genes with related function. Further follow up of these results and refinement of these gene sets could reveal more of the underlying biology of SCZ.

I next looked to develop a method specifically capitalizing on the polygenic nature of SCZ. I hoped to take the genome-wide polygenic signal identified previously and implicate subsets of genes carrying more true disease loci. I discovered significant issues relating to variant inclusion and correlation across

variants in the same region. Despite efforts to improve the method by using approaches designed to deal with dependence of variables I still identified significant correlation between numbers of variants within a test set and overall significance. With caveats aside I did identify a possible significant overlap of common SNP variants and genes hit by *de novo* CNVs, however neither this result nor any other was replicated in an independent dataset.

Exome sequencing has allowed for an extensive view of variation in the coding regions of the genome, the parts that I can currently best interpret. For the first time, I can survey very rare, deleterious variation across many individuals. Given the conclusive role of CNVs in SCZ, it seems likely that rare deleterious SNVs will also contribute to disease risk. A particular class of variation that historically has been strongly implicated in disease is the two-hit model where both chromosomes have a mutation, either at the same genomic position (recessive) or different positions (compound heterozygous). I have developed a method to identify these mutations and test them for association to disease. I identified a very small number of disruptive two-hit mutations across a set of over 5,000 people averaging out to about a tenth of an event per person, considerably rarer than *de novo* SNVs/indels. I identified a number of significant variants but none passing multiple test correction. In general, I saw no increase in two-hit disruptive events in cases compared to controls pointing to a limited role for these mutations in SCZ disease risk. This does not preclude mutations in specific genes playing a role, I identified a significant increase of these events in the same genes that contain disruptive *de novo* mutations

in SCZ probands. These genes represent good follow up candidates for further study in SCZ.

Finally, I aggregated all of the knowledge and data from the first three chapters into trying to identify specific genes for further study. I designed and implemented an analytical framework to test variants of any type or length to test genes for association. I showed a modest overlap between case specific rare deletions and case specific singleton disruptive mutations and identified a set of 76 genes showing nominal significance. I also pointed to a particularly significant gene (MS4A10) with prior disease implications in Alzheimer's disease. Most importantly I demonstrated a method for prioritizing genes underlying large associated CNV regions and specifically identified a set of four genes out of eleven in the 3q29 deletion region that are more significantly associated than the others. I believe with more data this will be a valuable approach to identifying disease genes and furthering the biological understanding of the schizophrenia.

6.3 Interpretation

The goal of this thesis was to contribute to the biological understanding of schizophrenia by utilizing genetic data. Here, I contributed to that goal by determining the role of multiple classes of variation not previously tested. I showed that CNVs on the shorter end of the spectrum (20kb – 100kb) but that are still capable of being accurately called by genotyping platforms were not enriched for case events similar to the larger CNVs. This does not preclude the contribution of

particular small CNVs in certain regions/genes to SCZ but that those may be hard to implicate amongst the many non-disease related shorter CNVs since as events get smaller there is an exponentially increasing number of them. Additionally, I identified a class of rare, likely deleterious recessive and compound heterozygous events in a large exome sequencing study (5,000 individuals). I identified no significant increase of these events in cases nor did I find any significant single variant after multiple test correction. These events are extremely rare and occur nearly 10 times less often than *de novo* mutations so power will be limited but this result is in direct contrast to what has been seen in autism, where a 50% increase in case two-hit events has been identified. There are numerous models for why this might be the case, recessive diseases are often early onset and severe and given the significant overlap between psychiatric diseases both phenotypically and genetically such a deleterious mutation might present with more dramatic symptoms. The true model is as yet unclear but I can conclude that these events are not a major contributor of SCZ risk in the general population.

I contribute several methods and knowledge regarding how one might approach identifying variants contributing to disease. I put great effort into taking the genome-wide polygenic signal previously identified in SCZ and attempting to refine this signal to a subset of genes or a gene set that is contributing more than the base level. The identification of this polygenic signal was a significant advancement in understanding the genetic underpinnings of this disorder but is limited in terms of its ability to inform the biology. We, as many do, believe that this signal is not a randomly distributed set of risk variants but are concentrated in certain biological

functions or pathways that when perturbed lead to altered disease risk. By testing certain pathways for enrichment of polygenic signal above baseline I hoped to contribute knowledge regarding SCZ disease biology. Unfortunately, I was confronted with multiple challenges in this approach that are a product of both the genome itself and the technology used to assay genetic markers. My challenges will provide insight into how to better this approach or convince others to rethink this strategy and utilize an alternative method or technology. I provide a method to identify and test recessive and compound heterozygous mutations that is accessible and easily run within the Plink/SEQ software package. I further describe a method to aggregate variants of many varieties and variable lengths to test specific genes for association. This method becomes immediately useful to any researcher with multiple class of rare variation. As sequencing becomes more widely used, this approach will be appropriate for more and more studies.

While I've contributed to knowledge of the genetic model underlying SCZ and have provided several methods of use to researchers in SCZ and other fields my main focus was on understanding the biology of this disorder. In that vein, I've approached each of the first three chapters with a focus on testing biological pathways. As time has passed and the chapters move forward the specific pathways I tested evolved to reflect the most current literature and my findings. I started in the first chapter testing pathways that had been previously implicated, mostly by CNVs, in the literature but that also represented brain related functions relevant for SCZ. These pathways ranged in specificity and size from the very large and broad set of brain expressed genes to the more specific sets of genes related to synaptic processes and

behavior/learning. I identified significant enrichment of case CNVs in many of these categories in particular the synaptic set. Research being done around the same time had specified a more complete and more specific set of genes related to synaptic function by isolating the postsynaptic density (PSD) from human neocortex and performing proteomic profiling. Subunits of the PSD complex were identified to be disproportionately affected by *de novo* CNVs. I included both the enriched subunits and the *de novo* CNVs into my polygenic pathway method hoping to understand the relationship between common SNPs and particular post synaptic density complexes. I identified a significant enrichment of polygenic signal in regions defined by *de novo* CNVs but not in the two PSD subunits implicated by those CNVs, possibly representing some evidence of overlap between common variation and *de novo* variation. Unfortunately, I was presented with difficulties in accurately assessing this overlap and was unable to replicate these results. I continued however along these lines of research into the very rare category of two-hit, disruptive mutations. Here, I again tested overlap with the PSD sets and a larger set of *de novo* points mutations. I identified a nominally significant enrichment of case two-hit mutations in genes carrying *de novo* SCZ mutations but no enrichment in the PSD sets. While definitively underpowered I started to see overlap between multiple classes of variation. Given this modest overlap and the ever increasing amount of SCZ genomic data I hoped to move beyond identifying pathways and or regions and implicate specific genes. While knowledge about the genetics underlying SCZ has advanced dramatically over the last several years, it is still extremely rare to have such a specific finding as to point to a single gene. Deletions in *NRXN1* and duplications in *VIPR2* are two of the

more specific results that have been presented, most other CNVs and GWAS results have implicated large regions of several to many genes. Pathway analysis has helped us identify biological functions related to disease but by definition implicate many genes. An identification of a single gene could provide an ideal candidate for functional work and possible therapeutic action. In combining rare deletions and singleton disruptive mutations I can start to pull out significant genes with no prior SCZ evidence such as *MS4A10* or begin to distill which genes among implicated regions are more worthwhile to follow up such as *DLG1* and *RNF168*. Each of these genes has prior evidence for involvement in psychiatric disease and with more data, we are likely to identify a gene with appropriate significance to begin the necessary follow up work to understand the biological implications of that gene's function on schizophrenia which has been the goal of this work.

6.4 Limitations and future research

I present my research with knowledge of limitations and areas for further study beyond the scope and/or timeframe of this degree. I presented results showing that CNVs between 20kb and 100kb are not contributing to the burden of CNVs in cases but this does not specifically ask how contribution is correlated with size. A full quantitative evaluation testing burden across the full range of CNV lengths would help answer this question more specifically. Additionally, the 1% frequency cutoff is somewhat arbitrary and a full evaluation might elucidate where the disease CNVs fall along the frequency distribution. A question related to CNV size that is not yet

understood is to what degree hitting multiple genes is necessary to be a disease-causing event. Given the limited evidence for strong effect CNVs hitting single genes it raises the question of whether even in CNV space disease risk is polygenic and only when a combination of genes is hit does risk increase. I also recognize that knowledge about the biology of SCZ has progressed over the years of research in this thesis and it would be appropriate to test the newest most relevant pathways for CNV enrichment.

I present my polygenic pathway method complete with the particular difficulties I faced regarding bias with number of SNPs and improperly handling LD between SNPs. There are several points along the process where different approaches could be taken and the results explored. I chose to include all SNPs to maximize information and with the approach that I would handle LD downstream. An alternative approach would be to employ a more careful SNP selection strategy, either by selectively pruning SNPs in LD to retain as many in genes as possible or by weighting SNPs based on LD with the most significant SNP in the region. The latter would be more difficult than the former. I also selected ridge regression as my approach to handle correlation across markers but there exists a wide literature and numerous alternative approaches to perform a similar task. Perhaps a different selection would work better in these conditions.

As is often the case, results from analyses present opportunities to perform further analyses. During the course of this research there were places where further research would be clearly of interest given unlimited time. One particular type of analysis that was beyond the scope of this thesis would be a thorough both inclusion

and comparison of results from SCZ and results from other disorders such as bipolar or autism. In particular, while bipolar and schizophrenia have shown a strong overlap of genetic risk loci there does not appear to be any CNV burden in bipolar cases compared to controls. A full comparison of CNV results as shown in chapter 2 with similar results from bipolar would be of particular interest. As mentioned previously, a 50% increase in two-hit disruptive mutations exists in autism cases compared to controls. Comparing the variants/genes hit across SCZ and autism could point to smaller set of genes related to psychiatric disease. In general, I could combine data from multiple diseases and ask more general questions about the biology of psychiatric disease. By doing so I'd also increase my ability to detect genes and variants but could be reducing the specificity of my results to SCZ.

Clearly power will play a large role in my ability to identify specific associated variants or genes. I present approaches to increase power either by aggregating variants across genes of a similar pathway or across multiple classes of variation. Including additional data will improve power, this could be by including more samples for all analyses or by increasing the number of variants. I specifically sought to test variants that were likely to be deleterious and restricted my analyses to those that were disruptive but the inclusion of predicted deleterious non-synonymous variants would increase my set of variants to test in both the two-hit model and the aggregation method. Further, I could include both duplications and *de novo* variants in my aggregation method to increase the likelihood of seeing more variants within a single gene. Finally, as I did in the first three chapters I could aggregate genes into pathways to test for specific biology associated in SCZ.

Appendix A: Complete Tables

Appendix A1. Full GO ridge regression results with permutation and replication p-values from independent Swedish sample

GO SET	#GENES	#SNPS	Ridge Stat	Ridge P	Perm P	Sweden Stat	Sweden P
SIGNAL_TRANSDUCTION	1634	78444	8.23	**1.91E-16**	0	-0.58	0.56
CELL_SURFACE_RECEPTOR_LINKED_SIGNAL_TRANSDUCTION_GO_0007166	641	34865	6.59	**4.36E-11**	0.01	0.82	0.42
SYSTEM_PROCESS	562	34478	6.25	**4.01E-10**	0.01	0.95	0.34
NEUROLOGICAL_SYSTEM_PROCESS	379	26955	6.16	**7.35E-10**	0	0.13	0.89
SYSTEM_DEVELOPMENT	856	55028	6.00	**2.02E-09**	0	0.43	0.67
ANATOMICAL_STRUCTURE_DEVELOPMENT	1008	61013	5.82	**5.88E-09**	0	0.83	0.41
SENSORY_PERCEPTION	190	9885	5.69	**1.27E-08**	0	-1.33	0.19
MULTICELLULAR_ORGANISMAL_DEVELOPMENT	1042	62415	5.62	**1.93E-08**	0.01	0.34	0.73
INTRACELLULAR_SIGNALING_CASCADE	667	28606	5.59	**2.31E-08**	0	0.67	0.50
G_PROTEIN_COUPLED_RECEPTOR_ACTIVITY	189	10076	5.59	**2.33E-08**	0	0.58	0.56
REGULATION_OF_BIOLOGICAL_QUALITY	418	19025	5.47	**4.46E-08**	0.01	0.82	0.41
BIOPOLYMER_MODIFICATION	649	36492	5.36	**8.10E-08**	0	-0.09	0.93
PROTEIN_MODIFICATION_PROCESS	630	36085	5.30	**1.17E-07**	0	-0.06	0.96
CENTRAL_NERVOUS_SYSTEM_DEVELOPMENT	122	13511	5.22	**1.75E-07**	0.01	0.70	0.48
PHOTORECEPTOR_CELL_MAINTENANCE	10	3008	5.14	**2.72E-07**	0.01	-0.32	0.75
CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS	209	12333	4.94	**7.77E-07**	0	-0.02	0.98
ACTIN_FILAMENT_BUNDLE_FORMATION	13	653	4.92	**8.70E-07**	0.04	-1.32	0.19
NERVOUS_SYSTEM_DEVELOPMENT	382	31283	4.87	**1.11E-06**	0.01	0.04	0.97
RHODOPSIN_LIKE_RECEPTOR_ACTIVITY	134	3878	4.76	**1.92E-06**	0.01	0.67	0.50
CELLULAR_MACROMOLECULE_METABOLIC_PROCESS	1128	50410	4.76	**1.93E-06**	0.02	0.46	0.65
CELL_CELL_SIGNALING	405	23543	4.75	**2.02E-06**	0	0.94	0.35
TRANSMEMBRANE_RECEPTOR_PROTEIN_TYROSINE_KINASE_SIGNALING_PATHWAY	83	8809	4.70	**2.60E-06**	0.03	0.86	0.39
CELLULAR_PROTEIN_METABOLIC_PROCESS	1114	50004	4.69	**2.69E-06**	0.04	0.44	0.66
ORGAN_DEVELOPMENT	568	29985	4.64	**3.46E-06**	0.04	1.56	0.12
ENZYME_LINKED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY	140	13845	4.55	**5.49E-06**	0.02	0.34	0.74
ACTIN_CYTOSKELETON	129	7916	4.54	**5.67E-06**	0.01	-1.59	0.11
HEMATOPOIETIN_INTERFERON_CLASS__D200_DOMAIN__CYTOKINE_RECEPTOR_ACTIVITY	34	1080	4.49	**7.08E-06**	0.07	0.55	0.58
PROTEIN_METABOLIC_PROCESS	1228	55652	4.48	**7.33E-06**	0.01	0.66	0.51
FEEDING_BEHAVIOR	24	647	4.43	**9.40E-06**	0.07	0.24	0.81
MEMBRANE	1985	110848	4.39	**1.15E-05**	0.02	-0.37	0.71
SKELETAL_DEVELOPMENT	101	5396	4.38	**1.17E-05**	0.06	1.05	0.29
PHOSPHORIC_DIESTER_HYDROLASE_ACTIVITY	39	3965	4.34	**1.42E-05**	0.01	-0.80	0.43
COLLAGEN_BINDING	13	1595	4.31	**1.66E-05**	0.06	0.62	0.54
SMALL_GTPASE_MEDIATED_SIGNAL_TRANSDUCTION	90	6224	4.28	**1.86E-05**	0.04	1.36	0.17
G_PROTEIN_COUPLED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY	342	12885	4.26	**2.01E-05**	0.01	1.63	0.10
SYNAPSE	27	6069	4.26	**2.08E-05**	0.04	0.62	0.53
PLASMA_MEMBRANE	1422	88107	4.25	**2.14E-05**	0.03	0.52	0.61
INTERCELLULAR_JUNCTION_ASSEMBLY	11	290	4.23	**2.38E-05**	0.06	1.72	0.08
PHOSPHORIC_ESTER_HYDROLASE_ACTIVITY	151	14140	4.15	**3.32E-05**	0.04	-0.76	0.45
POST_TRANSLATIONAL_PROTEIN_MODIFICATION	475	29297	4.13	**3.64E-05**	0.04	-0.20	0.84
VOLTAGE_GATED_CATION_CHANNEL_ACTIVITY	66	5716	4.10	**4.16E-05**	0.03	1.06	0.29
BEHAVIOR	155	5810	4.09	**4.23E-05**	0.04	1.73	0.08
MEMBRANE_ORGANIZATION_AND_BIOGENESIS	135	11178	4.06	**4.97E-05**	0.04	1.69	0.09
CELL_SUBSTRATE_ADHESION	39	4470	4.02	**5.94E-05**	0.07	1.47	0.14
ACTIN_FILAMENT_ORGANIZATION	24	909	4.00	**6.27E-05**	0.04	-0.43	0.66
PROTEIN_AMINO_ACID_PHOSPHORYLATION	278	15958	3.98	**6.92E-05**	0.07	0.26	0.79
VOLTAGE_GATED_CHANNEL_ACTIVITY	73	5800	3.98	**6.94E-05**	0.07	0.92	0.36
BRAIN_DEVELOPMENT	51	4685	3.97	**7.14E-05**	0.04	2.09	0.04
REGULATION_OF_TRANSLATIONAL_INITIATION	31	658	3.95	**7.95E-05**	0.04	-1.19	0.23
PLASMA_MEMBRANE_PART	1155	71117	3.93	**8.62E-05**	0.04	0.96	0.34
POSITIVE_REGULATION_OF_TRANSCRIPTION_FROM_RNA_POLYMERASE_II_PROMOTER	65	4057	3.90	**9.68E-05**	0.08	0.78	0.44
NEUROPEPTIDE_RECEPTOR_ACTIVITY	21	374	3.90	**9.73E-05**	0.07	1.77	0.08
CELL_MATRIX_ADHESION	38	4324	3.88	**1.04E-04**	0.05	1.00	0.32
CELL_DEVELOPMENT	577	30987	3.87	**1.09E-04**	0.03	-0.18	0.86
SECOND_MESSENGER_MEDIATED_SIGNALING	152	9622	3.84	**1.23E-04**	0.08	2.33	0.02
NEUROPEPTIDE_BINDING	22	394	3.83	**1.29E-04**	0.09	1.95	0.05
GENERATION_OF_PRECURSOR_METABOLITES_AND_ENERGY	122	4454	3.81	**1.40E-04**	0.05	1.63	0.10
RECEPTOR_ACTIVITY	579	40165	3.77	**1.63E-04**	0.03	-0.85	0.40
TRANSFERASE_ACTIVITY__TRANSFERRING_GLYCOSYL_GROUPS	111	5084	3.76	**1.72E-04**	0.04	0.56	0.58
SYNAPTIC_TRANSMISSION	174	16397	3.75	**1.74E-04**	0.05	0.84	0.40
NEURON_DIFFERENTIATION	76	10229	3.75	**1.80E-04**	0.05	-1.55	0.12
CYTOKINE_BINDING	49	748	3.73	**1.90E-04**	0.07	0.36	0.72
RAS_PROTEIN_SIGNAL_TRANSDUCTION	67	4701	3.71	**2.06E-04**	0.05	0.70	0.48
INTERPHASE_OF_MITOTIC_CELL_CYCLE	62	2721	3.67	**2.39E-04**	0.08	1.08	0.28
PHAGOCYTOSIS	17	1419	3.67	**2.41E-04**	0.07	1.30	0.19
REGULATION_OF_CELLULAR_COMPONENT_ORGANIZATION_AND_BIOGENESIS	125	7656	3.66	**2.56E-04**	0.06	0.72	0.47
LEADING_EDGE	48	2748	3.62	**2.98E-04**	0.07	-0.73	0.47
REGULATION_OF_SMALL_GTPASE_MEDIATED_SIGNAL_TRANSDUCTION	24	1640	3.57	**3.53E-04**	0.08	-1.29	0.20
TRANSMISSION_OF_NERVE_IMPULSE	189	17060	3.57	**3.61E-04**	0.05	0.97	0.33
ACTIN_CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS	106	7480	3.56	**3.69E-04**	0.08	0.30	0.76
INTERCELLULAR_JUNCTION_ASSEMBLY_AND_MAINTENANCE	13	319	3.56	**3.71E-04**	0.14	1.94	0.05
CYTOPLASM_ORGANIZATION_AND_BIOGENESIS	15	593	3.55	**3.90E-04**	0.09	1.57	0.12
ACTIN_FILAMENT	18	1413	3.54	**4.05E-04**	0.13	-1.70	0.09
PEPTIDE_RECEPTOR_ACTIVITY	53	1667	3.53	**4.15E-04**	0.04	0.55	0.58
MICROTUBULE_BASED_MOVEMENT	16	986	3.52	**4.28E-04**	0.08	-1.09	0.28
PERINUCLEAR_REGION_OF_CYTOPLASM	54	4458	3.52	**4.28E-04**	0.08	-0.44	0.66
CYCLIC_NUCLEOTIDE_PHOSPHODIESTERASE_ACTIVITY	14	2837	3.51	**4.42E-04**	0.07	-0.49	0.62
GATED_CHANNEL_ACTIVITY	122	10087	3.51	**4.45E-04**	0.1	0.71	0.48
ELECTRON_TRANSPORT_GO_0006118	51	2595	3.50	**4.57E-04**	0.05	0.60	0.55
3__5__CYCLIC_NUCLEOTIDE_PHOSPHODIESTERASE_ACTIVITY	13	2834	3.50	**4.68E-04**	0.06	-0.48	0.63
CYTOSKELETAL_PART	235	9737	3.50	**4.69E-04**	0.03	-2.07	0.04
LAMELLIPODIUM	26	1471	3.50	**4.69E-04**	0.07	-1.01	0.31
BASEMENT_MEMBRANE	37	4868	3.49	**4.81E-04**	0.02	-1.26	0.21
TRANSFERASE_ACTIVITY__TRANSFERRING_HEXOSYL_GROUPS	80	3470	3.48	**4.92E-04**	0.05	-0.08	0.93
REGULATION_OF_HYDROLASE_ACTIVITY	80	2431	3.48	**5.01E-04**	0.08	0.43	0.67
MICROTUBULE_BASED_PROCESS	82	4378	3.45	**5.56E-04**	0.07	-0.66	0.51
FOCAL_ADHESION_FORMATION	10	607	3.45	**5.65E-04**	0.05	-0.76	0.45
NEGATIVE_REGULATION_OF_CELLULAR_PROCESS	643	27214	3.43	**6.05E-04**	0.07	-0.10	0.92
PHOSPHORYLATION	312	16468	3.42	**6.31E-04**	0.06	0.25	0.80
REGULATION_OF_TRANSLATION	93	1679	3.42	**6.35E-04**	0.11	0.29	0.77
VESICLE_MEDIATED_TRANSPORT	194	13832	3.41	**6.41E-04**	0.07	-0.02	0.99
ACTIN_FILAMENT_BASED_PROCESS	116	7983	3.38	**7.17E-04**	0.11	0.33	0.74
CYTOSKELETAL_PROTEIN_BINDING	159	10669	3.38	**7.22E-04**	0.08	-2.30	0.02
GENERATION_OF_NEURONS	83	10519	3.38	**7.24E-04**	0.09	-1.80	0.07
NEURON_DEVELOPMENT	61	7425	3.38	**7.36E-04**	0.07	-0.82	0.41
MEMBRANE_PART	1663	90983	3.38	**7.36E-04**	0.1	-0.06	0.96
COFACTOR_BINDING	22	2295	3.37	**7.62E-04**	0.1	0.72	0.47
INOSITOL_OR_PHOSPHATIDYLINOSITOL_PHOSPHATASE_ACTIVITY	12	973	3.35	**8.07E-04**	0.14	0.01	0.99
REGULATION_OF_PHOSPHORYLATION	49	1250	3.34	**8.44E-04**	0.1	0.71	0.48
VOLTAGE_GATED_POTASSIUM_CHANNEL_ACTIVITY	36	2847	3.30	**9.79E-04**	0.06	1.69	0.09
RESPONSE_TO_ORGANIC_SUBSTANCE	30	3159	3.30	**9.82E-04**	0.11	1.64	0.10
NEGATIVE_REGULATION_OF_GROWTH	40	1540	3.28	**1.05E-03**	0.07	-0.14	0.89
EUKARYOTIC_TRANSLATION_INITIATION_FACTOR_3_COMPLEX	10	202	3.27	**1.09E-03**	0.13	-1.43	0.15
CATION_CHANNEL_ACTIVITY	119	10143	3.25	**1.16E-03**	0.05	0.71	0.48
BONE_REMODELING	28	1169	3.23	**1.26E-03**	0.14	2.06	0.04

TISSUE_REMODELING	29	1169	3.23	**1.26E-03**	0.2	2.06	0.04
NEGATIVE_REGULATION_OF_PHOSPHORYLATION	12	92	3.22	**1.29E-03**	0.12	0.88	0.38
NUCLEOBASE__NUCLEOSIDE_AND_NUCLEOTIDE_METABOLIC_PROCESS	51	4339	3.22	**1.29E-03**	0.05	0.50	0.62
LEARNING_AND_OR_MEMORY	14	710	3.22	**1.29E-03**	0.12	0.67	0.50
PROTEIN_TYROSINE_PHOSPHATASE_ACTIVITY	53	8209	3.21	**1.33E-03**	0.1	-0.43	0.66
NEGATIVE_REGULATION_OF_PHOSPHATE_METABOLIC_PROCESS	13	139	3.19	**1.45E-03**	0.07	0.91	0.36
REGULATION_OF_PROTEIN_METABOLIC_PROCESS	173	4195	3.18	**1.48E-03**	0.09	-0.06	0.95
REGULATION_OF_RAS_GTPASE_ACTIVITY	12	743	3.17	**1.50E-03**	0.1	-1.30	0.19
REGULATION_OF_DEVELOPMENTAL_PROCESS	440	21405	3.17	**1.50E-03**	0.13	1.08	0.28
GLYCOPROTEIN_METABOLIC_PROCESS	90	4576	3.16	**1.56E-03**	0.09	-0.65	0.51
S_PHASE_OF_MITOTIC_CELL_CYCLE	10	1137	3.16	**1.59E-03**	0.14	-1.14	0.26
REGULATION_OF_CELL_DIFFERENTIATION	59	4840	3.16	**1.60E-03**	0.15	1.34	0.18
G_PROTEIN_SIGNALING__COUPLED_TO_CYCLIC_NUCLEOTIDE_SECOND_MESSENGER	99	5027	3.12	**1.82E-03**	0.14	0.44	0.66
REGULATION_OF_RAS_PROTEIN_SIGNAL_TRANSDUCTION	20	1099	3.12	**1.83E-03**	0.11	-1.21	0.23
REGULATION_OF_GTPASE_ACTIVITY	16	838	3.11	**1.85E-03**	0.11	-1.13	0.26
INTEGRIN_COMPLEX	19	1819	3.11	**1.89E-03**	0.09	0.99	0.32
NEGATIVE_REGULATION_OF_CELL_PROLIFERATION	156	7218	3.10	**1.91E-03**	0.11	-0.11	0.91
PROTEIN_HOMODIMERIZATION_ACTIVITY	121	4852	3.10	**1.95E-03**	0.13	1.05	0.29
HYDROLASE_ACTIVITY__ACTING_ON_ESTER_BONDS	266	16714	3.10	**1.95E-03**	0.13	-0.54	0.59
REGULATION_OF_NEUROTRANSMITTER_LEVELS	24	2384	3.09	**1.97E-03**	0.13	0.53	0.60
HEMOSTASIS	47	2173	3.09	**1.98E-03**	0.1	0.34	0.74
ORGANELLE_ORGANIZATION_AND_BIOGENESIS	474	19549	3.09	**2.00E-03**	0.15	0.20	0.84
GROWTH	77	3466	3.08	**2.04E-03**	0.11	1.13	0.26
WOUND_HEALING	53	2254	3.08	**2.05E-03**	0.12	0.91	0.36
REGULATION_OF_RHO_GTPASE_ACTIVITY	11	640	3.07	**2.17E-03**	0.1	-1.06	0.29
SUBSTRATE_SPECIFIC_CHANNEL_ACTIVITY	156	10889	3.07	**2.17E-03**	0.1	0.29	0.77
NEGATIVE_REGULATION_OF_BIOLOGICAL_PROCESS	674	28517	3.06	**2.19E-03**	0.11	-0.24	0.81
TRANSMEMBRANE_RECEPTOR_PROTEIN_PHOSPHATASE_ACTIVITY	19	7030	3.06	**2.19E-03**	0.1	-0.36	0.72
CYTOSKELETON_DEPENDENT_INTRACELLULAR_TRANSPORT	26	1571	3.06	**2.21E-03**	0.08	-0.73	0.46
CELL_RECOGNITION	18	3494	3.05	**2.27E-03**	0.07	0.05	0.96
GLYCOPROTEIN_BIOSYNTHETIC_PROCESS	74	4172	3.04	**2.37E-03**	0.05	-0.75	0.45
METAL_ION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	147	12220	3.03	**2.41E-03**	0.14	-0.17	0.87
CYCLIC_NUCLEOTIDE_MEDIATED_SIGNALING	101	5142	3.03	**2.42E-03**	0.09	0.49	0.63
DETECTION_OF_ABIOTIC_STIMULUS	19	468	3.03	**2.47E-03**	0.16	-0.36	0.72
REGULATION_OF_METABOLIC_PROCESS	796	29671	3.02	**2.54E-03**	0.11	1.06	0.29
REGULATION_OF_CELLULAR_METABOLIC_PROCESS	784	29444	3.01	**2.57E-03**	0.13	1.25	0.21
CALMODULIN_BINDING	25	2178	3.01	**2.59E-03**	0.07	-2.53	0.01
ION_CHANNEL_ACTIVITY	149	10775	3.01	**2.61E-03**	0.09	0.36	0.72
NEUROGENESIS	93	10792	3.00	**2.66E-03**	0.14	-1.75	0.08
NUCLEOTIDE_METABOLIC_PROCESS	41	4221	3.00	**2.73E-03**	0.09	0.38	0.71
NEGATIVE_REGULATION_OF_MYELOID_CELL_DIFFERENTIATION	10	142	2.97	**2.93E-03**	0.11	1.65	0.10
NEURITE_DEVELOPMENT	53	6820	2.96	**3.12E-03**	0.13	-0.83	0.41
INTERPHASE	68	3119	2.95	**3.13E-03**	0.17	0.90	0.37
POTASSIUM_CHANNEL_ACTIVITY	50	3674	2.95	**3.16E-03**	0.14	2.31	0.02
GLUTAMATE_RECEPTOR_ACTIVITY	20	6357	2.95	**3.20E-03**	0.16	0.01	1.00
REGULATION_OF_NUCLEOCYTOPLASMIC_TRANSPORT	22	786	2.94	**3.32E-03**	0.11	-0.89	0.37
KERATINOCYTE_DIFFERENTIATION	15	326	2.93	**3.37E-03**	0.08	1.32	0.19
REGULATION_OF_CELL_GROWTH	46	1866	2.93	**3.39E-03**	0.13	0.16	0.88
TRANSCRIPTION_FACTOR_BINDING	307	11501	2.93	**3.42E-03**	0.15	-0.22	0.83
TRANSLATIONAL_INITIATION	39	781	2.93	**3.44E-03**	0.11	-1.62	0.10
INSOLUBLE_FRACTION	15	1769	2.92	**3.51E-03**	0.16	-1.74	0.08
REGULATION_OF_CELLULAR_PROTEIN_METABOLIC_PROCESS	162	3940	2.92	**3.53E-03**	0.18	0.44	0.66
G_PROTEIN_SIGNALING__ADENYLATE_CYCLASE_INHIBITING_PATHWAY	10	292	2.90	**3.69E-03**	0.13	-0.05	0.96
GROWTH_FACTOR_ACTIVITY	55	1835	2.90	**3.70E-03**	0.12	1.43	0.15
CARBOHYDRATE_TRANSPORT	19	607	2.90	**3.76E-03**	0.14	0.20	0.84
SPHINGOLIPID_BIOSYNTHETIC_PROCESS	10	1145	2.89	**3.91E-03**	0.17	-0.15	0.88
INSULIN_RECEPTOR_SIGNALING_PATHWAY	19	1214	2.87	**4.17E-03**	0.14	0.71	0.48
POSITIVE_REGULATION_OF_CELL_ADHESION	13	2354	2.86	**4.22E-03**	0.2	2.30	0.02
ANCHORED_TO_PLASMA_MEMBRANE	13	1193	2.86	**4.29E-03**	0.08	0.03	0.98
ANCHORED_TO_MEMBRANE	13	1193	2.86	**4.29E-03**	0.2	0.03	0.98
VOLTAGE_GATED_CALCIUM_CHANNEL_COMPLEX	15	2244	2.85	**4.31E-03**	0.12	0.60	0.55
TELOMERIC_DNA_BINDING	10	308	2.85	**4.32E-03**	0.14	0.41	0.68
G_PROTEIN_SIGNALING__COUPLED_TO_CAMP_NUCLEOTIDE_SECOND_MESSENGER	63	4159	2.85	**4.34E-03**	0.14	0.22	0.83
REGULATION_OF_TRANSPORT	67	4185	2.85	**4.37E-03**	0.18	1.22	0.22
INTRINSIC_TO_PLASMA_MEMBRANE	988	56568	2.85	**4.39E-03**	0.16	0.59	0.55
ENZYME_REGULATOR_ACTIVITY	323	16827	2.85	**4.39E-03**	0.13	0.39	0.70
COAGULATION	43	1982	2.84	**4.54E-03**	0.17	0.85	0.39
OXIDOREDUCTASE_ACTIVITY__ACTING_ON_SULFUR_GROUP_OF_DONORS	10	899	2.84	**4.57E-03**	0.17	-0.54	0.59
REGULATION_OF_RHO_PROTEIN_SIGNAL_TRANSDUCTION	15	698	2.82	**4.77E-03**	0.13	-1.25	0.21
VOLTAGE_GATED_CALCIUM_CHANNEL_ACTIVITY	18	2383	2.82	**4.79E-03**	0.15	0.45	0.65
SEQUENCE_SPECIFIC_DNA_BINDING	58	1735	2.82	**4.79E-03**	0.1	-0.08	0.94
REGULATION_OF_INTRACELLULAR_TRANSPORT	25	788	2.81	**4.95E-03**	0.15	-0.88	0.38
POSITIVE_REGULATION_OF_HYDROLASE_ACTIVITY	55	1528	2.80	**5.12E-03**	0.16	0.72	0.47
NON_MEMBRANE_BOUND_ORGANELLE	629	24962	2.80	**5.13E-03**	0.16	-0.57	0.57
INTRACELLULAR_NON_MEMBRANE_BOUND_ORGANELLE	629	24962	2.80	**5.13E-03**	0.21	-0.57	0.57
CAMP_MEDIATED_SIGNALING	64	4269	2.78	**5.36E-03**	0.14	0.25	0.80
BLOOD_COAGULATION	42	1975	2.78	**5.46E-03**	0.19	0.84	0.40
RHO_PROTEIN_SIGNAL_TRANSDUCTION	39	3089	2.78	**5.48E-03**	0.14	1.27	0.21
GTPASE_REGULATOR_ACTIVITY	126	11193	2.75	**5.92E-03**	0.2	0.63	0.53
AXONOGENESIS	43	6661	2.75	**5.96E-03**	0.18	-0.80	0.42
HYDROLASE_ACTIVITY__ACTING_ON_CARBON_NITROGEN__BUT_NOT_PEPTIDE__BONDS__IN_CYCLIC_AMIDINES	16	842	2.74	**6.08E-03**	0.2	-1.73	0.08
METABOTROPIC_GLUTAMATE__GABA_B_LIKE_RECEPTOR_ACTIVITY	10	3386	2.74	**6.19E-03**	0.18	0.93	0.35
NEUROPEPTIDE_SIGNALING_PATHWAY	14	257	2.73	**6.40E-03**	0.19	-0.33	0.74
CELL_PROJECTION_BIOGENESIS	26	3204	2.72	**6.62E-03**	0.18	1.65	0.10
CELL_CYCLE_ARREST_GO_0007050	57	1861	2.71	**6.64E-03**	0.25	-1.32	0.19
MITOTIC_CELL_CYCLE	152	5299	2.71	**6.72E-03**	0.17	0.87	0.38
COENZYME_BINDING	16	1858	2.71	**6.73E-03**	0.2	-0.24	0.81
TRANSCRIPTION_ACTIVATOR_ACTIVITY	172	6804	2.71	**6.74E-03**	0.14	0.75	0.45
CELLULAR_MORPHOGENESIS_DURING_DIFFERENTIATION	49	6948	2.70	**6.99E-03**	0.16	-1.06	0.29
RNA_POLYMERASE_II_TRANSCRIPTION_FACTOR_ACTIVITY__ENHANCER_BINDING	14	607	2.69	**7.22E-03**	0.13	0.10	0.92
POSITIVE_REGULATION_OF_DEVELOPMENTAL_PROCESS	218	9997	2.66	**7.91E-03**	0.09	-0.38	0.71
CATION_TRANSPORT	147	8981	2.64	**8.23E-03**	0.24	0.62	0.54
NEGATIVE_REGULATION_OF_INTRACELLULAR_TRANSPORT	11	419	2.64	**8.33E-03**	0.16	-2.02	0.04
REGULATION_OF_MITOTIC_CELL_CYCLE	23	1475	2.64	**8.36E-03**	0.23	-1.86	0.06
POSITIVE_REGULATION_OF_BIOLOGICAL_PROCESS	708	28558	2.63	**8.42E-03**	0.15	1.42	0.16
CATION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	213	14075	2.63	**8.50E-03**	0.19	0.38	0.70
REGULATION_OF_GROWTH	58	2006	2.61	**9.13E-03**	0.18	0.11	0.91
PROTEIN_AMINO_ACID_AUTOPHOSPHORYLATION	31	1756	2.61	**9.19E-03**	0.17	-0.93	0.35
POSITIVE_REGULATION_OF_CELL_DIFFERENTIATION	25	1300	2.60	**9.33E-03**	0.18	0.93	0.35
PYRIMIDINE_NUCLEOTIDE_METABOLIC_PROCESS	9	417	2.59	**9.55E-03**	0.19	-3.05	0.00
CELL_PROJECTION	109	9285	2.58	**9.85E-03**	0.19	0.19	0.85
APICAL_JUNCTION_COMPLEX	34	1445	2.58	**1.00E-02**	0.15	0.75	0.45
APICOLATERAL_PLASMA_MEMBRANE	34	1445	2.58	**1.00E-02**	0.2	0.75	0.45
ENDOTHELIAL_CELL_PROLIFERATION	12	2366	2.56	**1.04E-02**	0.21	2.10	0.04
REGULATION_OF_GENE_EXPRESSION	670	27068	2.56	**1.05E-02**	0.2	1.72	0.09
EXTRACELLULAR_REGION	446	19333	2.56	**1.06E-02**	0.19	1.85	0.06
PROTEIN_KINASE_CASCADE	293	9977	2.55	**1.09E-02**	0.25	-2.01	0.04

CYTOSKELETON	367	17331	2.54	**1.10E-02**	0.22	-1.39	0.16
PROTEIN_HOMOOLIGOMERIZATION	22	1136	2.53	**1.15E-02**	0.15	-0.51	0.61
SMALL_GTPASE_REGULATOR_ACTIVITY	67	4758	2.52	**1.16E-02**	0.2	0.22	0.83
INTEGRAL_TO_PLASMA_MEMBRANE	974	55375	2.52	**1.19E-02**	0.19	0.60	0.55
PROTEIN_AUTOPROCESSING	32	1778	2.51	**1.22E-02**	0.21	-0.97	0.33
CIRCADIAN_RHYTHM	14	427	2.50	**1.23E-02**	0.17	1.89	0.06
PROTEIN_DOMAIN_SPECIFIC_BINDING	72	6127	2.50	**1.23E-02**	0.21	0.52	0.61
PROTEIN_PROCESSING	49	2419	2.50	**1.25E-02**	0.23	-1.18	0.24
NEGATIVE_REGULATION_OF_CELL_CYCLE	79	2943	2.49	**1.26E-02**	0.22	-1.25	0.21
REGULATION_OF_ENDOTHELIAL_CELL_PROLIFERATION	10	2263	2.49	**1.28E-02**	0.22	2.14	0.03
INTRINSIC_TO_ORGANELLE_MEMBRANE	52	2431	2.48	**1.30E-02**	0.14	-1.19	0.24
EXOCYTOSIS	25	2466	2.48	**1.32E-02**	0.18	-1.90	0.06
ESTABLISHMENT_OF_LOCALIZATION	871	39566	2.48	**1.32E-02**	0.2	0.05	0.96
ION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	278	15613	2.47	**1.35E-02**	0.17	0.69	0.49
NEGATIVE_REGULATION_OF_CELL_ADHESION	17	2471	2.47	**1.36E-02**	0.24	1.86	0.06
CALCIUM_MEDIATED_SIGNALING	16	2638	2.46	**1.39E-02**	0.23	2.08	0.04
REGULATION_OF_PROTEIN_AMINO_ACID_PHOSPHORYLATION	30	927	2.46	**1.40E-02**	0.17	0.58	0.56
CALCIUM_ION_TRANSPORT	27	2921	2.45	**1.43E-02**	0.19	-0.88	0.38
SOLUTE_SODIUM_SYMPORTER_ACTIVITY	13	384	2.44	**1.45E-02**	0.15	1.03	0.30
STRUCTURAL_CONSTITUENT_OF_CYTOSKELETON	57	2372	2.44	**1.48E-02**	0.11	-0.43	0.67
MAGNESIUM_ION_BINDING	61	2716	2.43	**1.51E-02**	0.18	-0.48	0.63
STEROID_METABOLIC_PROCESS	71	2886	2.43	**1.52E-02**	0.24	0.20	0.84
CYTOPLASM	2124	88292	2.43	**1.52E-02**	0.27	-0.79	0.43
GOLGI_APPARATUS	226	15099	2.42	**1.54E-02**	0.21	0.81	0.42
LOW_DENSITY_LIPOPROTEIN_BINDING	12	2734	2.42	**1.55E-02**	0.24	1.63	0.10
INTERCELLULAR_JUNCTION	65	5279	2.39	**1.67E-02**	0.25	0.22	0.82
INTEGRAL_TO_ORGANELLE_MEMBRANE	50	2418	2.38	**1.72E-02**	0.22	-1.22	0.22
REGULATION_OF_CELL_ADHESION	36	3020	2.38	**1.73E-02**	0.19	1.57	0.12
PROTEIN_SERINE_THREONINE_TYROSINE_KINASE_ACTIVITY	10	706	2.37	**1.77E-02**	0.17	-0.32	0.75
CDC42_PROTEIN_SIGNAL_TRANSDUCTION	13	669	2.36	**1.82E-02**	0.22	-0.20	0.84
RESPONSE_TO_CHEMICAL_STIMULUS	314	11247	2.36	**1.82E-02**	0.25	0.56	0.57
ADP_BINDING	11	944	2.36	**1.83E-02**	0.26	-1.23	0.22
NEGATIVE_REGULATION_OF_CELL_DIFFERENTIATION	27	2626	2.36	**1.84E-02**	0.25	1.19	0.23
MOTOR_ACTIVITY	27	1587	2.35	**1.86E-02**	0.21	-1.61	0.11
TRANSPORT	796	37525	2.34	**1.92E-02**	0.27	-0.21	0.83
IDENTICAL_PROTEIN_BINDING	304	12745	2.34	**1.94E-02**	0.27	-0.88	0.38
POSITIVE_REGULATION_OF_CELL_MIGRATION	10	2335	2.34	**1.95E-02**	0.18	2.41	0.02
LIGASE_ACTIVITY__FORMING_CARBON_NITROGEN_BONDS	68	3621	2.33	**1.98E-02**	0.15	-0.17	0.86
FUCOSYLTRANSFERASE_ACTIVITY	10	568	2.31	**2.08E-02**	0.26	0.36	0.72
MACROMOLECULAR_COMPLEX_ASSEMBLY	279	10262	2.31	**2.08E-02**	0.35	1.17	0.24
PROTEIN_OLIGOMERIZATION	40	1937	2.31	**2.10E-02**	0.25	0.43	0.66
REGULATION_OF_MYELOID_CELL_DIFFERENTIATION	19	632	2.30	**2.14E-02**	0.23	0.96	0.34
ACTIN_BINDING	76	6375	2.30	**2.15E-02**	0.3	-2.31	0.02
INTRINSIC_TO_MEMBRANE	1342	72016	2.28	**2.24E-02**	0.25	-0.76	0.44
JAK_STAT_CASCADE	31	673	2.28	**2.26E-02**	0.23	1.21	0.23
LIPID_TRANSPORTER_ACTIVITY	28	819	2.28	**2.29E-02**	0.23	0.17	0.87
TRANSCRIPTION_COFACTOR_ACTIVITY	228	8553	2.28	**2.29E-02**	0.24	0.92	0.36
REGULATION_OF_PROTEIN_STABILITY	19	527	2.27	**2.32E-02**	0.22	-0.77	0.44
ADENYL_RIBONUCLEOTIDE_BINDING	162	7754	2.27	**2.33E-02**	0.22	-1.71	0.09
REGULATION_OF_CELL_SHAPE	14	608	2.26	**2.35E-02**	0.27	-0.15	0.88
PDZ_DOMAIN_BINDING	14	2891	2.26	**2.37E-02**	0.27	-0.04	0.97
SULFUR_METABOLIC_PROCESS	37	2277	2.25	**2.42E-02**	0.22	0.61	0.54
CELL_CELL_ADHESION	86	9753	2.25	**2.47E-02**	0.27	0.78	0.44
NEGATIVE_REGULATION_OF_SECRETION	13	233	2.24	**2.48E-02**	0.34	-0.65	0.52
LIPOPROTEIN_BINDING	18	2905	2.24	**2.52E-02**	0.27	1.51	0.13
LOCOMOTORY_BEHAVIOR	97	3949	2.23	**2.54E-02**	0.3	1.12	0.26
MOLECULAR_ADAPTOR_ACTIVITY	49	3362	2.23	**2.58E-02**	0.23	-0.72	0.47
PEPTIDE_BINDING	85	2700	2.22	**2.62E-02**	0.2	0.04	0.97
PROTEIN_COMPLEX_ASSEMBLY	167	7687	2.22	**2.64E-02**	0.27	1.79	0.07
ANATOMICAL_STRUCTURE_MORPHOGENESIS	376	21324	2.22	**2.67E-02**	0.32	1.34	0.18
ADENYL_NUCLEOTIDE_BINDING	168	7878	2.21	**2.68E-02**	0.28	-1.62	0.11
INDUCTION_OF_APOPTOSIS_BY_EXTRACELLULAR_SIGNALS	27	1293	2.21	**2.68E-02**	0.32	-0.18	0.86
CALCIUM_ION_BINDING	103	4138	2.21	**2.72E-02**	0.23	0.64	0.52
UDP_GLYCOSYLTRANSFERASE_ACTIVITY	42	2193	2.20	**2.75E-02**	0.28	0.65	0.51
PHOSPHOINOSITIDE_BINDING	20	1174	2.20	**2.76E-02**	0.14	0.04	0.97
EXTRACELLULAR_REGION_PART	337	16198	2.20	**2.79E-02**	0.24	1.23	0.22
REGULATION_OF_PROTEIN_MODIFICATION_PROCESS	44	1333	2.20	**2.81E-02**	0.27	0.31	0.76
POSITIVE_REGULATION_OF_CELLULAR_PROCESS	667	27857	2.18	**2.91E-02**	0.23	1.66	0.10
BASAL_LAMINA	21	2325	2.18	**2.91E-02**	0.29	-1.94	0.05
REGULATION_OF_CELL_PROLIFERATION	308	11539	2.18	**2.96E-02**	0.25	1.25	0.21
PHOSPHORIC_MONOESTER_HYDROLASE_ACTIVITY	111	10136	2.17	**2.98E-02**	0.25	-0.38	0.70
STRUCTURAL_CONSTITUENT_OF_MUSCLE	33	1714	2.17	**3.01E-02**	0.19	0.30	0.76
DETECTION_OF_STIMULUS	47	2191	2.16	**3.07E-02**	0.31	0.58	0.56
TRANSCRIPTION_FROM_RNA_POLYMERASE_II_PROMOTER	454	16224	2.16	**3.11E-02**	0.29	1.88	0.06
DEAMINASE_ACTIVITY	14	752	2.15	**3.13E-02**	0.26	-1.82	0.07
NF_KAPPAB_BINDING	11	418	2.15	**3.15E-02**	0.26	-2.10	0.04
ENDOTHELIAL_CELL_MIGRATION	11	2380	2.13	**3.28E-02**	0.28	2.73	0.01
SMALL_PROTEIN_CONJUGATING_ENZYME_ACTIVITY	52	3421	2.13	**3.35E-02**	0.24	0.47	0.64
PROTEIN_COMPLEX	814	31980	2.12	**3.40E-02**	0.26	1.57	0.12
POSITIVE_REGULATION_OF_CASPASE_ACTIVITY	30	789	2.12	**3.41E-02**	0.27	0.81	0.42
SISTER_CHROMATID_SEGREGATION	17	395	2.11	**3.44E-02**	0.28	0.91	0.36
MITOTIC_SISTER_CHROMATID_SEGREGATION	16	382	2.11	**3.49E-02**	0.33	0.62	0.54
REGULATION_OF_SYNAPSE_STRUCTURE_AND_ACTIVITY	11	1428	2.11	**3.53E-02**	0.32	2.29	0.02
TRANSMEMBRANE_RECEPTOR_ACTIVITY	414	30514	2.10	**3.61E-02**	0.3	-0.35	0.72
TRANSCRIPTION_INITIATION_FROM_RNA_POLYMERASE_II_PROMOTER	29	734	2.08	**3.76E-02**	0.21	-0.34	0.73
METAL_ION_TRANSPORT	117	7399	2.08	**3.80E-02**	0.28	0.59	0.56
S_PHASE	14	1224	2.07	**3.83E-02**	0.28	-0.24	0.81
RHO_GTPASE_ACTIVATOR_ACTIVITY	19	1419	2.07	**3.85E-02**	0.31	-0.43	0.67
CELL_MATURATION	16	1717	2.07	**3.85E-02**	0.3	0.52	0.61
GENERATION_OF_A_SIGNAL_INVOLVED_IN_CELL_CELL_SIGNALING	29	2263	2.07	**3.86E-02**	0.27	0.03	0.98
RNA_ELONGATION	10	297	2.07	**3.87E-02**	0.26	0.23	0.82
ENDOSOME_TRANSPORT	23	1029	2.06	**3.92E-02**	0.26	2.30	0.02
CALCIUM_CHANNEL_ACTIVITY	33	4323	2.06	**3.94E-02**	0.34	-0.18	0.85
ACETYLGLUCOSAMINYLTRANSFERASE_ACTIVITY	16	1029	2.06	**3.97E-02**	0.23	0.92	0.36
AMINE_RECEPTOR_ACTIVITY	34	1157	2.06	**3.98E-02**	0.37	0.96	0.34
GLUCOSAMINE_METABOLIC_PROCESS	13	725	2.05	**4.01E-02**	0.21	-0.07	0.94
NEGATIVE_REGULATION_OF_BIOSYNTHETIC_PROCESS	30	1386	2.05	**4.01E-02**	0.31	0.19	0.85
CARBOHYDRATE_METABOLIC_PROCESS	179	6629	2.05	**4.03E-02**	0.23	1.52	0.13
REGULATION_OF_BODY_FLUID_LEVELS	56	2306	2.05	**4.07E-02**	0.25	0.61	0.54
PHOSPHOPROTEIN_PHOSPHATASE_ACTIVITY	81	8711	2.04	**4.15E-02**	0.22	-0.19	0.85
PROGRAMMED_CELL_DEATH	432	16688	2.04	**4.17E-02**	0.29	1.43	0.15
APOPTOSIS_GO	431	16687	2.03	**4.20E-02**	0.3	1.43	0.15
REGULATION_OF_CELL_MORPHOGENESIS	15	636	2.03	**4.23E-02**	0.32	0.00	1.00
REGULATION_OF_BLOOD_PRESSURE	21	372	2.02	**4.35E-02**	0.25	0.35	0.73
CELLULAR_HOMEOSTASIS	147	6640	2.00	**4.56E-02**	0.31	-0.01	0.99
CASPASE_ACTIVATION	26	724	1.99	**4.66E-02**	0.31	0.74	0.46
EXTRACELLULAR_MATRIX_PART	57	6204	1.99	**4.67E-02**	0.35	-0.78	0.43

REGULATION_OF_PROTEIN_IMPORT_INTO_NUCLEUS	16	571	1.99	**4.70E-02**	0.23	-1.78	0.08
EMBRYONIC_MORPHOGENESIS	17	1155	1.98	**4.72E-02**	0.31	1.55	0.12
INTERLEUKIN_BINDING	26	459	1.98	**4.74E-02**	0.26	-0.02	0.98
TRANSCRIPTION_FACTOR_ACTIVITY	350	11244	1.98	**4.78E-02**	0.38	-0.16	0.87
N_ACETYLGLUCOSAMINE_METABOLIC_PROCESS	12	715	1.98	**4.78E-02**	0.24	-0.02	0.99
PROTEASE_INHIBITOR_ACTIVITY	40	1128	1.98	**4.82E-02**	0.24	1.21	0.23
GTPASE_ACTIVITY	99	2470	1.97	**4.84E-02**	0.25	0.70	0.48
DELAYED_RECTIFIER_POTASSIUM_CHANNEL_ACTIVITY	12	1134	1.97	**4.88E-02**	0.31	1.36	0.17
GUANYL_NUCLEOTIDE_EXCHANGE_FACTOR_ACTIVITY	47	5201	1.97	**4.92E-02**	0.39	0.44	0.66
SUBSTRATE_SPECIFIC_TRANSPORTER_ACTIVITY	392	18628	1.97	**4.94E-02**	0.34	0.82	0.41
MACROMOLECULAR_COMPLEX_DISASSEMBLY	15	1360	1.96	**4.97E-02**	0.28	-0.38	0.70
MYELOID_LEUKOCYTE_DIFFERENTIATION	15	367	1.95	5.07E-02	NA	1.41	0.16
SUGAR_BINDING	34	302	1.95	5.11E-02	NA	-0.13	0.90
INTERLEUKIN_RECEPTOR_ACTIVITY	21	403	1.94	5.27E-02	NA	0.38	0.70
GTPASE_ACTIVATOR_ACTIVITY	60	5676	1.93	5.39E-02	NA	0.91	0.36
EXTRINSIC_TO_PLASMA_MEMBRANE	13	494	1.92	5.46E-02	NA	0.25	0.80
TRANSLATION_INITIATION_FACTOR_ACTIVITY	24	351	1.92	5.48E-02	NA	-0.97	0.33
RNA_POLYMERASE_II_TRANSCRIPTION_FACTOR_ACTIVITY	182	6296	1.92	5.52E-02	NA	0.10	0.92
INTEGRAL_TO_MEMBRANE	1324	70793	1.91	5.65E-02	NA	-0.82	0.41
ICOSANOID_METABOLIC_PROCESS	17	273	1.90	5.75E-02	NA	-0.39	0.70
ACTIN_FILAMENT_BINDING	25	2314	1.90	5.77E-02	NA	-1.23	0.22
MICROTUBULE_MOTOR_ACTIVITY	15	634	1.90	5.78E-02	NA	-0.97	0.33
SYNAPSE_PART	13	3315	1.89	5.82E-02	NA	1.18	0.24
CELL_MIGRATION	96	9205	1.89	5.86E-02	NA	1.40	0.16
MALE_GONAD_DEVELOPMENT	12	310	1.89	5.86E-02	NA	2.30	0.02
RUFFLE	32	1383	1.89	5.87E-02	NA	0.05	0.96
RESPONSE_TO_OXIDATIVE_STRESS	46	1352	1.88	5.96E-02	NA	0.85	0.39
MACROMOLECULE_BIOSYNTHETIC_PROCESS	321	10077	1.88	6.06E-02	NA	-0.09	0.93
AMINO_SUGAR_METABOLIC_PROCESS	19	844	1.87	6.09E-02	NA	-0.25	0.80
VOLTAGE_GATED_POTASSIUM_CHANNEL_COMPLEX	40	4022	1.87	6.10E-02	NA	2.17	0.03
PROTEIN_DIMERIZATION_ACTIVITY	181	8822	1.87	6.13E-02	NA	0.82	0.41
PHOSPHOLIPASE_A2_ACTIVITY	13	271	1.86	6.27E-02	NA	0.50	0.62
METALLOPEPTIDASE_ACTIVITY	50	2362	1.86	6.27E-02	NA	0.19	0.85
REGULATION_OF_PROGRAMMED_CELL_DEATH	342	13592	1.86	6.35E-02	NA	0.50	0.62
EPIDERMAL_GROWTH_FACTOR_RECEPTOR_SIGNALING_PATHWAY	22	3191	1.85	6.38E-02	NA	1.90	0.06
REGULATION_OF_APOPTOSIS	341	13591	1.85	6.38E-02	NA	0.49	0.62
DEVELOPMENTAL_MATURATION	18	1907	1.85	6.45E-02	NA	0.89	0.38
NEUROTRANSMITTER_SECRETION	13	2065	1.84	6.53E-02	NA	0.06	0.95
PHOSPHATASE_BINDING	13	2194	1.84	6.56E-02	NA	0.60	0.55
PERIPHERAL_NERVOUS_SYSTEM_DEVELOPMENT	12	1545	1.84	6.60E-02	NA	-1.76	0.08
MICROVILLUS	11	2335	1.84	6.61E-02	NA	0.48	0.63
AEROBIC_RESPIRATION	15	476	1.84	6.62E-02	NA	0.30	0.76
G2_M_TRANSITION_OF_MITOTIC_CELL_CYCLE	13	303	1.82	6.83E-02	NA	1.94	0.05
NUCLEOTIDE_BIOSYNTHETIC_PROCESS	19	2084	1.82	6.84E-02	NA	-1.39	0.16
TRANSFERASE_ACTIVITY__TRANSFERRING_PENTOSYL_GROUPS	20	842	1.82	6.85E-02	NA	0.76	0.45
OLIGOSACCHARIDE_METABOLIC_PROCESS	11	708	1.82	6.94E-02	NA	-0.35	0.72
PROTEIN_SERINE_THREONINE_KINASE_ACTIVITY	204	10695	1.81	6.98E-02	NA	-1.93	0.05
PROTEOGLYCAN_METABOLIC_PROCESS	21	1178	1.81	6.99E-02	NA	1.65	0.10
GLYCOSPHINGOLIPID_METABOLIC_PROCESS	12	867	1.81	7.01E-02	NA	0.74	0.46
EXOPEPTIDASE_ACTIVITY	32	1778	1.81	7.03E-02	NA	0.57	0.57
PHOSPHOLIPID_TRANSPORTER_ACTIVITY	12	585	1.81	7.06E-02	NA	-0.30	0.77
EXTRACELLULAR_MATRIX_STRUCTURAL_CONSTITUENT	27	1529	1.81	7.07E-02	NA	1.41	0.16
CATION_BINDING	211	7261	1.80	7.13E-02	NA	-0.76	0.45
POSITIVE_REGULATION_OF_CELL_PROLIFERATION	149	6459	1.80	7.17E-02	NA	2.79	0.01
NEUROTRANSMITTER_RECEPTOR_ACTIVITY	49	1320	1.80	7.21E-02	NA	1.52	0.13
RECEPTOR_COMPLEX	56	3168	1.80	7.24E-02	NA	1.24	0.22
INTEGRAL_TO_ENDOPLASMIC_RETICULUM_MEMBRANE	24	1020	1.79	7.28E-02	NA	-1.51	0.13
INTRINSIC_TO_ENDOPLASMIC_RETICULUM_MEMBRANE	24	1020	1.79	7.28E-02	NA	-1.51	0.13
LIPASE_ACTIVITY	50	1707	1.79	7.34E-02	NA	-0.48	0.63
ENZYME_INHIBITOR_ACTIVITY	118	2785	1.79	7.41E-02	NA	0.64	0.52
GLYCOLIPID_METABOLIC_PROCESS	16	930	1.78	7.53E-02	NA	0.27	0.79
PROTEIN_BINDING__BRIDGING	59	3087	1.78	7.56E-02	NA	-0.58	0.56
DI___TRI_VALENT_INORGANIC_CATION_TRANSPORT	32	2993	1.76	7.89E-02	NA	-0.72	0.47
CORNIFIED_ENVELOPE	13	289	1.75	8.03E-02	NA	1.10	0.27
REGULATION_OF_TRANSCRIPTION_FROM_RNA_POLYMERASE_II_PROMOTER	288	11133	1.75	8.09E-02	NA	2.35	0.02
REGULATION_OF_TRANSCRIPTION	563	22951	1.74	8.10E-02	NA	1.20	0.23
REGULATION_OF_ACTIN_FILAMENT_LENGTH	13	776	1.74	8.10E-02	NA	0.89	0.37
SUBSTRATE_SPECIFIC_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	344	17465	1.74	8.23E-02	NA	1.07	0.29
SERINE_TYPE_ENDOPEPTIDASE_INHIBITOR_ACTIVITY	24	829	1.73	8.36E-02	NA	1.01	0.31
MUSCLE_DEVELOPMENT	93	6866	1.73	8.38E-02	NA	-0.05	0.96
ORGAN_MORPHOGENESIS	144	7067	1.73	8.39E-02	NA	1.24	0.22
G1_S_TRANSITION_OF_MITOTIC_CELL_CYCLE	27	763	1.73	8.39E-02	NA	0.68	0.50
U12_DEPENDENT_SPLICEOSOME	11	148	1.73	8.42E-02	NA	1.27	0.20
REGULATED_SECRETORY_PATHWAY	15	2141	1.73	8.42E-02	NA	-0.18	0.86
NEUROTRANSMITTER_BINDING	52	1343	1.73	8.45E-02	NA	1.65	0.10
DNA_MODIFICATION	11	263	1.72	8.63E-02	NA	-0.71	0.48
UBIQUITIN_PROTEIN_LIGASE_ACTIVITY	49	3016	1.72	8.63E-02	NA	-0.05	0.96
EXTRACELLULAR_SPACE	244	8997	1.71	8.71E-02	NA	1.36	0.17
TRANS_GOLGI_NETWORK_TRANSPORT_VESICLE	13	332	1.71	8.72E-02	NA	0.96	0.34
REGULATION_OF_CELLULAR_COMPONENT_SIZE	14	850	1.70	8.82E-02	NA	0.41	0.68
PROTEOGLYCAN_BIOSYNTHETIC_PROCESS	15	1113	1.70	8.99E-02	NA	1.05	0.29
SMALL_CONJUGATING_PROTEIN_LIGASE_ACTIVITY	51	3047	1.70	9.00E-02	NA	-0.20	0.84
RESPONSE_TO_EXTERNAL_STIMULUS	311	10275	1.68	9.22E-02	NA	1.78	0.08
ESTABLISHMENT_OF_ORGANELLE_LOCALIZATION	18	1121	1.68	9.34E-02	NA	-0.58	0.56
ANTI_APOPTOSIS	118	5329	1.67	9.40E-02	NA	0.30	0.77
ENZYME_ACTIVATOR_ACTIVITY	125	7428	1.67	9.42E-02	NA	0.47	0.64
SODIUM_CHANNEL_ACTIVITY	17	1909	1.67	9.49E-02	NA	-1.55	0.12
MONOSACCHARIDE_BINDING	11	119	1.67	9.54E-02	NA	0.14	0.89
RAS_GTPASE_ACTIVATOR_ACTIVITY	29	2037	1.67	9.58E-02	NA	0.04	0.96
BIOPOLYMER_METABOLIC_PROCESS	1678	69167	1.66	9.69E-02	NA	0.32	0.75
MACROMOLECULAR_COMPLEX	943	33584	1.65	9.80E-02	NA	1.74	0.08
TRNA_PROCESSING	10	596	1.65	9.81E-02	NA	0.28	0.78
CELL_JUNCTION	82	6193	1.65	9.92E-02	NA	0.74	0.46
MRNA_BINDING	23	391	1.64	1.00E-01	NA	0.12	0.91
NEGATIVE_REGULATION_OF_CYTOKINE_BIOSYNTHETIC_PROCESS	12	50	1.64	1.00E-01	NA	0.14	0.89
N_TERMINAL_PROTEIN_AMINO_ACID_MODIFICATION	11	261	1.64	1.00E-01	NA	0.57	0.57
AMINE_TRANSPORT	38	1348	1.64	1.00E-01	NA	-0.16	0.87
BIOGENIC_AMINE_METABOLIC_PROCESS	17	421	1.64	1.02E-01	NA	0.20	0.84
LIPID_TRANSPORT	28	590	1.63	1.04E-01	NA	1.33	0.18
RAS_GUANYL_NUCLEOTIDE_EXCHANGE_FACTOR_ACTIVITY	19	2480	1.63	1.04E-01	NA	0.57	0.57
CARBOHYDRATE_BIOSYNTHETIC_PROCESS	49	2831	1.62	1.05E-01	NA	0.42	0.68
RECEPTOR_SIGNALING_PROTEIN_ACTIVITY	82	4248	1.62	1.05E-01	NA	1.05	0.29
NEGATIVE_REGULATION_OF_CELLULAR_METABOLIC_PROCESS	257	10459	1.62	1.05E-01	NA	0.90	0.37
PROTEIN_COMPLEX_DISASSEMBLY	14	1343	1.62	1.06E-01	NA	-0.27	0.79
POSITIVE_REGULATION_OF_EPITHELIAL_CELL_PROLIFERATION	10	685	1.62	1.06E-01	NA	0.45	0.65
ENZYME_BINDING	179	8756	1.61	1.07E-01	NA	0.45	0.66
HOMEOSTATIC_PROCESS	207	8108	1.61	1.08E-01	NA	0.05	0.96

OXIDOREDUCTASE_ACTIVITY	286	8971	1.60	1.10E-01	NA	0.81	0.42
PROTEIN_KINASE_INHIBITOR_ACTIVITY	24	273	1.59	1.11E-01	NA	0.79	0.43
CARBOXYPEPTIDASE_ACTIVITY	12	650	1.59	1.11E-01	NA	0.26	0.80
MYOSIN_COMPLEX	16	536	1.59	1.11E-01	NA	0.96	0.33
SULFUR_COMPOUND_BIOSYNTHETIC_PROCESS	18	869	1.59	1.12E-01	NA	0.62	0.53
EXTRACELLULAR_MATRIX	100	7541	1.59	1.12E-01	NA	-0.22	0.82
CELLULAR_PROTEIN_COMPLEX_DISASSEMBLY	13	1323	1.59	1.12E-01	NA	-0.29	0.77
MONOCARBOXYLIC_ACID_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	11	411	1.59	1.13E-01	NA	1.21	0.23
PROTEINACEOUS_EXTRACELLULAR_MATRIX	98	7537	1.58	1.14E-01	NA	-0.23	0.82
DETECTION_OF_STIMULUS_INVOLVED_IN_SENSORY_PERCEPTION	21	224	1.58	1.15E-01	NA	-0.79	0.43
REGULATION_OF_CELL_MIGRATION	28	2855	1.58	1.15E-01	NA	1.94	0.05
REGULATION_OF_MOLECULAR_FUNCTION	325	11952	1.57	1.16E-01	NA	0.27	0.79
MITOCHONDRIAL_MEMBRANE_ORGANIZATION_AND_BIOGENESIS	11	311	1.57	1.17E-01	NA	-1.02	0.31
DYSTROPHIN_ASSOCIATED_GLYCOPROTEIN_COMPLEX	14	2003	1.56	1.19E-01	NA	-2.04	0.04
NEURON_PROJECTION	21	2697	1.56	1.20E-01	NA	2.77	0.01
CELLULAR_CARBOHYDRATE_METABOLIC_PROCESS	126	4934	1.55	1.21E-01	NA	0.89	0.37
CELL_CORTEX	39	2309	1.54	1.23E-01	NA	-1.10	0.27
ER_NUCLEAR_SIGNALING_PATHWAY	16	638	1.54	1.23E-01	NA	-0.45	0.65
POTASSIUM_ION_TRANSPORT	58	3588	1.53	1.26E-01	NA	2.01	0.04
REGULATION_OF_MAPKKK_CASCADE	20	529	1.53	1.27E-01	NA	0.62	0.53
REGULATION_OF_MUSCLE_CONTRACTION	19	923	1.51	1.30E-01	NA	0.38	0.71
METALLOEXOPEPTIDASE_ACTIVITY	13	708	1.51	1.30E-01	NA	0.12	0.91
REGULATION_OF_NUCLEOBASE__NUCLEOSIDE__NUCLEOTIDE_AND_NUCLEIC_ACID_METABOLIC_PROCESS	615	25303	1.51	1.31E-01	NA	1.24	0.21
ATP_BINDING	155	6981	1.51	1.32E-01	NA	-1.24	0.21
PROTEIN_TRANSPORTER_ACTIVITY	14	378	1.50	1.33E-01	NA	-1.61	0.11
ACID_AMINO_ACID_LIGASE_ACTIVITY	57	3116	1.50	1.33E-01	NA	-0.21	0.84
TRANSCRIPTION_COACTIVATOR_ACTIVITY	123	3636	1.50	1.34E-01	NA	-0.03	0.98
DETECTION_OF_CHEMICAL_STIMULUS	18	1361	1.50	1.34E-01	NA	0.68	0.50
AXON_GUIDANCE	22	4373	1.50	1.34E-01	NA	-0.67	0.50
CHROMOSOME_SEGREGATION	32	822	1.50	1.34E-01	NA	1.68	0.09
ORGANELLE_LOCALIZATION	25	1276	1.49	1.35E-01	NA	-1.03	0.30
METALLOENDOPEPTIDASE_ACTIVITY	27	762	1.49	1.38E-01	NA	-0.03	0.98
TRANSCRIPTION	748	27438	1.48	1.38E-01	NA	1.07	0.29
CASPASE_REGULATOR_ACTIVITY	14	528	1.48	1.39E-01	NA	2.13	0.03
ION_TRANSPORT	185	10271	1.48	1.39E-01	NA	0.49	0.62
PHOSPHOINOSITIDE_MEDIATED_SIGNALING	48	2667	1.48	1.40E-01	NA	1.65	0.10
TRANSFERASE_ACTIVITY__TRANSFERRING_SULFUR_CONTAINING_GROUPS	32	1344	1.48	1.40E-01	NA	0.14	0.89
PROTEIN_AMINO_ACID_ADP_RIBOSYLATION	10	243	1.48	1.40E-01	NA	-0.92	0.36
CELL_CORTEX_PART	24	1768	1.46	1.43E-01	NA	-1.19	0.23
NEGATIVE_REGULATION_OF_METABOLIC_PROCESS	260	10489	1.46	1.44E-01	NA	0.91	0.37
PHOSPHOLIPASE_ACTIVITY	42	1480	1.46	1.44E-01	NA	-0.84	0.40
REGULATION_OF_CATALYTIC_ACTIVITY	277	10669	1.46	1.45E-01	NA	0.56	0.58
SPERM_MOTILITY	11	123	1.46	1.45E-01	NA	1.13	0.26
MUSCLE_CELL_DIFFERENTIATION	22	1934	1.45	1.48E-01	NA	0.62	0.54
PHOTOTRANSDUCTION	13	210	1.44	1.49E-01	NA	-1.12	0.26
NEGATIVE_REGULATION_OF_TRANSLATION	23	230	1.44	1.49E-01	NA	1.07	0.29
SPECIFIC_RNA_POLYMERASE_II_TRANSCRIPTION_FACTOR_ACTIVITY	35	1423	1.44	1.50E-01	NA	1.13	0.26
SULFOTRANSFERASE_ACTIVITY	28	1268	1.43	1.53E-01	NA	-0.03	0.98
DEVELOPMENT_OF_PRIMARY_SEXUAL_CHARACTERISTICS	26	1179	1.43	1.53E-01	NA	2.62	0.01
ADHERENS_JUNCTION	23	3106	1.43	1.53E-01	NA	-0.69	0.49
REGULATION_OF_SIGNAL_TRANSDUCTION	223	8628	1.43	1.53E-01	NA	0.05	0.96
LIPID_RAFT	29	997	1.43	1.54E-01	NA	-1.15	0.25
NEGATIVE_REGULATION_OF_TRANSPORT	20	564	1.42	1.55E-01	NA	-2.09	0.04
PROTEIN_DNA_COMPLEX_ASSEMBLY	48	1388	1.42	1.56E-01	NA	0.95	0.34
SMALL_GTPASE_BINDING	34	1914	1.42	1.57E-01	NA	0.01	0.99
IONOTROPIC_GLUTAMATE_RECEPTOR_ACTIVITY	10	2971	1.41	1.58E-01	NA	-0.87	0.38
CYCLIC_NUCLEOTIDE_METABOLIC_PROCESS	11	1231	1.41	1.59E-01	NA	0.22	0.83
EXTERNAL_SIDE_OF_PLASMA_MEMBRANE	16	783	1.41	1.60E-01	NA	0.28	0.78
NLS_BEARING_SUBSTRATE_IMPORT_INTO_NUCLEUS	13	844	1.41	1.60E-01	NA	0.39	0.70
POSITIVE_REGULATION_OF_METABOLIC_PROCESS	235	10308	1.40	1.61E-01	NA	0.38	0.70
PHOSPHOTRANSFERASE_ACTIVITY__PHOSPHATE_GROUP_AS_ACCEPTOR	18	1928	1.40	1.61E-01	NA	-1.18	0.24
INWARD_RECTIFIER_POTASSIUM_CHANNEL_ACTIVITY	12	1014	1.40	1.63E-01	NA	1.87	0.06
G_PROTEIN_SIGNALING__COUPLED_TO_IP3_SECOND_MESSENGER__PHOSPHOLIPASE_C_ACTIVATING	45	2360	1.39	1.64E-01	NA	1.75	0.08
CYCLASE_ACTIVITY	10	747	1.39	1.66E-01	NA	1.75	0.08
NEGATIVE_REGULATION_OF_CELLULAR_BIOSYNTHETIC_PROCESS	29	1339	1.38	1.68E-01	NA	0.46	0.65
MICROTUBULE_BINDING	33	1570	1.37	1.69E-01	NA	-0.68	0.49
GTPASE_BINDING	35	1954	1.37	1.70E-01	NA	0.09	0.93
THYROID_HORMONE_RECEPTOR_BINDING	17	466	1.37	1.71E-01	NA	-0.12	0.90
POSITIVE_REGULATION_OF_NUCLEOBASE__NUCLEOSIDE__NUCLEOTIDE_AND_NUCLEIC_ACID_METABOLIC_PROCESS	153	8421	1.37	1.72E-01	NA	0.21	0.83
NUCLEOTIDE_KINASE_ACTIVITY	13	1912	1.36	1.73E-01	NA	-1.33	0.18
POSITIVE_REGULATION_OF_ANGIOGENESIS	10	158	1.36	1.75E-01	NA	1.05	0.29
INTEGRATOR_COMPLEX	13	405	1.36	1.75E-01	NA	-0.52	0.60
BETA_TUBULIN_BINDING	10	197	1.36	1.75E-01	NA	-1.05	0.29
BIOPOLYMER_BIOSYNTHETIC_PROCESS	14	160	1.35	1.76E-01	NA	0.53	0.60
NEURON_APOPTOSIS	17	1080	1.35	1.77E-01	NA	0.96	0.34
MAP_KINASE_ACTIVITY	12	682	1.35	1.78E-01	NA	0.50	0.61
ION_BINDING	269	9729	1.34	1.80E-01	NA	-1.16	0.25
REGULATION_OF_RESPONSE_TO_EXTERNAL_STIMULUS	15	2353	1.34	1.80E-01	NA	2.30	0.02
ACTIN_POLYMERIZATION_AND_OR_DEPOLYMERIZATION	23	1091	1.34	1.81E-01	NA	1.00	0.32
INTRINSIC_TO_GOLGI_MEMBRANE	15	1172	1.33	1.83E-01	NA	0.04	0.97
GALACTOSYLTRANSFERASE_ACTIVITY	15	310	1.32	1.85E-01	NA	0.57	0.57
MONOOXYGENASE_ACTIVITY	30	1017	1.32	1.87E-01	NA	-0.60	0.55
LIGAND_DEPENDENT_NUCLEAR_RECEPTOR_ACTIVITY	25	1361	1.32	1.87E-01	NA	-0.52	0.61
APICAL_PART_OF_CELL	17	1224	1.32	1.88E-01	NA	1.14	0.26
NEUROPEPTIDE_HORMONE_ACTIVITY	12	86	1.32	1.88E-01	NA	-1.10	0.27
KINASE_INHIBITOR_ACTIVITY	25	285	1.31	1.90E-01	NA	0.79	0.43
POSITIVE_REGULATION_OF_RNA_METABOLIC_PROCESS	120	7031	1.30	1.94E-01	NA	-0.13	0.90
POSITIVE_REGULATION_OF_RESPONSE_TO_STIMULUS	41	3149	1.29	1.97E-01	NA	2.05	0.04
RHO_GUANYL_NUCLEOTIDE_EXCHANGE_FACTOR_ACTIVITY	12	2001	1.29	1.99E-01	NA	0.49	0.63
REGULATION_OF_CELL_CYCLE	182	5322	1.28	2.00E-01	NA	-0.65	0.52
CYTOKINE_AND_CHEMOKINE_MEDIATED_SIGNALING_PATHWAY	23	254	1.28	2.02E-01	NA	-0.27	0.79
NUCLEAR_MATRIX	12	603	1.27	2.04E-01	NA	-0.11	0.91
POSITIVE_REGULATION_OF_TRANSCRIPTION	143	8112	1.27	2.04E-01	NA	0.00	1.00
VIRAL_REPRODUCTIVE_PROCESS	36	1023	1.27	2.05E-01	NA	-2.17	0.03
PEPTIDYL_AMINO_ACID_MODIFICATION	63	2445	1.27	2.06E-01	NA	0.53	0.60
STRESS_ACTIVATED_PROTEIN_KINASE_SIGNALING_PATHWAY	49	2088	1.26	2.06E-01	NA	-2.13	0.03
SMOOTH_MUSCLE_CONTRACTION_GO_0006939	12	1007	1.26	2.08E-01	NA	0.43	0.67
INTEGRAL_TO_GOLGI_MEMBRANE	10	1138	1.26	2.09E-01	NA	0.29	0.77
SINGLE_STRANDED_RNA_BINDING	13	950	1.26	2.09E-01	NA	-0.84	0.40
POSITIVE_REGULATION_OF_PROTEIN_AMINO_ACID_PHOSPHORYLATION	20	285	1.25	2.11E-01	NA	1.25	0.21
MYOBLAST_DIFFERENTIATION	17	1436	1.24	2.14E-01	NA	0.90	0.37
REGULATION_OF_ANATOMICAL_STRUCTURE_MORPHOGENESIS	26	2021	1.24	2.14E-01	NA	-0.40	0.69
CELLULAR_PROTEIN_COMPLEX_ASSEMBLY	33	1025	1.24	2.14E-01	NA	2.27	0.02
INOSITOL_OR_PHOSPHATIDYLINOSITOL_PHOSPHODIESTERASE_ACTIVITY	13	755	1.24	2.17E-01	NA	-0.71	0.48
CELL_SUBSTRATE_ADHERENS_JUNCTION	16	878	1.23	2.18E-01	NA	-0.21	0.83
NEGATIVE_REGULATION_OF_DEVELOPMENTAL_PROCESS	196	9554	1.23	2.19E-01	NA	1.19	0.23
UNFOLDED_PROTEIN_RESPONSE	10	389	1.23	2.20E-01	NA	-0.98	0.33

AGING	13	372	1.23	2.20E-01	NA	0.99	0.32
LYSOSOMAL_TRANSPORT	10	388	1.22	2.22E-01	NA	1.74	0.08
COPPER_ION_BINDING	15	210	1.22	2.22E-01	NA	0.82	0.41
POSITIVE_REGULATION_OF_TRANSCRIPTION__DNA_DEPENDENT	118	6953	1.22	2.23E-01	NA	-0.18	0.86
STRUCTURAL_MOLECULE_ACTIVITY	243	7056	1.20	2.28E-01	NA	0.90	0.37
TRANSCRIPTION_INITIATION	35	911	1.20	2.28E-01	NA	0.44	0.66
POSITIVE_REGULATION_OF_CELLULAR_METABOLIC_PROCESS	228	10147	1.20	2.29E-01	NA	0.74	0.46
PROTEIN_N_TERMINUS_BINDING	38	1604	1.20	2.31E-01	NA	-0.54	0.59
CYTOKINE_ACTIVITY	113	1506	1.20	2.32E-01	NA	-0.44	0.66
IMMUNOLOGICAL_SYNAPSE	11	417	1.19	2.35E-01	NA	0.55	0.58
TRANSMEMBRANE_TRANSPORTER_ACTIVITY	375	18379	1.18	2.37E-01	NA	0.47	0.64
JNK_CASCADE	47	1997	1.18	2.39E-01	NA	-2.11	0.03
TRANSFORMING_GROWTH_FACTOR_BETA_RECEPTOR_SIGNALING_PATHWAY	36	1175	1.18	2.39E-01	NA	-0.27	0.78
PHOSPHORUS_OXYGEN_LYASE_ACTIVITY	9	730	1.17	2.44E-01	NA	1.60	0.11
CENTROSOME_CYCLE	11	218	1.16	2.45E-01	NA	1.40	0.16
GOLGI_MEMBRANE	45	2138	1.16	2.46E-01	NA	0.35	0.73
RHYTHMIC_PROCESS	28	884	1.16	2.47E-01	NA	2.52	0.01
CELLULAR_COMPONENT_ASSEMBLY	297	10871	1.16	2.47E-01	NA	1.13	0.26
HISTONE_ACETYLTRANSFERASE_ACTIVITY	16	484	1.15	2.51E-01	NA	1.86	0.06
TRANS_GOLGI_NETWORK	21	3518	1.14	2.53E-01	NA	2.55	0.01
STEROID_HORMONE_RECEPTOR_ACTIVITY	13	628	1.14	2.54E-01	NA	-1.17	0.24
VACUOLAR_TRANSPORT	13	460	1.14	2.56E-01	NA	1.72	0.09
AMINO_ACID_DERIVATIVE_METABOLIC_PROCESS	24	499	1.13	2.57E-01	NA	-0.04	0.97
GLYCOPROTEIN_CATABOLIC_PROCESS	12	337	1.13	2.57E-01	NA	0.83	0.41
NUCLEOBASE__NUCLEOSIDE__NUCLEOTIDE_KINASE_ACTIVITY	24	2309	1.13	2.57E-01	NA	-1.52	0.13
PML_BODY	14	435	1.13	2.58E-01	NA	1.96	0.05
MYELOID_CELL_DIFFERENTIATION	37	1165	1.13	2.58E-01	NA	1.92	0.05
RECEPTOR_MEDIATED_ENDOCYTOSIS	33	3181	1.13	2.59E-01	NA	-0.21	0.84
ESTABLISHMENT_OF_VESICLE_LOCALIZATION	10	927	1.13	2.60E-01	NA	-1.34	0.18
ENDOCYTIC_VESICLE	14	807	1.12	2.62E-01	NA	-0.05	0.96
EXCRETION	36	805	1.11	2.67E-01	NA	0.49	0.62
REGULATION_OF_ENDOCYTOSIS	16	2393	1.10	2.70E-01	NA	2.72	0.01
VESICLE_LOCALIZATION	11	942	1.10	2.70E-01	NA	-1.40	0.16
GOLGI_APPARATUS_PART	100	7043	1.10	2.72E-01	NA	2.08	0.04
POSITIVE_REGULATION_OF_PHOSPHORYLATION	26	595	1.10	2.73E-01	NA	0.69	0.49
REGULATION_OF_ORGANELLE_ORGANIZATION_AND_BIOGENESIS	40	1933	1.09	2.74E-01	NA	0.16	0.87
AMINO_ACID_DERIVATIVE_BIOSYNTHETIC_PROCESS	10	207	1.09	2.74E-01	NA	1.42	0.16
PROTEIN_STABILIZATION	11	332	1.08	2.78E-01	NA	-0.45	0.65
HORMONE_SECRETION	17	199	1.08	2.82E-01	NA	0.09	0.93
DOUBLE_STRANDED_RNA_BINDING	17	1049	1.07	2.83E-01	NA	-0.80	0.42
NUCLEOLUS	126	2767	1.07	2.85E-01	NA	0.65	0.52
CELLULAR_RESPIRATION	19	647	1.06	2.87E-01	NA	0.36	0.72
POSITIVE_REGULATION_OF_PHOSPHATE_METABOLIC_PROCESS	28	643	1.05	2.93E-01	NA	0.51	0.61
REGULATION_OF_CATABOLIC_PROCESS	16	300	1.05	2.95E-01	NA	-2.52	0.01
POSITIVE_REGULATION_OF_MAPKKK_CASCADE	11	383	1.04	2.96E-01	NA	1.36	0.17
INTERACTION_WITH_HOST	17	513	1.04	2.96E-01	NA	-2.10	0.04
TRANSCRIPTION_COREPRESSOR_ACTIVITY	94	3100	1.04	2.97E-01	NA	0.79	0.43
GABA_RECEPTOR_ACTIVITY	11	1323	1.04	2.98E-01	NA	-0.31	0.75
GROWTH_FACTOR_BINDING	32	1433	1.04	3.00E-01	NA	0.22	0.83
PURINE_NUCLEOTIDE_METABOLIC_PROCESS	12	672	1.04	3.00E-01	NA	-1.12	0.26
PROTEIN_AMINO_ACID_DEPHOSPHORYLATION	63	7676	1.03	3.02E-01	NA	-0.59	0.56
CHROMOSOMAL_PART	95	2035	1.03	3.03E-01	NA	1.11	0.27
MONOVALENT_INORGANIC_CATION_TRANSPORT	94	4972	1.02	3.08E-01	NA	0.83	0.40
REGULATION_OF_CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS	30	1717	1.02	3.10E-01	NA	0.20	0.84
POSITIVE_REGULATION_OF_CELLULAR_COMPONENT_ORGANIZATION_AND_BIOGENESIS	36	1993	1.01	3.12E-01	NA	-1.03	0.30
NEGATIVE_REGULATION_OF_TRANSCRIPTION	186	8278	1.00	3.16E-01	NA	0.81	0.42
RIBONUCLEOTIDE_METABOLIC_PROCESS	15	990	1.00	3.17E-01	NA	-1.56	0.12
CELLULAR_COMPONENT_DISASSEMBLY	33	1610	0.99	3.20E-01	NA	0.42	0.67
N_ACYLTRANSFERASE_ACTIVITY	24	547	0.99	3.22E-01	NA	1.87	0.06
NEGATIVE_REGULATION_OF_NUCLEOBASE__NUCLEOSIDE__NUCLEOTIDE_AND_NUCLEIC_ACID_METABOLIC_PROCESS	209	9620	0.98	3.26E-01	NA	0.75	0.45
LYSOSOME_ORGANIZATION_AND_BIOGENESIS	11	512	0.97	3.30E-01	NA	0.33	0.74
VACUOLE_ORGANIZATION_AND_BIOGENESIS	12	512	0.97	3.30E-01	NA	0.33	0.74
PHOSPHOLIPASE_C_ACTIVITY	15	846	0.97	3.33E-01	NA	-0.77	0.44
NEGATIVE_REGULATION_OF_PROTEIN_METABOLIC_PROCESS	48	959	0.95	3.41E-01	NA	0.39	0.70
SEROTONIN_RECEPTOR_ACTIVITY	11	382	0.95	3.43E-01	NA	1.13	0.26
RECEPTOR_BINDING	377	10879	0.95	3.43E-01	NA	1.29	0.20
POSITIVE_REGULATION_OF_CELL_CYCLE	16	374	0.95	3.44E-01	NA	-1.05	0.29
PROTEIN_TETRAMERIZATION	14	749	0.95	3.45E-01	NA	0.51	0.61
TRANSFERASE_ACTIVITY__TRANSFERRING_GROUPS_OTHER_THAN_AMINO_ACYL_GROUPS	49	1304	0.94	3.45E-01	NA	1.54	0.12
CORTICAL_ACTIN_CYTOSKELETON	13	987	0.94	3.45E-01	NA	-1.30	0.20
LATE_ENDOSOME	12	227	0.94	3.45E-01	NA	-0.53	0.60
CELL_MATRIX_JUNCTION	18	1138	0.94	3.47E-01	NA	-0.15	0.88
AMYLOID_PRECURSOR_PROTEIN_METABOLIC_PROCESS	10	175	0.93	3.51E-01	NA	-0.42	0.68
OLIGOSACCHARYL_TRANSFERASE_COMPLEX	10	539	0.93	3.54E-01	NA	-1.05	0.29
TRANSFERASE_ACTIVITY__TRANSFERRING_ACYL_GROUPS	59	1834	0.92	3.55E-01	NA	1.02	0.31
PURINE_RIBONUCLEOTIDE_METABOLIC_PROCESS	10	642	0.92	3.56E-01	NA	-1.28	0.20
APICAL_PLASMA_MEMBRANE	14	1136	0.92	3.57E-01	NA	0.60	0.55
TRANSMEMBRANE_RECEPTOR_PROTEIN_SERINE_THREONINE_KINASE_SIGNALING_PATHWAY	47	1532	0.92	3.59E-01	NA	-0.28	0.78
ATP_DEPENDENT_DNA_HELICASE_ACTIVITY	11	332	0.91	3.63E-01	NA	-0.03	0.98
N_ACETYLTRANSFERASE_ACTIVITY	21	528	0.91	3.65E-01	NA	1.83	0.07
TRANSCRIPTION_REPRESSOR_ACTIVITY	151	5840	0.91	3.65E-01	NA	1.13	0.26
VIRAL_INFECTIOUS_CYCLE	32	860	0.90	3.67E-01	NA	-1.77	0.08
CALCIUM_ION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	11	1604	0.90	3.67E-01	NA	-2.22	0.03
CORTICAL_CYTOSKELETON	20	1608	0.88	3.76E-01	NA	-1.64	0.10
POSITIVE_REGULATION_OF_PROTEIN_MODIFICATION_PROCESS	29	618	0.88	3.77E-01	NA	0.53	0.59
PROTEIN_AMINO_ACID_N_LINKED_GLYCOSYLATION	30	2177	0.88	3.77E-01	NA	-0.73	0.47
SH3_SH2_ADAPTOR_ACTIVITY	43	2875	0.88	3.79E-01	NA	-0.63	0.53
CELL_PROJECTION_PART	19	1042	0.87	3.82E-01	NA	1.16	0.24
BIOSYNTHETIC_PROCESS	468	16286	0.87	3.83E-01	NA	-0.82	0.41
TRNA_METABOLIC_PROCESS	19	781	0.85	3.95E-01	NA	0.11	0.91
SIALYLTRANSFERASE_ACTIVITY	10	761	0.85	3.96E-01	NA	0.53	0.60
LIGAND_GATED_CHANNEL_ACTIVITY	40	3644	0.85	3.98E-01	NA	-0.96	0.33
OUTER_MEMBRANE	25	582	0.84	4.03E-01	NA	0.12	0.90
NEGATIVE_REGULATION_OF_CELLULAR_PROTEIN_METABOLIC_PROCESS	45	909	0.83	4.07E-01	NA	0.41	0.68
RESPONSE_TO_DRUG	22	751	0.83	4.09E-01	NA	-0.05	0.96
SIGNAL_SEQUENCE_BINDING	15	543	0.83	4.09E-01	NA	-1.35	0.18
REGULATION_OF_TRANSFORMING_GROWTH_FACTOR_BETA_RECEPTOR_SIGNALING_PATHWAY	14	382	0.82	4.11E-01	NA	0.12	0.91
SMAD_BINDING	12	866	0.82	4.14E-01	NA	-0.24	0.81
GLUTAMATE_SIGNALING_PATHWAY	17	4127	0.81	4.15E-01	NA	-0.25	0.80
CLATHRIN_COATED_VESICLE	37	2243	0.81	4.15E-01	NA	1.85	0.06
REGULATION_OF_JNK_CASCADE	12	346	0.80	4.25E-01	NA	0.17	0.86
DEPHOSPHORYLATION	70	7993	0.79	4.32E-01	NA	-0.50	0.62
SMALL_CONJUGATING_PROTEIN_SPECIFIC_PROTEASE_ACTIVITY	10	257	0.78	4.36E-01	NA	-0.21	0.84
VACUOLAR_PART	13	241	0.78	4.37E-01	NA	-0.01	0.99
NEGATIVE_REGULATION_OF_PROGRAMMED_CELL_DEATH	151	5798	0.78	4.38E-01	NA	0.51	0.61
MEMBRANE_FUSION	28	1844	0.77	4.39E-01	NA	-0.23	0.82
REGULATION_OF_RESPONSE_TO_STIMULUS	59	3424	0.77	4.40E-01	NA	2.20	0.03

NEGATIVE_REGULATION_OF_APOPTOSIS	150	5797	0.77	4.41E-01	NA	0.50	0.62
PATTERN_RECOGNITION_RECEPTOR_ACTIVITY	10	313	0.77	4.42E-01	NA	-0.71	0.48
REGULATION_OF_CYTOKINE_BIOSYNTHETIC_PROCESS	38	435	0.77	4.42E-01	NA	0.16	0.87
HYDROLASE_ACTIVITY__ACTING_ON_CARBON_NITROGEN__BUT_NOT_PEPTIDE__BONDS	46	1449	0.77	4.42E-01	NA	-0.76	0.45
SECRETIN_LIKE_RECEPTOR_ACTIVITY	10	237	0.77	4.42E-01	NA	-1.42	0.16
LYSOSOMAL_MEMBRANE	10	238	0.76	4.46E-01	NA	-0.01	1.00
BASOLATERAL_PLASMA_MEMBRANE	35	1821	0.76	4.47E-01	NA	0.42	0.67
RNA_CATABOLIC_PROCESS	23	697	0.76	4.49E-01	NA	0.63	0.53
HYDROLASE_ACTIVITY__HYDROLYZING_O_GLYCOSYL_COMPOUNDS	37	1754	0.75	4.51E-01	NA	-0.22	0.83
EXTRINSIC_TO_MEMBRANE	25	848	0.75	4.51E-01	NA	0.23	0.82
TRICARBOXYLIC_ACID_CYCLE_INTERMEDIATE_METABOLIC_PROCESS	11	679	0.75	4.51E-01	NA	0.60	0.55
RESPONSE_TO_HYPOXIA	28	838	0.75	4.52E-01	NA	0.47	0.64
SENSORY_ORGAN_DEVELOPMENT	14	556	0.75	4.53E-01	NA	-0.46	0.65
RESPIRATORY_GASEOUS_EXCHANGE	13	368	0.75	4.53E-01	NA	-0.03	0.97
PROTEIN_AMINO_ACID_O_LINKED_GLYCOSYLATION	18	551	0.75	4.54E-01	NA	0.35	0.73
HETEROGENEOUS_NUCLEAR_RIBONUCLEOPROTEIN_COMPLEX	14	62	0.75	4.54E-01	NA	0.19	0.85
VACUOLAR_MEMBRANE	11	240	0.75	4.54E-01	NA	-0.03	0.97
POSITIVE_REGULATION_OF_CATALYTIC_ACTIVITY	165	5783	0.74	4.58E-01	NA	0.56	0.58
GLUCOSE_CATABOLIC_PROCESS	11	154	0.74	4.61E-01	NA	-0.35	0.73
REGULATION_OF_VIRAL_REPRODUCTION	12	243	0.73	4.65E-01	NA	-1.23	0.22
PHOSPHOLIPID_BINDING	46	2751	0.73	4.66E-01	NA	-0.20	0.84
HEART_DEVELOPMENT	36	1505	0.73	4.67E-01	NA	-0.32	0.75
VIRAL_REPRODUCTION	41	1080	0.72	4.70E-01	NA	-2.31	0.02
CHROMATIN_REMODELING_COMPLEX	17	583	0.72	4.74E-01	NA	0.80	0.42
RAS_GTPASE_BINDING	25	1307	0.71	4.75E-01	NA	-0.19	0.85
AUXILIARY_TRANSPORT_PROTEIN_ACTIVITY	26	1399	0.71	4.77E-01	NA	0.18	0.85
SPHINGOLIPID_METABOLIC_PROCESS	29	1645	0.70	4.85E-01	NA	1.07	0.29
KINETOCHORE	24	778	0.69	4.92E-01	NA	0.74	0.46
CARBOHYDRATE_BINDING	72	2185	0.68	4.94E-01	NA	0.07	0.94
ACETYLGALACTOSAMINYLTRANSFERASE_ACTIVITY	12	454	0.68	4.95E-01	NA	-0.49	0.62
RESPONSE_TO_EXTRACELLULAR_STIMULUS	32	573	0.68	4.95E-01	NA	-0.20	0.84
REPLICATION_FORK	18	324	0.68	4.98E-01	NA	0.23	0.82
PEROXISOME_ORGANIZATION_AND_BIOGENESIS	16	359	0.68	4.99E-01	NA	-1.37	0.17
CELL_DIVISION	21	1174	0.68	4.99E-01	NA	1.93	0.05
OXYGEN_AND_REACTIVE_OXYGEN_SPECIES_METABOLIC_PROCESS	20	989	0.68	4.99E-01	NA	0.17	0.86
METHYLTRANSFERASE_ACTIVITY	36	1291	0.67	5.01E-01	NA	-1.17	0.24
LIGASE_ACTIVITY	97	5186	0.67	5.02E-01	NA	0.77	0.44
DETECTION_OF_EXTERNAL_STIMULUS	23	363	0.67	5.02E-01	NA	-0.64	0.52
REGULATION_OF_ACTIN_POLYMERIZATION_AND_OR_DEPOLYMERIZATION	12	456	0.67	5.04E-01	NA	0.96	0.34
SKELETAL_MUSCLE_DEVELOPMENT	31	2394	0.67	5.06E-01	NA	0.54	0.59
TRANSFERASE_ACTIVITY__TRANSFERRING_ONE_CARBON_GROUPS	37	1303	0.66	5.10E-01	NA	-1.19	0.23
PROTEIN_CATABOLIC_PROCESS	69	3316	0.65	5.14E-01	NA	-1.12	0.26
ACTIVATION_OF_IMMUNE_RESPONSE	17	397	0.64	5.22E-01	NA	0.12	0.90
SH3_DOMAIN_BINDING	15	991	0.63	5.26E-01	NA	0.97	0.33
MACROMOLECULE_CATABOLIC_PROCESS	137	4732	0.63	5.28E-01	NA	-0.27	0.79
GROWTH_CONE	10	1317	0.63	5.31E-01	NA	-0.06	0.95
OXIDOREDUCTASE_ACTIVITY_GO_0016705	37	1434	0.63	5.31E-01	NA	0.19	0.85
NUCLEOTIDE_BINDING	223	9338	0.62	5.32E-01	NA	-2.08	0.04
SECRETION	178	7060	0.61	5.43E-01	NA	-1.06	0.29
TRANSCRIPTION__DNA_DEPENDENT	632	22121	0.61	5.44E-01	NA	1.76	0.08
PROTEIN_LOCALIZATION	214	10472	0.60	5.46E-01	NA	-0.36	0.72
CYTOPLASMIC_VESICLE_MEMBRANE	28	1082	0.60	5.51E-01	NA	0.85	0.39
CYTOPLASMIC_VESICLE_PART	28	1082	0.60	5.51E-01	NA	0.85	0.39
DRUG_BINDING	16	224	0.59	5.54E-01	NA	-0.77	0.44
CELLULAR_MACROMOLECULE_CATABOLIC_PROCESS	104	4153	0.59	5.56E-01	NA	-0.12	0.90
REGULATION_OF_HEART_CONTRACTION	25	1612	0.58	5.59E-01	NA	2.39	0.02
MITOCHONDRIAL_TRANSPORT	21	527	0.58	5.59E-01	NA	-1.14	0.25
ACTIN_FILAMENT_POLYMERIZATION	14	823	0.58	5.60E-01	NA	0.89	0.37
PROTEIN_C_TERMINUS_BINDING	73	5900	0.58	5.60E-01	NA	1.33	0.18
CERAMIDE_METABOLIC_PROCESS	11	282	0.58	5.61E-01	NA	1.31	0.19
SPHINGOID_METABOLIC_PROCESS	12	282	0.58	5.61E-01	NA	1.31	0.19
REGULATION_OF_TYROSINE_PHOSPHORYLATION_OF_STAT_PROTEIN	10	266	0.57	5.68E-01	NA	0.78	0.44
CYTOPLASMIC_VESICLE	119	5291	0.57	5.69E-01	NA	1.50	0.13
AMINE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	41	1750	0.57	5.69E-01	NA	-0.74	0.46
ALDO_KETO_REDUCTASE_ACTIVITY	11	134	0.57	5.72E-01	NA	0.52	0.60
MEDIATOR_COMPLEX	18	436	0.57	5.72E-01	NA	1.28	0.20
BIOPOLYMER_CATABOLIC_PROCESS	117	4466	0.56	5.73E-01	NA	-0.33	0.74
DNA_DAMAGE_CHECKPOINT	20	944	0.56	5.78E-01	NA	1.07	0.28
COATED_VESICLE	47	2491	0.55	5.79E-01	NA	1.74	0.08
ESTABLISHMENT_OF_PROTEIN_LOCALIZATION	190	9448	0.55	5.83E-01	NA	-1.00	0.32
CARBOXYLIC_ACID_TRANSPORT	41	1123	0.55	5.84E-01	NA	0.85	0.39
REGULATION_OF_JAK_STAT_CASCADE	12	321	0.54	5.86E-01	NA	0.90	0.37
FEMALE_GAMETE_GENERATION	17	873	0.54	5.87E-01	NA	1.22	0.22
POST_GOLGI_VESICLE_MEDIATED_TRANSPORT	14	516	0.54	5.88E-01	NA	-0.20	0.84
INNATE_IMMUNE_RESPONSE	23	628	0.54	5.91E-01	NA	0.62	0.54
VITAMIN_METABOLIC_PROCESS	17	512	0.53	5.96E-01	NA	-0.59	0.56
INTERMEDIATE_FILAMENT	24	400	0.53	5.97E-01	NA	-0.68	0.50
INTERMEDIATE_FILAMENT_CYTOSKELETON	24	400	0.53	5.97E-01	NA	-0.68	0.50
DI___TRI_VALENT_INORGANIC_CATION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	22	1798	0.53	5.97E-01	NA	-1.76	0.08
SECRETORY_GRANULE	18	1067	0.52	6.02E-01	NA	0.98	0.33
PHOSPHOLIPASE_C_ACTIVATION	14	473	0.52	6.05E-01	NA	1.18	0.24
PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION	43	2601	0.51	6.10E-01	NA	0.19	0.85
LIPOPROTEIN_BIOSYNTHETIC_PROCESS	26	455	0.51	6.10E-01	NA	-0.26	0.80
DNA_HELICASE_ACTIVITY	25	981	0.51	6.11E-01	NA	0.92	0.36
REGULATION_OF_TRANSCRIPTION__DNA_DEPENDENT	459	17702	0.51	6.13E-01	NA	2.05	0.04
NUCLEAR_UBIQUITIN_LIGASE_COMPLEX	11	468	0.50	6.14E-01	NA	-1.04	0.30
STRIATED_MUSCLE_DEVELOPMENT	40	3131	0.50	6.16E-01	NA	0.14	0.89
LIPID_BINDING	86	3472	0.49	6.22E-01	NA	-0.51	0.61
PURINE_RIBONUCLEOTIDE_BINDING	204	8813	0.49	6.23E-01	NA	-2.17	0.03
AMINOPEPTIDASE_ACTIVITY	14	1047	0.49	6.25E-01	NA	1.16	0.25
SPECIFIC_TRANSCRIPTIONAL_REPRESSOR_ACTIVITY	12	548	0.49	6.26E-01	NA	1.36	0.17
PROTEIN_FOLDING	58	1479	0.49	6.26E-01	NA	-0.26	0.79
CELLULAR_RESPONSE_TO_STRESS	10	219	0.48	6.29E-01	NA	-0.11	0.92
HOMEOSTASIS_OF_NUMBER_OF_CELLS	20	525	0.48	6.30E-01	NA	1.05	0.30
HORMONE_ACTIVITY	44	259	0.48	6.34E-01	NA	0.08	0.94
NON_MEMBRANE_SPANNING_PROTEIN_TYROSINE_KINASE_ACTIVITY	11	714	0.47	6.37E-01	NA	-0.66	0.51
CYTOKINESIS	19	507	0.47	6.39E-01	NA	1.00	0.32
ORGANIC_ACID_TRANSPORT	42	1131	0.46	6.44E-01	NA	0.88	0.38
VITAMIN_BINDING	12	449	0.46	6.45E-01	NA	1.13	0.26
PURINE_NUCLEOTIDE_BINDING	210	8937	0.46	6.48E-01	NA	-2.08	0.04
POSITIVE_REGULATION_OF_DNA_METABOLIC_PROCESS	11	217	0.45	6.52E-01	NA	0.68	0.50
RNA_POLYMERASE_II_TRANSCRIPTION_MEDIATOR_ACTIVITY	13	422	0.45	6.53E-01	NA	-0.16	0.87
POSITIVE_REGULATION_OF_PEPTIDYL_TYROSINE_PHOSPHORYLATION	13	233	0.45	6.54E-01	NA	0.78	0.43
REGULATION_OF_RNA_METABOLIC_PROCESS	469	17876	0.45	6.54E-01	NA	2.01	0.04
TIGHT_JUNCTION	31	1103	0.44	6.57E-01	NA	1.99	0.05
STEROL_BINDING	11	255	0.44	6.58E-01	NA	-0.78	0.44
CELLULAR_LOCALIZATION	372	17113	0.44	6.59E-01	NA	-1.09	0.28

RNA_BIOSYNTHETIC_PROCESS	634	22257	0.44	6.59E-01	NA	1.69	0.09
POSITIVE_REGULATION_OF_IMMUNE_SYSTEM_PROCESS	51	1401	0.44	6.62E-01	NA	0.13	0.90
DNA_DEPENDENT_ATPASE_ACTIVITY	22	780	0.44	6.64E-01	NA	0.28	0.78
POSITIVE_REGULATION_OF_MAP_KINASE_ACTIVITY	47	1688	0.43	6.66E-01	NA	0.06	0.95
REGULATION_OF_PEPTIDYL_TYROSINE_PHOSPHORYLATION	18	507	0.43	6.68E-01	NA	0.85	0.40
NUCLEAR_HORMONE_RECEPTOR_BINDING	28	936	0.43	6.68E-01	NA	0.30	0.77
PROTEIN_TRANSPORT	157	8200	0.43	6.68E-01	NA	-1.43	0.15
REGULATION_OF_CHROMOSOME_ORGANIZATION_AND_BIOGENESIS	10	216	0.42	6.74E-01	NA	-0.09	0.92
TYROSINE_PHOSPHORYLATION_OF_STAT_PROTEIN	13	327	0.41	6.80E-01	NA	0.89	0.37
REGULATION_OF_MULTICELLULAR_ORGANISMAL_PROCESS	150	6392	0.41	6.81E-01	NA	1.18	0.24
POSITIVE_REGULATION_OF_IMMUNE_RESPONSE	29	858	0.41	6.81E-01	NA	0.77	0.44
SYNAPTOGENESIS	18	1343	0.41	6.82E-01	NA	0.10	0.92
PROTEIN_EXPORT_FROM_NUCLEUS	12	454	0.40	6.89E-01	NA	0.68	0.50
POSITIVE_REGULATION_OF_TRANSPORT	23	692	0.40	6.91E-01	NA	0.83	0.41
PROTEASOME_COMPLEX	23	305	0.40	6.93E-01	NA	0.13	0.89
EXONUCLEASE_ACTIVITY	19	555	0.39	6.94E-01	NA	0.69	0.49
CELL_CYCLE_PROCESS	192	6965	0.39	6.94E-01	NA	0.94	0.34
REGULATION_OF_MITOSIS	41	968	0.39	6.94E-01	NA	0.14	0.89
UBIQUITIN_CYCLE	48	2620	0.39	6.99E-01	NA	-0.25	0.80
VESICLE	124	5406	0.39	7.00E-01	NA	1.25	0.21
REGULATION_OF_IMMUNE_SYSTEM_PROCESS	67	1601	0.39	7.00E-01	NA	0.45	0.66
PROTEIN_AMINO_ACID_LIPIDATION	24	453	0.38	7.01E-01	NA	-0.34	0.74
ATPASE_ACTIVITY__COUPLED_TO_TRANSMEMBRANE_MOVEMENT_OF_IONS	24	513	0.38	7.01E-01	NA	-0.35	0.72
CHAPERONE_BINDING	12	537	0.38	7.01E-01	NA	0.65	0.52
PROTEIN_POLYMERIZATION	18	1046	0.37	7.10E-01	NA	0.67	0.50
LIPOPROTEIN_METABOLIC_PROCESS	33	505	0.37	7.10E-01	NA	0.73	0.47
CHANNEL_REGULATOR_ACTIVITY	24	1253	0.37	7.11E-01	NA	0.54	0.59
POSITIVE_REGULATION_OF_CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS	10	542	0.35	7.23E-01	NA	0.36	0.72
STRIATED_MUSCLE_CONTRACTION_GO_0006941	14	851	0.35	7.28E-01	NA	0.29	0.77
TRANSITION_METAL_ION_BINDING	110	3137	0.34	7.35E-01	NA	-2.08	0.04
HYDROLASE_ACTIVITY__ACTING_ON_GLYCOSYL_BONDS	47	2047	0.33	7.39E-01	NA	-0.28	0.78
PEPTIDYL_TYROSINE_MODIFICATION	29	1192	0.33	7.40E-01	NA	0.71	0.48
INORGANIC_CATION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	58	2688	0.33	7.42E-01	NA	-1.72	0.09
REGULATION_OF_NEURON_APOPTOSIS	12	363	0.33	7.44E-01	NA	1.93	0.05
AMINE_BIOSYNTHETIC_PROCESS	15	498	0.33	7.45E-01	NA	0.62	0.54
REPRODUCTIVE_PROCESS	161	3836	0.32	7.48E-01	NA	1.21	0.22
DEVELOPMENTAL_GROWTH	11	803	0.32	7.53E-01	NA	2.97	0.00
ACETYLTRANSFERASE_ACTIVITY	25	664	0.31	7.55E-01	NA	1.39	0.17
POSITIVE_REGULATION_OF_MULTICELLULAR_ORGANISMAL_PROCESS	66	1622	0.31	7.56E-01	NA	-0.33	0.74
PROTEOLYSIS	190	7480	0.31	7.56E-01	NA	0.21	0.83
NEGATIVE_REGULATION_OF_CELL_MIGRATION	15	415	0.30	7.61E-01	NA	-0.55	0.58
PROTEIN_UBIQUITINATION	40	2462	0.30	7.63E-01	NA	0.28	0.78
MAPKKK_CASCADE_GO_0000165	104	3583	0.29	7.69E-01	NA	-1.10	0.27
SITE_OF_POLARIZED_GROWTH	11	1346	0.29	7.70E-01	NA	0.14	0.89
REGULATION_OF_SECRETION	40	1184	0.28	7.76E-01	NA	-0.14	0.88
CELLULAR_PROTEIN_CATABOLIC_PROCESS	58	3083	0.28	7.78E-01	NA	-0.43	0.67
REGULATION_OF_HORMONE_SECRETION	14	175	0.27	7.83E-01	NA	0.72	0.47
CYTOKINE_METABOLIC_PROCESS	42	521	0.26	7.91E-01	NA	0.24	0.81
OXIDOREDUCTASE_ACTIVITY__ACTING_ON_THE_ALDEHYDE_OR_OXO_GROUP_OF_DONORS	22	805	0.26	7.95E-01	NA	0.69	0.49
PROTEIN_TARGETING_TO_MITOCHONDRION	11	258	0.26	7.96E-01	NA	-0.40	0.69
SECRETION_BY_CELL	116	5628	0.26	7.96E-01	NA	-1.78	0.08
CHROMOSOME__PERICENTRIC_REGION	30	933	0.25	8.01E-01	NA	0.58	0.56
POSITIVE_REGULATION_OF_PROTEIN_METABOLIC_PROCESS	75	1647	0.25	8.06E-01	NA	0.15	0.88
VIRAL_GENOME_REPLICATION	21	489	0.25	8.06E-01	NA	-1.69	0.09
CHROMOSOME	122	2820	0.23	8.15E-01	NA	0.92	0.36
TUBULIN_BINDING	47	2445	0.23	8.17E-01	NA	-1.91	0.06
DEACETYLASE_ACTIVITY	11	84	0.23	8.22E-01	NA	-0.44	0.66
NEGATIVE_REGULATION_OF_CATALYTIC_ACTIVITY	69	3919	0.22	8.23E-01	NA	0.04	0.97
CYTOKINE_BIOSYNTHETIC_PROCESS	41	511	0.22	8.26E-01	NA	0.22	0.82
POLYSACCHARIDE_BINDING	36	1734	0.22	8.27E-01	NA	0.06	0.96
SECRETORY_PATHWAY	84	4603	0.22	8.29E-01	NA	-1.80	0.07
CYTOPLASMIC_MEMBRANE_BOUND_VESICLE	115	5061	0.21	8.31E-01	NA	1.37	0.17
MANNOSYLTRANSFERASE_ACTIVITY	10	78	0.21	8.33E-01	NA	0.74	0.46
CALCIUM_INDEPENDENT_CELL_CELL_ADHESION	22	316	0.21	8.35E-01	NA	0.46	0.64
NUCLEAR_ENVELOPE_ENDOPLASMIC_RETICULUM_NETWORK	94	3276	0.21	8.35E-01	NA	-0.90	0.37
HORMONE_RECEPTOR_BINDING	29	939	0.21	8.37E-01	NA	0.35	0.73
REGULATION_OF_JNK_ACTIVITY	20	684	0.21	8.37E-01	NA	-1.62	0.11
POSITIVE_REGULATION_OF_BINDING	28	785	0.20	8.43E-01	NA	-0.44	0.66
PEPTIDASE_ACTIVITY	176	6035	0.20	8.45E-01	NA	1.19	0.23
CARBOHYDRATE_CATABOLIC_PROCESS	24	374	0.19	8.51E-01	NA	0.09	0.93
MITOCHONDRIAL_OUTER_MEMBRANE	18	383	0.18	8.55E-01	NA	-0.46	0.64
MITOTIC_SPINDLE_ORGANIZATION_AND_BIOGENESIS	10	195	0.18	8.56E-01	NA	2.67	0.01
SPINDLE_ORGANIZATION_AND_BIOGENESIS	11	195	0.18	8.56E-01	NA	2.67	0.01
CELL_SURFACE	79	4658	0.18	8.57E-01	NA	-0.44	0.66
ENDOPLASMIC_RETICULUM_MEMBRANE	85	3099	0.17	8.62E-01	NA	-0.90	0.37
REGULATION_OF_PROTEIN_SECRETION	22	653	0.17	8.68E-01	NA	-0.21	0.84
NUCLEOTIDYLTRANSFERASE_ACTIVITY	47	1149	0.16	8.69E-01	NA	0.29	0.77
NOTCH_SIGNALING_PATHWAY	12	1021	0.16	8.70E-01	NA	-0.87	0.39
CHROMATIN_BINDING	32	525	0.16	8.72E-01	NA	0.68	0.50
CARBOXYLIC_ACID_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	44	1678	0.16	8.75E-01	NA	0.27	0.79
TRANSPORT_VESICLE	33	744	0.16	8.76E-01	NA	0.89	0.38
ACTIN_FILAMENT_BASED_MOVEMENT	10	585	0.16	8.76E-01	NA	0.21	0.83
MICROTUBULE_ASSOCIATED_COMPLEX	47	1888	0.16	8.77E-01	NA	-0.91	0.36
LEUKOCYTE_MIGRATION	16	739	0.15	8.78E-01	NA	-0.12	0.91
COENZYME_BIOSYNTHETIC_PROCESS	10	472	0.15	8.79E-01	NA	0.79	0.43
REGULATION_OF_PROTEIN_POLYMERIZATION	11	485	0.15	8.80E-01	NA	0.40	0.69
CELLULAR_RESPONSE_TO_EXTRACELLULAR_STIMULUS	12	108	0.14	8.85E-01	NA	-1.05	0.29
REGULATION_OF_MAP_KINASE_ACTIVITY	67	2316	0.14	8.86E-01	NA	0.08	0.94
PROTEIN_SECRETION	32	1025	0.14	8.89E-01	NA	-0.34	0.73
ANION_CHANNEL_ACTIVITY	20	571	0.14	8.91E-01	NA	-1.25	0.21
CHLORIDE_CHANNEL_ACTIVITY	19	571	0.14	8.91E-01	NA	-1.25	0.21
CATION_TRANSPORTING_ATPASE_ACTIVITY	11	263	0.13	8.94E-01	NA	-0.61	0.54
MONOCARBOXYLIC_ACID_TRANSPORT	10	197	0.13	8.95E-01	NA	0.70	0.48
NUCLEAR_REPLICATION_FORK	10	158	0.13	8.99E-01	NA	0.44	0.66
REGULATION_OF_CELL_CELL_ADHESION	10	222	0.13	9.00E-01	NA	-1.09	0.28
ORGANELLE_OUTER_MEMBRANE	24	477	0.11	9.11E-01	NA	-0.59	0.56
INTERLEUKIN_1_SECRETION	10	262	0.10	9.17E-01	NA	0.48	0.63
INTERCALATED_DISC	9	2018	0.10	9.19E-01	NA	-1.17	0.24
NEGATIVE_REGULATION_OF_CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS	11	460	0.10	9.19E-01	NA	-0.57	0.57
SENSORY_PERCEPTION_OF_CHEMICAL_STIMULUS	20	445	0.09	9.27E-01	NA	0.82	0.41
ORGANIC_ACID_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	45	1686	0.08	9.33E-01	NA	0.30	0.76
TRIACYLGLYCEROL_METABOLIC_PROCESS	10	139	0.08	9.33E-01	NA	-2.06	0.04
CELLULAR_CARBOHYDRATE_CATABOLIC_PROCESS	23	373	0.08	9.33E-01	NA	0.03	0.97
CELL_CYCLE_PHASE	169	6397	0.08	9.33E-01	NA	0.89	0.37
VESICLE_MEMBRANE	30	1129	0.08	9.33E-01	NA	0.51	0.61
NUCLEAR_CHROMOSOME_PART	34	498	0.08	9.38E-01	NA	1.20	0.23
COATED_MEMBRANE	17	433	0.08	9.39E-01	NA	1.60	0.11

MEMBRANE_COAT	17	433	0.08	9.39E-01	NA	1.60	0.11
PROTEIN_DEACETYLASE_ACTIVITY	10	53	0.08	9.40E-01	NA	-0.30	0.77
GLYCOSAMINOGLYCAN_BINDING	34	1680	0.07	9.41E-01	NA	0.12	0.90
PROTEIN_POLYUBIQUITINATION	11	285	0.07	9.41E-01	NA	0.53	0.59
HOMOPHILIC_CELL_ADHESION	16	1365	0.06	9.49E-01	NA	-0.45	0.65
RESPONSE_TO_NUTRIENT	16	408	0.06	9.50E-01	NA	0.47	0.64
DENDRITE	16	688	0.06	9.51E-01	NA	-0.32	0.75
REGULATION_OF_AXONOGENESIS	10	1373	0.06	9.54E-01	NA	-0.54	0.59
INTRACELLULAR_RECEPTOR_MEDIATED_SIGNALING_PATHWAY	21	576	0.05	9.61E-01	NA	1.14	0.25
SODIUM_ION_TRANSPORT	22	1187	0.05	9.63E-01	NA	-1.92	0.05
CHROMOSOME_CONDENSATION	10	118	0.04	9.69E-01	NA	0.11	0.92
REGULATION_OF_NEUROGENESIS	14	1534	0.04	9.71E-01	NA	-0.61	0.54
RESPONSE_TO_NUTRIENT_LEVELS	28	551	0.04	9.72E-01	NA	-0.42	0.67
TRANSLATION	180	3008	0.04	9.72E-01	NA	0.02	0.98
POSITIVE_REGULATION_OF_DNA_BINDING	26	775	0.04	9.72E-01	NA	-0.46	0.65
POSITIVE_REGULATION_OF_TRANSCRIPTION_FACTOR_ACTIVITY	24	772	0.04	9.72E-01	NA	-0.46	0.65
VOLTAGE_GATED_SODIUM_CHANNEL_ACTIVITY	11	577	0.03	9.72E-01	NA	-0.90	0.37
SYNAPSE_ORGANIZATION_AND_BIOGENESIS	23	1842	0.03	9.73E-01	NA	0.04	0.97
SPERMATID_DIFFERENTIATION	11	262	0.03	9.73E-01	NA	-1.64	0.10
KINASE_BINDING	70	2539	0.03	9.75E-01	NA	-0.02	0.98
ADENYLATE_CYCLASE_ACTIVATION	19	354	0.03	9.75E-01	NA	-0.47	0.64
STEROID_HORMONE_RECEPTOR_SIGNALING_PATHWAY	20	575	0.03	9.80E-01	NA	1.14	0.25
STEROID_BIOSYNTHETIC_PROCESS	23	745	0.02	9.83E-01	NA	0.25	0.80
COATED_VESICLE_MEMBRANE	17	424	0.01	9.89E-01	NA	1.48	0.14
VESICLE_COAT	16	424	0.01	9.89E-01	NA	1.48	0.14
POLYSACCHARIDE_METABOLIC_PROCESS	18	423	0.01	9.90E-01	NA	-0.28	0.78
REGIONALIZATION	15	364	0.01	9.95E-01	NA	0.11	0.91
OXIDOREDUCTASE_ACTIVITY__ACTING_ON_PEROXIDE_AS_ACCEPTOR	12	284	0.01	9.95E-01	NA	0.31	0.76
LIPID_KINASE_ACTIVITY	12	378	0.00	9.97E-01	NA	-0.27	0.78
DNA_INTEGRITY_CHECKPOINT	24	989	0.00	9.99E-01	NA	1.10	0.27
L_AMINO_ACID_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	17	992	0.00	9.96E-01	NA	-0.56	0.57
CONDENSED_CHROMOSOME	33	1081	-0.01	9.92E-01	NA	0.49	0.62
RESPONSE_TO_TOXIN	10	257	-0.01	9.91E-01	NA	-2.93	0.00
AMINO_ACID_TRANSPORT	26	805	-0.01	9.89E-01	NA	0.91	0.36
POSITIVE_REGULATION_OF_JNK_ACTIVITY	18	626	-0.02	9.88E-01	NA	-1.41	0.16
CHROMOSOME_ORGANIZATION_AND_BIOGENESIS	124	2989	-0.02	9.87E-01	NA	1.23	0.22
GOLGI_ASSOCIATED_VESICLE	29	881	-0.02	9.85E-01	NA	1.04	0.30
ATPASE_ACTIVITY__COUPLED_TO_TRANSMEMBRANE_MOVEMENT_OF_IONS__PHOSPHORYLATIVE_MECHANISM	20	429	-0.02	9.84E-01	NA	-0.42	0.68
LIPID_HOMEOSTASIS	16	324	-0.02	9.83E-01	NA	-0.14	0.89
NUCLEOLAR_PART	18	204	-0.02	9.82E-01	NA	-1.90	0.06
HYDROLASE_ACTIVITY__ACTING_ON_ACID_ANHYDRIDES	228	8168	-0.03	9.78E-01	NA	0.68	0.49
MEMBRANE_BOUND_VESICLE	117	5108	-0.03	9.74E-01	NA	1.20	0.23
NUCLEOSOME_ASSEMBLY	10	202	-0.04	9.70E-01	NA	1.14	0.25
ANTIOXIDANT_ACTIVITY	18	414	-0.04	9.68E-01	NA	-0.16	0.87
CELLULAR_POLYSACCHARIDE_METABOLIC_PROCESS	16	410	-0.05	9.62E-01	NA	-0.12	0.90
GLUTATHIONE_TRANSFERASE_ACTIVITY	15	146	-0.05	9.59E-01	NA	-0.04	0.97
ORGANIC_ANION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	10	517	-0.06	9.55E-01	NA	1.77	0.08
REGULATION_OF_IMMUNE_RESPONSE	33	1003	-0.06	9.55E-01	NA	1.02	0.31
EPITHELIAL_TO_MESENCHYMAL_TRANSITION	10	270	-0.08	9.37E-01	NA	0.10	0.92
M_PHASE_OF_MITOTIC_CELL_CYCLE	84	2170	-0.10	9.24E-01	NA	0.42	0.67
RESPONSE_TO_ABIOTIC_STIMULUS	89	2234	-0.10	9.21E-01	NA	-1.20	0.23
NUCLEAR_CHROMOSOME	54	1022	-0.10	9.21E-01	NA	0.38	0.70
MICROTUBULE_ORGANIZING_CENTER	67	2739	-0.10	9.18E-01	NA	-1.00	0.32
MONOVALENT_INORGANIC_CATION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	34	1216	-0.10	9.17E-01	NA	-1.25	0.21
MAINTENANCE_OF_PROTEIN_LOCALIZATION	13	484	-0.11	9.15E-01	NA	0.29	0.77
OXIDOREDUCTASE_ACTIVITY__ACTING_ON_THE_CH_CH_GROUP_OF_DONORS	23	421	-0.11	9.14E-01	NA	1.48	0.14
SYNAPTIC_VESICLE	15	1370	-0.11	9.14E-01	NA	0.67	0.50
SPERMATID_DEVELOPMENT	10	114	-0.11	9.14E-01	NA	-2.52	0.01
SENSORY_PERCEPTION_OF_TASTE	11	20	-0.11	9.14E-01	NA	1.69	0.09
ACTIVATION_OF_MAPK_ACTIVITY	41	1320	-0.11	9.11E-01	NA	-0.28	0.78
RECEPTOR_SIGNALING_PROTEIN_SERINE_THREONINE_KINASE_ACTIVITY	34	1648	-0.11	9.11E-01	NA	0.23	0.82
ENDOSOME	67	2175	-0.11	9.10E-01	NA	0.06	0.95
ELECTRON_CARRIER_ACTIVITY	78	2237	-0.11	9.10E-01	NA	0.65	0.52
ENERGY_DERIVATION_BY_OXIDATION_OF_ORGANIC_COMPOUNDS	37	1257	-0.12	9.06E-01	NA	0.10	0.92
APOPTOTIC_PROGRAM	60	1273	-0.12	9.02E-01	NA	2.01	0.04
PROTON_TRANSPORTING_TWO_SECTOR_ATPASE_COMPLEX	15	183	-0.14	8.87E-01	NA	0.20	0.84
LEUKOCYTE_CHEMOTAXIS	13	688	-0.15	8.84E-01	NA	-0.06	0.95
BILE_ACID_METABOLIC_PROCESS	11	391	-0.15	8.80E-01	NA	0.94	0.35
PEPTIDE_METABOLIC_PROCESS	10	479	-0.16	8.72E-01	NA	-2.43	0.02
NUCLEAR_IMPORT	50	2095	-0.17	8.69E-01	NA	-1.78	0.07
INTERLEUKIN_8_BIOSYNTHETIC_PROCESS	10	97	-0.17	8.63E-01	NA	-0.06	0.95
CELLULAR_RESPONSE_TO_STIMULUS	19	329	-0.18	8.57E-01	NA	-0.72	0.47
NEGATIVE_REGULATION_OF_SIGNAL_TRANSDUCTION	37	738	-0.19	8.47E-01	NA	0.19	0.85
CHROMATIN_ASSEMBLY	16	273	-0.19	8.46E-01	NA	0.37	0.71
POSITIVE_REGULATION_OF_DEFENSE_RESPONSE	10	305	-0.20	8.39E-01	NA	0.34	0.73
DNA_CATABOLIC_PROCESS	23	415	-0.21	8.37E-01	NA	1.40	0.16
NUCLEAR_BODY	33	840	-0.21	8.34E-01	NA	0.81	0.42
MITOSIS	81	2076	-0.22	8.25E-01	NA	0.48	0.63
POSITIVE_REGULATION_OF_TRANSFERASE_ACTIVITY	86	3522	-0.22	8.25E-01	NA	0.51	0.61
PROTEIN_TARGETING	109	4549	-0.22	8.24E-01	NA	-1.83	0.07
NEGATIVE_REGULATION_OF_TRANSCRIPTION__DNA_DEPENDENT	128	4917	-0.22	8.22E-01	NA	1.42	0.15
RESPONSE_TO_WOUNDING	190	5204	-0.23	8.20E-01	NA	1.13	0.26
PEPTIDYL_TYROSINE_PHOSPHORYLATION	27	1039	-0.23	8.20E-01	NA	0.92	0.36
INTERLEUKIN_2_PRODUCTION	11	260	-0.23	8.15E-01	NA	-0.66	0.51
MAINTENANCE_OF_CELLULAR_PROTEIN_LOCALIZATION	11	464	-0.24	8.11E-01	NA	0.46	0.65
MICROBODY_MEMBRANE	13	304	-0.24	8.07E-01	NA	-0.30	0.77
PEROXISOMAL_MEMBRANE	13	304	-0.24	8.07E-01	NA	-0.30	0.77
RESPONSE_TO_VIRUS	50	457	-0.25	8.03E-01	NA	0.51	0.61
ESTABLISHMENT_OF_CELLULAR_LOCALIZATION	354	14248	-0.25	8.00E-01	NA	-2.36	0.02
REGULATION_OF_MEMBRANE_POTENTIAL	15	696	-0.26	7.92E-01	NA	0.74	0.46
LEUKOCYTE_DIFFERENTIATION	38	1438	-0.27	7.89E-01	NA	1.23	0.22
CELL_CYCLE_GO_0007049	314	9677	-0.28	7.83E-01	NA	0.08	0.94
PROTEIN_RNA_COMPLEX_ASSEMBLY	67	1281	-0.29	7.69E-01	NA	-2.22	0.03
CYTOPLASMIC_PART	1380	54143	-0.30	7.67E-01	NA	-0.79	0.43
PYROPHOSPHATASE_ACTIVITY	226	7943	-0.30	7.61E-01	NA	0.51	0.61
REGULATION_OF_ANGIOGENESIS	26	874	-0.30	7.61E-01	NA	0.32	0.75
HORMONE_METABOLIC_PROCESS	32	856	-0.31	7.59E-01	NA	0.05	0.96
RESPONSE_TO_IONIZING_RADIATION	10	454	-0.31	7.55E-01	NA	-1.31	0.19
DEOXYRIBONUCLEASE_ACTIVITY	22	299	-0.32	7.46E-01	NA	1.01	0.31
G_PROTEIN_COUPLED_RECEPTOR_BINDING	54	447	-0.32	7.45E-01	NA	-0.09	0.92
CARBOXYLESTERASE_ACTIVITY	32	732	-0.33	7.42E-01	NA	0.00	1.00
MAP_KINASE_KINASE_KINASE_ACTIVITY	10	654	-0.33	7.38E-01	NA	0.08	0.93
REGULATION_OF_DEFENSE_RESPONSE	20	395	-0.34	7.35E-01	NA	0.33	0.74
POSITIVE_REGULATION_OF_CYTOKINE_PRODUCTION	15	319	-0.34	7.34E-01	NA	-0.62	0.53
STEROID_BINDING	18	452	-0.35	7.29E-01	NA	-0.51	0.61
ZINC_ION_BINDING	89	2505	-0.35	7.27E-01	NA	-2.64	0.01
NEGATIVE_REGULATION_OF_RNA_METABOLIC_PROCESS	130	4927	-0.36	7.20E-01	NA	1.35	0.18

CHROMATIN_ASSEMBLY_OR_DISASSEMBLY	26	444	-0.37	7.15E-01	NA	-0.34	0.73
GLUCAN_METABOLIC_PROCESS	10	347	-0.37	7.12E-01	NA	0.04	0.97
REGULATION_OF_DNA_REPLICATION	20	609	-0.37	7.08E-01	NA	1.71	0.09
POSITIVE_REGULATION_OF_CELLULAR_PROTEIN_METABOLIC_PROCESS	73	1489	-0.38	7.05E-01	NA	1.14	0.25
OXIDOREDUCTASE_ACTIVITY_GO_0016706	10	511	-0.38	7.04E-01	NA	0.98	0.33
CELL_PROLIFERATION_GO_0008283	511	17974	-0.38	7.03E-01	NA	1.37	0.17
PROTEIN_KINASE_BINDING	62	2117	-0.38	7.02E-01	NA	-1.12	0.26
HUMORAL_IMMUNE_RESPONSE	31	923	-0.40	6.92E-01	NA	0.46	0.64
ENDOMEMBRANE_SYSTEM	220	8609	-0.40	6.91E-01	NA	-0.86	0.39
POSITIVE_REGULATION_OF_SIGNAL_TRANSDUCTION	126	5065	-0.41	6.83E-01	NA	0.35	0.72
VESICULAR_FRACTION	44	975	-0.41	6.81E-01	NA	0.30	0.77
S_ADENOSYLMETHIONINE_DEPENDENT_METHYLTRANSFERASE_ACTIVITY	23	773	-0.41	6.80E-01	NA	0.01	0.99
UNFOLDED_PROTEIN_BINDING	42	580	-0.42	6.72E-01	NA	0.65	0.52
ISOMERASE_ACTIVITY	36	896	-0.43	6.68E-01	NA	-0.63	0.53
MAINTENANCE_OF_CELLULAR_LOCALIZATION	12	476	-0.43	6.67E-01	NA	0.26	0.79
TRANSFERASE_ACTIVITY__TRANSFERRING_PHOSPHORUS_CONTAINING_GROUPS	422	23272	-0.44	6.61E-01	NA	-1.76	0.08
TRANSCRIPTION_FACTOR_TFIID_COMPLEX	14	228	-0.44	6.60E-01	NA	-1.56	0.12
TRANSCRIPTION_ELONGATION_REGULATOR_ACTIVITY	12	377	-0.44	6.59E-01	NA	-0.70	0.48
ESTABLISHMENT_AND_OR_MAINTENANCE_OF_CELL_POLARITY	19	1421	-0.44	6.58E-01	NA	0.84	0.40
PROTEIN_IMPORT_INTO_NUCLEUS	48	2075	-0.44	6.56E-01	NA	-1.68	0.09
REGULATION_OF_LYMPHOCYTE_ACTIVATION	35	802	-0.45	6.55E-01	NA	0.05	0.96
COFACTOR_CATABOLIC_PROCESS	10	319	-0.45	6.51E-01	NA	-2.09	0.04
CHEMOKINE_ACTIVITY	42	258	-0.46	6.49E-01	NA	-0.45	0.65
CHEMOKINE_RECEPTOR_BINDING	43	258	-0.46	6.49E-01	NA	-0.45	0.65
NUCLEAR_SPECK	11	329	-0.46	6.45E-01	NA	-0.43	0.67
SH2_DOMAIN_BINDING	15	585	-0.47	6.38E-01	NA	-0.86	0.39
PATTERN_BINDING	45	1858	-0.48	6.35E-01	NA	0.45	0.65
NEGATIVE_REGULATION_OF_IMMUNE_SYSTEM_PROCESS	14	345	-0.48	6.30E-01	NA	1.05	0.29
EMBRYO_IMPLANTATION	13	249	-0.48	6.29E-01	NA	1.39	0.16
MICROSOME	42	843	-0.48	6.29E-01	NA	0.99	0.32
ISOPRENOID_METABOLIC_PROCESS	12	366	-0.49	6.27E-01	NA	-0.52	0.60
POSITIVE_REGULATION_OF_SECRETION	20	449	-0.49	6.24E-01	NA	-0.23	0.82
GENERAL_RNA_POLYMERASE_II_TRANSCRIPTION_FACTOR_ACTIVITY	32	749	-0.50	6.15E-01	NA	-1.10	0.27
SINGLE_STRANDED_DNA_BINDING	34	925	-0.51	6.11E-01	NA	1.80	0.07
THIOLESTER_HYDROLASE_ACTIVITY	16	441	-0.51	6.09E-01	NA	-0.95	0.34
RIBONUCLEOPROTEIN_BINDING	12	173	-0.52	6.04E-01	NA	0.99	0.32
REGULATION_OF_BINDING	58	1397	-0.52	6.03E-01	NA	-0.54	0.59
PROTEIN_PHOSPHATASE_BINDING	10	333	-0.52	6.01E-01	NA	0.62	0.53
SOLUBLE_FRACTION	160	4994	-0.53	5.98E-01	NA	0.32	0.75
FATTY_ACID_OXIDATION	18	427	-0.53	5.96E-01	NA	1.34	0.18
REGULATION_OF_G_PROTEIN_COUPLED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY	23	984	-0.55	5.85E-01	NA	0.34	0.73
POSITIVE_REGULATION_OF_CYTOKINE_BIOSYNTHETIC_PROCESS	25	225	-0.55	5.83E-01	NA	-0.24	0.81
REGULATION_OF_CYTOKINE_PRODUCTION	25	611	-0.55	5.82E-01	NA	-0.63	0.53
SMALL_CONJUGATING_PROTEIN_BINDING	12	223	-0.55	5.79E-01	NA	-1.08	0.28
B_CELL_DIFFERENTIATION	12	761	-0.55	5.79E-01	NA	0.36	0.72
MICROTUBULE_ORGANIZING_CENTER_PART	19	492	-0.57	5.72E-01	NA	-1.27	0.20
AMINE_BINDING	23	767	-0.59	5.55E-01	NA	0.76	0.45
I_KAPPAB_KINASE_NF_KAPPAB_CASCADE	114	2513	-0.60	5.49E-01	NA	-1.33	0.18
PROTEIN_COMPLEX_BINDING	54	2856	-0.60	5.48E-01	NA	1.12	0.26
MULTI_ORGANISM_PROCESS	163	2897	-0.61	5.39E-01	NA	-0.24	0.81
ENDOPEPTIDASE_ACTIVITY	117	3209	-0.62	5.38E-01	NA	0.77	0.44
MACROMOLECULE_LOCALIZATION	236	11207	-0.62	5.37E-01	NA	-0.80	0.42
CELLULAR_CATION_HOMEOSTASIS	106	2290	-0.62	5.37E-01	NA	-0.34	0.73
CELLULAR_BIOSYNTHETIC_PROCESS	319	9636	-0.62	5.36E-01	NA	-0.47	0.64
MEMBRANE_LIPID_BIOSYNTHETIC_PROCESS	49	1758	-0.62	5.36E-01	NA	-0.81	0.42
INTRA_GOLGI_VESICLE_MEDIATED_TRANSPORT	12	320	-0.62	5.35E-01	NA	0.02	0.98
CONDENSED_NUCLEAR_CHROMOSOME	18	556	-0.62	5.33E-01	NA	0.01	0.99
CARBOHYDRATE_KINASE_ACTIVITY	15	415	-0.63	5.30E-01	NA	-0.19	0.85
KINASE_ACTIVITY	367	21622	-0.63	5.27E-01	NA	-1.81	0.07
INTRACELLULAR_TRANSPORT	281	10720	-0.64	5.23E-01	NA	-2.19	0.03
FATTY_ACID_BIOSYNTHETIC_PROCESS	14	171	-0.64	5.22E-01	NA	-0.14	0.89
CENTROSOME_ORGANIZATION_AND_BIOGENESIS	15	291	-0.64	5.21E-01	NA	0.80	0.42
PROTEIN_KINASE_REGULATOR_ACTIVITY	39	928	-0.65	5.18E-01	NA	-1.62	0.10
CATION_HOMEOSTASIS	109	2304	-0.65	5.18E-01	NA	-0.43	0.67
ANTIPORTER_ACTIVITY	10	111	-0.65	5.17E-01	NA	-1.07	0.29
CYSTEINE_TYPE_PEPTIDASE_ACTIVITY	54	1363	-0.65	5.16E-01	NA	0.47	0.64
SPINDLE_POLE	18	466	-0.65	5.16E-01	NA	1.57	0.12
MICROTUBULE_ORGANIZING_CENTER_ORGANIZATION_AND_BIOGENESIS	16	341	-0.65	5.15E-01	NA	1.52	0.13
INOSITOL_OR_PHOSPHATIDYLINOSITOL_KINASE_ACTIVITY	18	645	-0.65	5.14E-01	NA	-0.14	0.89
CELL_FRACTION	492	20216	-0.66	5.10E-01	NA	1.57	0.12
ACTIVATION_OF_JNK_ACTIVITY	16	565	-0.66	5.09E-01	NA	-1.06	0.29
CYSTEINE_TYPE_ENDOPEPTIDASE_ACTIVITY	40	1044	-0.66	5.08E-01	NA	0.13	0.90
REGULATION_OF_DNA_BINDING	47	1102	-0.66	5.07E-01	NA	-0.93	0.35
NEGATIVE_REGULATION_OF_CELLULAR_COMPONENT_ORGANIZATION_AND_BIOGENESIS	28	755	-0.67	5.04E-01	NA	-0.61	0.54
CELL_SOMA	10	287	-0.67	5.03E-01	NA	-0.67	0.50
REGULATION_OF_TRANSCRIPTION_FACTOR_ACTIVITY	40	1023	-0.68	4.99E-01	NA	-0.87	0.39
UBIQUITIN_LIGASE_COMPLEX	26	955	-0.68	4.96E-01	NA	-0.98	0.32
CARBONATE_DEHYDRATASE_ACTIVITY	13	172	-0.68	4.94E-01	NA	1.08	0.28
APOPTOTIC_NUCLEAR_CHANGES	19	251	-0.69	4.92E-01	NA	1.56	0.12
ARYLSULFATASE_ACTIVITY	12	575	-0.69	4.87E-01	NA	0.89	0.37
NUCLEASE_ACTIVITY	55	934	-0.70	4.87E-01	NA	0.00	1.00
MORPHOGENESIS_OF_AN_EPITHELIUM	16	431	-0.70	4.86E-01	NA	1.70	0.09
ALCOHOL_METABOLIC_PROCESS	88	1959	-0.70	4.83E-01	NA	-0.69	0.49
GOLGI_STACK	13	632	-0.70	4.82E-01	NA	-0.33	0.74
PROTEIN_IMPORT_INTO_NUCLEUS__TRANSLOCATION	11	444	-0.71	4.79E-01	NA	-0.06	0.95
GLUCOSE_METABOLIC_PROCESS	28	364	-0.72	4.74E-01	NA	-0.10	0.92
PROTEIN_IMPORT	62	2400	-0.72	4.71E-01	NA	-2.04	0.04
DNA_PACKAGING	35	499	-0.72	4.70E-01	NA	0.61	0.54
OXIDOREDUCTASE_ACTIVITY__ACTING_ON_CH_OH_GROUP_OF_DONORS	62	1397	-0.73	4.66E-01	NA	0.32	0.75
SARCOMERE	14	589	-0.74	4.62E-01	NA	-0.85	0.40
ONE_CARBON_COMPOUND_METABOLIC_PROCESS	26	707	-0.74	4.60E-01	NA	1.75	0.08
MONOCARBOXYLIC_ACID_METABOLIC_PROCESS	88	2287	-0.75	4.56E-01	NA	1.42	0.16
ION_HOMEOSTASIS	129	2944	-0.75	4.56E-01	NA	0.36	0.72
COFACTOR_TRANSPORTER_ACTIVITY	10	165	-0.75	4.53E-01	NA	3.63	0.00
CENTROSOME	58	2183	-0.75	4.53E-01	NA	-1.27	0.20
VASCULATURE_DEVELOPMENT	55	3756	-0.75	4.52E-01	NA	2.10	0.04
RIBONUCLEOPROTEIN_COMPLEX_BIOGENESIS_AND_ASSEMBLY	86	1454	-0.75	4.51E-01	NA	-1.87	0.06
RESPONSE_TO_CARBOHYDRATE_STIMULUS	12	350	-0.75	4.51E-01	NA	-0.91	0.36
CYTOKINE_PRODUCTION	73	1560	-0.76	4.50E-01	NA	-0.14	0.89
NEGATIVE_REGULATION_OF_TRANSCRIPTION_FROM_RNA_POLYMERASE_II_PROMOTER	83	3288	-0.76	4.50E-01	NA	1.25	0.21
MICROTUBULE_CYTOSKELETON	152	5648	-0.76	4.49E-01	NA	-1.27	0.20
SULFURIC_ESTER_HYDROLASE_ACTIVITY	16	626	-0.76	4.48E-01	NA	1.02	0.31
KINASE_REGULATOR_ACTIVITY	46	1096	-0.76	4.46E-01	NA	-1.04	0.30
ANGIOGENESIS	48	3590	-0.76	4.46E-01	NA	2.20	0.03
STEROID_DEHYDROGENASE_ACTIVITY__ACTING_ON_THE_CH_OH_GROUP_OF_DONORS__NAD_OR_NADP_AS_ACCEPTOR	9	94	-0.76	4.45E-01	NA	0.24	0.81
POTASSIUM_CHANNEL_REGULATOR_ACTIVITY	14	799	-0.77	4.44E-01	NA	1.89	0.06
G_PROTEIN_SIGNALING__ADENYLATE_CYCLASE_ACTIVATING_PATHWAY	25	484	-0.77	4.41E-01	NA	-1.00	0.32

EPIDERMIS_DEVELOPMENT	71	1546	-0.77	4.40E-01	NA	0.26	0.79
PROTEIN_KINASE_ACTIVITY	284	17705	-0.77	4.39E-01	NA	-1.91	0.06
CYTOKINE_SECRETION	18	550	-0.79	4.31E-01	NA	0.16	0.87
CELLULAR_RESPONSE_TO_NUTRIENT_LEVELS	10	102	-0.79	4.30E-01	NA	-1.17	0.24
HYDROGEN_ION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	27	449	-0.80	4.26E-01	NA	-0.04	0.97
REGULATION_OF_GENE_SPECIFIC_TRANSCRIPTION	11	621	-0.80	4.24E-01	NA	-1.22	0.22
HYDROLASE_ACTIVITY__HYDROLYZING_N_GLYCOSYL_COMPOUNDS	10	293	-0.80	4.24E-01	NA	-0.20	0.84
CYTOSOL	205	7618	-0.80	4.22E-01	NA	-1.22	0.22
ENDODEOXYRIBONUCLEASE_ACTIVITY	11	208	-0.81	4.20E-01	NA	0.85	0.40
HEMATOPOIETIN_INTERFERON_CLASS__D200_DOMAIN__CYTOKINE_RECEPTOR_BINDING	29	255	-0.81	4.20E-01	NA	1.05	0.29
POSITIVE_REGULATION_OF_LYMPHOCYTE_ACTIVATION	24	620	-0.81	4.16E-01	NA	-0.51	0.61
CELLULAR_LIPID_METABOLIC_PROCESS	255	8048	-0.82	4.15E-01	NA	0.00	1.00
OXIDOREDUCTASE_ACTIVITY_GO_0016616	56	1163	-0.83	4.06E-01	NA	0.04	0.96
NEUTRAL_AMINO_ACID_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	12	209	-0.84	4.03E-01	NA	0.48	0.63
REGULATION_OF_DNA_METABOLIC_PROCESS	45	1130	-0.84	4.00E-01	NA	1.01	0.31
PROTEIN_TARGETING_TO_MEMBRANE	12	183	-0.85	3.96E-01	NA	-0.58	0.56
3__5__EXONUCLEASE_ACTIVITY	13	309	-0.85	3.95E-01	NA	0.59	0.56
HETEROPHILIC_CELL_ADHESION	10	309	-0.86	3.88E-01	NA	0.28	0.78
HISTONE_DEACETYLASE_BINDING	10	70	-0.87	3.87E-01	NA	0.90	0.37
SPINDLE	39	1140	-0.87	3.83E-01	NA	1.66	0.10
TRANSITION_METAL_ION_TRANSPORT	12	204	-0.87	3.82E-01	NA	1.48	0.14
OVULATION_CYCLE	13	457	-0.88	3.82E-01	NA	1.63	0.10
DETECTION_OF_BIOTIC_STIMULUS	10	154	-0.88	3.80E-01	NA	0.58	0.56
MITOCHONDRIAL_SMALL_RIBOSOMAL_SUBUNIT	11	137	-0.88	3.80E-01	NA	-1.40	0.16
ORGANELLAR_SMALL_RIBOSOMAL_SUBUNIT	11	137	-0.88	3.80E-01	NA	-1.40	0.16
SMALL_RIBOSOMAL_SUBUNIT	11	137	-0.88	3.80E-01	NA	-1.40	0.16
MICROBODY	47	1233	-0.88	3.78E-01	NA	-0.10	0.92
PEROXISOME	47	1233	-0.88	3.78E-01	NA	-0.10	0.92
DNA_DAMAGE_RESPONSE__SIGNAL_TRANSDUCTION	35	1339	-0.88	3.77E-01	NA	1.44	0.15
LIPID_BIOSYNTHETIC_PROCESS	97	3420	-0.89	3.73E-01	NA	-0.80	0.42
CELL_CYCLE_CHECKPOINT_GO_0000075	48	1430	-0.89	3.73E-01	NA	0.95	0.34
TRANSITION_METAL_ION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	10	175	-0.90	3.71E-01	NA	0.81	0.42
UBIQUITIN_BINDING	11	194	-0.90	3.68E-01	NA	-1.23	0.22
MYOFIBRIL	19	1122	-0.90	3.67E-01	NA	-1.18	0.24
CYTOCHROME_C_OXIDASE_ACTIVITY	13	214	-0.91	3.63E-01	NA	1.28	0.20
REGULATION_OF_INTERFERON_GAMMA_BIOSYNTHETIC_PROCESS	11	80	-0.91	3.63E-01	NA	0.97	0.33
REGULATION_OF_HOMEOSTATIC_PROCESS	14	332	-0.92	3.59E-01	NA	-1.30	0.19
POSITIVE_REGULATION_OF_T_CELL_ACTIVATION	21	614	-0.92	3.57E-01	NA	-0.40	0.69
TISSUE_DEVELOPMENT	137	4065	-0.92	3.56E-01	NA	0.78	0.43
SUPEROXIDE_METABOLIC_PROCESS	10	417	-0.93	3.54E-01	NA	-0.63	0.53
INTERLEUKIN_8_PRODUCTION	12	148	-0.93	3.51E-01	NA	-1.13	0.26
RESPONSE_TO_HEAT	10	145	-0.94	3.49E-01	NA	-1.17	0.24
MICROBODY_PART	14	349	-0.94	3.49E-01	NA	-0.13	0.90
PEROXISOMAL_PART	14	349	-0.94	3.49E-01	NA	-0.13	0.90
AMINO_ACID_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	29	1161	-0.94	3.47E-01	NA	-0.32	0.75
RNA_BINDING	258	6044	-0.94	3.46E-01	NA	0.36	0.72
INTERFERON_GAMMA_PRODUCTION	14	98	-0.94	3.45E-01	NA	0.84	0.40
NEGATIVE_REGULATION_OF_DNA_BINDING	17	309	-0.95	3.41E-01	NA	-0.66	0.51
RESPONSE_TO_TEMPERATURE_STIMULUS	16	248	-0.96	3.39E-01	NA	-1.59	0.11
REGULATION_OF_CYCLIN_DEPENDENT_PROTEIN_KINASE_ACTIVITY	43	488	-0.96	3.39E-01	NA	-0.72	0.47
REGULATION_OF_CYTOKINE_SECRETION	16	387	-0.96	3.35E-01	NA	0.00	1.00
PIGMENT_BIOSYNTHETIC_PROCESS	17	665	-0.97	3.33E-01	NA	-0.57	0.57
EPITHELIAL_CELL_DIFFERENTIATION	10	236	-0.97	3.30E-01	NA	0.96	0.34
CHROMATIN	35	466	-0.98	3.28E-01	NA	0.69	0.49
ENDOPLASMIC_RETICULUM	292	12121	-0.98	3.26E-01	NA	-0.71	0.48
EMBRYONIC_DEVELOPMENT	57	4038	-0.99	3.23E-01	NA	1.11	0.27
DAMAGED_DNA_BINDING	20	649	-0.99	3.22E-01	NA	0.42	0.68
NITROGEN_COMPOUND_METABOLIC_PROCESS	154	6072	-1.00	3.19E-01	NA	1.93	0.05
NEGATIVE_REGULATION_OF_DNA_METABOLIC_PROCESS	18	262	-1.00	3.17E-01	NA	0.67	0.50
RIBOSOME_BIOGENESIS_AND_ASSEMBLY	18	165	-1.00	3.16E-01	NA	0.36	0.72
ACTIVATION_OF_NF_KAPPAB_TRANSCRIPTION_FACTOR	18	673	-1.00	3.16E-01	NA	-0.99	0.32
SECONDARY_ACTIVE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	47	1633	-1.00	3.16E-01	NA	1.68	0.09
PHOSPHOTRANSFERASE_ACTIVITY__ALCOHOL_GROUP_AS_ACCEPTOR	333	19215	-1.00	3.16E-01	NA	-1.51	0.13
OXYGEN_BINDING	22	404	-1.01	3.15E-01	NA	-0.05	0.96
NEGATIVE_REGULATION_OF_MAP_KINASE_ACTIVITY	17	601	-1.01	3.13E-01	NA	-0.21	0.83
SYMPORTER_ACTIVITY	31	1117	-1.01	3.12E-01	NA	0.87	0.39
FEMALE_PREGNANCY	52	931	-1.01	3.11E-01	NA	0.73	0.47
STEROID_HORMONE_RECEPTOR_BINDING	10	310	-1.03	3.03E-01	NA	0.73	0.46
CHEMICAL_HOMEOSTASIS	155	3393	-1.03	3.03E-01	NA	0.17	0.86
DIGESTION	44	706	-1.04	2.98E-01	NA	1.92	0.06
ANION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	59	1875	-1.04	2.96E-01	NA	0.52	0.60
ECTODERM_DEVELOPMENT	80	1651	-1.05	2.93E-01	NA	0.55	0.58
GOLGI_VESICLE_TRANSPORT	48	1386	-1.07	2.86E-01	NA	-1.57	0.12
RESPONSE_TO_OTHER_ORGANISM	83	912	-1.07	2.84E-01	NA	-0.27	0.79
FATTY_ACID_BETA_OXIDATION	11	212	-1.08	2.82E-01	NA	1.12	0.26
SECONDARY_METABOLIC_PROCESS	26	791	-1.08	2.82E-01	NA	-1.96	0.05
NUCLEOTIDE_EXCISION_REPAIR	20	518	-1.08	2.81E-01	NA	-0.21	0.84
MAINTENANCE_OF_LOCALIZATION	22	620	-1.09	2.74E-01	NA	-0.01	0.99
RESPONSE_TO_LIGHT_STIMULUS	46	1012	-1.10	2.72E-01	NA	-0.15	0.88
MITOCHONDRION_ORGANIZATION_AND_BIOGENESIS	48	817	-1.11	2.68E-01	NA	-1.09	0.27
CELL_STRUCTURE_DISASSEMBLY_DURING_APOPTOSIS	18	250	-1.11	2.67E-01	NA	1.26	0.21
CARBOHYDRATE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	16	332	-1.12	2.63E-01	NA	3.37	0.00
PHOSPHATASE_INHIBITOR_ACTIVITY	11	326	-1.12	2.63E-01	NA	-0.28	0.78
NEGATIVE_REGULATION_OF_TRANSCRIPTION_FACTOR_ACTIVITY	15	235	-1.12	2.62E-01	NA	-0.96	0.34
RNA_3__END_PROCESSING	10	137	-1.12	2.61E-01	NA	-1.14	0.25
PIGMENT_METABOLIC_PROCESS	18	690	-1.13	2.59E-01	NA	-0.58	0.56
CONTRACTILE_FIBER	25	1128	-1.13	2.58E-01	NA	-1.21	0.23
APOPTOTIC_MITOCHONDRIAL_CHANGES	11	126	-1.13	2.58E-01	NA	0.14	0.89
ACETYLCHOLINE_BINDING	17	598	-1.13	2.57E-01	NA	1.16	0.24
COLLAGEN	23	1772	-1.13	2.57E-01	NA	0.99	0.32
NUCLEAR_CHROMATIN	14	214	-1.13	2.57E-01	NA	0.67	0.50
FOCAL_ADHESION	13	422	-1.14	2.56E-01	NA	0.89	0.38
POSITIVE_REGULATION_OF_CYTOKINE_SECRETION	10	358	-1.15	2.52E-01	NA	0.19	0.85
COENZYME_METABOLIC_PROCESS	38	913	-1.16	2.45E-01	NA	-0.68	0.50
REGULATION_OF_T_CELL_ACTIVATION	28	791	-1.18	2.37E-01	NA	-0.07	0.94
RESPONSE_TO_BIOTIC_STIMULUS	120	1619	-1.18	2.37E-01	NA	-0.24	0.81
NUCLEOSIDE_TRIPHOSPHATASE_ACTIVITY	212	5969	-1.18	2.36E-01	NA	-0.39	0.70
POSITIVE_REGULATION_OF_TRANSLATION	35	462	-1.21	2.28E-01	NA	1.15	0.25
MICROTUBULE_CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS	35	1102	-1.21	2.28E-01	NA	-0.42	0.67
CYTOSOLIC_PART	23	543	-1.22	2.24E-01	NA	-0.51	0.61
ER_TO_GOLGI_VESICLE_MEDIATED_TRANSPORT	18	498	-1.22	2.22E-01	NA	-1.47	0.14
SMALL_NUCLEAR_RIBONUCLEOPROTEIN_COMPLEX	22	212	-1.22	2.21E-01	NA	1.30	0.19
REGULATION_OF_TRANSFERASE_ACTIVITY	161	5009	-1.24	2.16E-01	NA	0.11	0.91
STEROID_DEHYDROGENASE_ACTIVITY	10	121	-1.24	2.15E-01	NA	0.14	0.89
RRNA_METABOLIC_PROCESS	16	118	-1.24	2.14E-01	NA	0.48	0.63
SPLICEOSOME	51	679	-1.24	2.13E-01	NA	2.41	0.02
BASE_EXCISION_REPAIR	16	423	-1.25	2.10E-01	NA	0.24	0.81

PROTEIN_METHYLTRANSFERASE_ACTIVITY	14	469	-1.26	2.09E-01	NA	1.06	0.29
POSITIVE_REGULATION_OF_PROTEIN_SECRETION	12	359	-1.26	2.06E-01	NA	0.12	0.90
AMINE_METABOLIC_PROCESS	141	5786	-1.27	2.03E-01	NA	1.70	0.09
HISTONE_METHYLTRANSFERASE_ACTIVITY	11	389	-1.28	2.00E-01	NA	1.35	0.18
N_METHYLTRANSFERASE_ACTIVITY	13	405	-1.28	2.00E-01	NA	1.65	0.10
DNA_BINDING	597	19374	-1.29	1.98E-01	NA	-0.64	0.52
MITOTIC_CELL_CYCLE_CHECKPOINT	21	420	-1.29	1.97E-01	NA	0.72	0.47
REGULATION_OF_PROTEIN_KINASE_ACTIVITY	155	4829	-1.29	1.96E-01	NA	0.30	0.76
RRNA_PROCESSING	15	111	-1.29	1.96E-01	NA	0.27	0.79
NUCLEAR_TRANSPORT	89	3748	-1.30	1.94E-01	NA	-1.66	0.10
NUCLEOCYTOPLASMIC_TRANSPORT	88	3747	-1.30	1.92E-01	NA	-1.66	0.10
INTRACELLULAR_ORGANELLE_PART	1190	39518	-1.31	1.92E-01	NA	0.40	0.69
NEGATIVE_REGULATION_OF_HYDROLASE_ACTIVITY	14	177	-1.31	1.91E-01	NA	-0.12	0.91
SERINE_TYPE_ENDOPEPTIDASE_ACTIVITY	41	1208	-1.32	1.88E-01	NA	0.77	0.44
IMMUNE_RESPONSE	234	5095	-1.32	1.88E-01	NA	-0.40	0.69
PHOSPHATASE_REGULATOR_ACTIVITY	26	1309	-1.32	1.87E-01	NA	-1.67	0.10
ER_GOLGI_INTERMEDIATE_COMPARTMENT	24	793	-1.32	1.86E-01	NA	1.48	0.14
REGULATION_OF_KINASE_ACTIVITY	157	4876	-1.33	1.85E-01	NA	0.29	0.77
SERINE_TYPE_PEPTIDASE_ACTIVITY	45	1360	-1.33	1.83E-01	NA	0.79	0.43
OXIDOREDUCTASE_ACTIVITY__ACTING_ON_THE_ALDEHYDE_OR_OXO_GROUP_OF_DONORS__NAD_OR_NADP_AS_ACCEPTOR	16	402	-1.33	1.83E-01	NA	0.48	0.63
NEGATIVE_REGULATION_OF_BINDING	18	344	-1.34	1.80E-01	NA	-0.55	0.58
INTERFERON_GAMMA_BIOSYNTHETIC_PROCESS	12	97	-1.35	1.76E-01	NA	0.85	0.40
NUCLEOTIDE_SUGAR_METABOLIC_PROCESS	10	622	-1.35	1.76E-01	NA	-0.21	0.83
MEMBRANE_FRACTION	339	15230	-1.36	1.75E-01	NA	1.79	0.07
LYMPHOCYTE_DIFFERENTIATION	26	1088	-1.39	1.65E-01	NA	0.76	0.45
TASTE_RECEPTOR_ACTIVITY	15	16	-1.39	1.65E-01	NA	0.88	0.38
DNA_FRAGMENTATION_DURING_APOPTOSIS	13	227	-1.39	1.65E-01	NA	1.27	0.21
RIBONUCLEASE_ACTIVITY	25	278	-1.39	1.64E-01	NA	-0.16	0.87
SERINE_HYDROLASE_ACTIVITY	46	1361	-1.40	1.62E-01	NA	0.75	0.45
INORGANIC_ANION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	19	603	-1.40	1.60E-01	NA	1.27	0.21
INTRACELLULAR_PROTEIN_TRANSPORT	145	5805	-1.41	1.60E-01	NA	-1.86	0.06
STRUCTURE_SPECIFIC_DNA_BINDING	55	1589	-1.41	1.58E-01	NA	1.36	0.17
G1_PHASE_OF_MITOTIC_CELL_CYCLE	13	174	-1.42	1.56E-01	NA	1.19	0.24
ORGANELLE_PART	1195	39569	-1.42	1.55E-01	NA	0.33	0.74
PHOSPHOINOSITIDE_BIOSYNTHETIC_PROCESS	24	385	-1.43	1.52E-01	NA	-0.31	0.76
MICROTUBULE	32	1248	-1.46	1.45E-01	NA	-0.25	0.80
IMMUNE_SYSTEM_DEVELOPMENT	81	3272	-1.46	1.43E-01	NA	1.51	0.13
REGULATION_OF_CELLULAR_PH	10	256	-1.46	1.43E-01	NA	-1.25	0.21
LIPID_METABOLIC_PROCESS	324	9813	-1.47	1.42E-01	NA	-0.08	0.93
ENDOPLASMIC_RETICULUM_PART	97	3744	-1.47	1.42E-01	NA	-0.38	0.70
REGULATION_OF_ACTION_POTENTIAL	17	986	-1.47	1.41E-01	NA	1.02	0.31
DNA_POLYMERASE_ACTIVITY	18	400	-1.47	1.40E-01	NA	-1.09	0.28
RNA_METABOLIC_PROCESS	837	27093	-1.48	1.39E-01	NA	1.38	0.17
PROTEIN_PHOSPHATASE_TYPE_2A_REGULATOR_ACTIVITY	14	907	-1.48	1.38E-01	NA	-1.75	0.08
COFACTOR_BIOSYNTHETIC_PROCESS	21	689	-1.49	1.37E-01	NA	0.95	0.34
PROTEIN_MATURATION	11	109	-1.49	1.35E-01	NA	-0.06	0.95
HORMONE_BINDING	13	473	-1.50	1.34E-01	NA	1.18	0.24
HEMOPOIESIS	75	3198	-1.50	1.34E-01	NA	1.65	0.10
CELLULAR_MONOVALENT_INORGANIC_CATION_HOMEOSTASIS	11	270	-1.50	1.33E-01	NA	-1.52	0.13
HEMOPOIETIC_OR_LYMPHOID_ORGAN_DEVELOPMENT	77	3201	-1.51	1.31E-01	NA	1.55	0.12
RESPONSE_TO_BACTERIUM	31	413	-1.51	1.31E-01	NA	-0.60	0.55
VITAMIN_TRANSPORT	13	531	-1.51	1.31E-01	NA	2.33	0.02
RESPONSE_TO_RADIATION	60	1600	-1.51	1.30E-01	NA	-1.12	0.26
G1_PHASE	15	495	-1.54	1.24E-01	NA	-0.15	0.88
REGULATION_OF_PH	13	270	-1.55	1.20E-01	NA	-1.57	0.12
MEMBRANE_LIPID_METABOLIC_PROCESS	101	3065	-1.56	1.20E-01	NA	-0.27	0.79
DEFENSE_RESPONSE_TO_BACTERIUM	25	358	-1.58	1.13E-01	NA	-0.56	0.57
NEGATIVE_REGULATION_OF_ANGIOGENESIS	13	594	-1.59	1.12E-01	NA	0.38	0.70
SPINDLE_MICROTUBULE	16	675	-1.59	1.12E-01	NA	1.56	0.12
MONOVALENT_INORGANIC_CATION_HOMEOSTASIS	14	284	-1.59	1.12E-01	NA	-1.81	0.07
MICROTUBULE_POLYMERIZATION_OR_DEPOLYMERIZATION	11	578	-1.59	1.12E-01	NA	-1.18	0.24
AROMATIC_COMPOUND_METABOLIC_PROCESS	27	721	-1.59	1.11E-01	NA	0.18	0.86
ENERGY_RESERVE_METABOLIC_PROCESS	15	600	-1.59	1.11E-01	NA	-0.26	0.80
MITOCHONDRION	341	6094	-1.60	1.10E-01	NA	0.39	0.69
CARBON_CARBON_LYASE_ACTIVITY	18	631	-1.60	1.09E-01	NA	1.21	0.23
RESPONSE_TO_HORMONE_STIMULUS	33	1104	-1.61	1.08E-01	NA	-0.02	0.98
CONTRACTILE_FIBER_PART	23	989	-1.61	1.08E-01	NA	-1.29	0.20
LYMPHOCYTE_ACTIVATION	61	1862	-1.61	1.08E-01	NA	0.04	0.97
LEUKOCYTE_ACTIVATION	69	2157	-1.61	1.08E-01	NA	-0.31	0.76
ANDROGEN_RECEPTOR_SIGNALING_PATHWAY	12	137	-1.61	1.08E-01	NA	0.93	0.35
RIBONUCLEOPROTEIN_COMPLEX	143	1757	-1.61	1.07E-01	NA	0.56	0.57
MONOSACCHARIDE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	10	131	-1.62	1.06E-01	NA	2.43	0.02
PROTEIN_HETERODIMERIZATION_ACTIVITY	77	3460	-1.62	1.05E-01	NA	-0.03	0.98
INSULIN_LIKE_GROWTH_FACTOR_RECEPTOR_BINDING	10	192	-1.63	1.04E-01	NA	0.94	0.35
NUCLEAR_ENVELOPE	73	2255	-1.65	9.95E-02	NA	-1.59	0.11
RNA_POLYMERASE_ACTIVITY	16	263	-1.68	9.38E-02	NA	0.87	0.38
NEGATIVE_REGULATION_OF_MULTICELLULAR_ORGANISMAL_PROCESS	31	519	-1.68	9.28E-02	NA	-0.05	0.96
SUGAR_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	11	213	-1.70	8.92E-02	NA	2.53	0.01
IMMUNE_SYSTEM_PROCESS	331	9248	-1.70	8.85E-02	NA	-0.04	0.97
CELL_ACTIVATION	76	2612	-1.71	8.79E-02	NA	0.11	0.91
ENDONUCLEASE_ACTIVITY_GO_0016893	11	172	-1.71	8.78E-02	NA	-0.94	0.35
ENDORIBONUCLEASE_ACTIVITY	13	172	-1.71	8.78E-02	NA	-0.94	0.35
REGULATION_OF_T_CELL_PROLIFERATION	16	486	-1.71	8.66E-02	NA	0.09	0.93
LYASE_ACTIVITY	67	2414	-1.72	8.61E-02	NA	1.85	0.06
ATPASE_ACTIVITY__COUPLED_TO_MOVEMENT_OF_SUBSTANCES	40	1131	-1.72	8.48E-02	NA	0.01	0.99
TRANSCRIPTION_FROM_RNA_POLYMERASE_III_PROMOTER	19	281	-1.72	8.47E-02	NA	1.29	0.20
NEUTRAL_AMINO_ACID_TRANSPORT	10	144	-1.73	8.38E-02	NA	1.52	0.13
NICOTINIC_ACETYLCHOLINE_GATED_RECEPTOR_CHANNEL_COMPLEX	11	136	-1.75	7.97E-02	NA	0.53	0.60
NICOTINIC_ACETYLCHOLINE_ACTIVATED_CATION_SELECTIVE_CHANNEL_ACTIVITY	11	136	-1.75	7.97E-02	NA	0.53	0.60
TRANSFERASE_ACTIVITY__TRANSFERRING_ALKYL_OR_ARYL__OTHER_THAN_METHYL__GROUPS	30	623	-1.76	7.83E-02	NA	-0.97	0.33
AXON	12	480	-1.76	7.82E-02	NA	0.95	0.34
DEFENSE_RESPONSE_TO_VIRUS	12	209	-1.77	7.76E-02	NA	-1.91	0.06
HISTONE_MODIFICATION	24	605	-1.78	7.51E-02	NA	2.74	0.01
REGULATION_OF_LIPID_METABOLIC_PROCESS	12	325	-1.79	7.38E-02	NA	0.29	0.77
PROTEIN_SERINE_THREONINE_PHOSPHATASE_COMPLEX	10	271	-1.80	7.21E-02	NA	-0.28	0.78
INDUCTION_OF_APOPTOSIS_BY_INTRACELLULAR_SIGNALS	24	676	-1.80	7.16E-02	NA	0.76	0.45
NEGATIVE_REGULATION_OF_RESPONSE_TO_STIMULUS	11	316	-1.81	7.09E-02	NA	1.28	0.20
T_CELL_DIFFERENTIATION	15	397	-1.84	6.61E-02	NA	0.49	0.62
NUCLEOBASE__NUCLEOSIDE__NUCLEOTIDE_AND_NUCLEIC_ACID_METABOLIC_PROCESS	1237	44280	-1.84	6.59E-02	NA	0.32	0.75
FATTY_ACID_METABOLIC_PROCESS	63	1336	-1.84	6.56E-02	NA	1.30	0.19
REGULATION_OF_I_KAPPAB_KINASE_NF_KAPPAB_CASCADE	93	1758	-1.86	6.35E-02	NA	-0.05	0.96
EXTRACELLULAR_STRUCTURE_ORGANIZATION_AND_BIOGENESIS	32	2535	-1.86	6.34E-02	NA	1.03	0.30
GENE_SILENCING	10	280	-1.86	6.27E-02	NA	-0.77	0.44
T_CELL_ACTIVATION	44	1154	-1.86	6.25E-02	NA	-0.32	0.75
INTRAMOLECULAR_OXIDOREDUCTASE_ACTIVITY	20	361	-1.86	6.22E-02	NA	0.02	0.98
NITROGEN_COMPOUND_BIOSYNTHETIC_PROCESS	25	757	-1.87	6.18E-02	NA	0.59	0.55
COVALENT_CHROMATIN_MODIFICATION	25	629	-1.88	5.99E-02	NA	2.52	0.01

PHOSPHATE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	13	336	-1.90	5.72E-02	NA	1.48	0.14
LIGASE_ACTIVITY__FORMING_CARBON_OXYGEN_BONDS	14	682	-1.91	5.60E-02	NA	0.95	0.34
ACTIVATION_OF_PROTEIN_KINASE_ACTIVITY	28	1502	-1.92	5.51E-02	NA	1.04	0.30
INACTIVATION_OF_MAPK_ACTIVITY	14	440	-1.92	5.51E-02	NA	0.51	0.61
DOUBLE_STRANDED_DNA_BINDING	32	1173	-1.92	5.49E-02	NA	0.23	0.81
POSITIVE_REGULATION_OF_T_CELL_PROLIFERATION	13	451	-1.94	5.27E-02	NA	0.15	0.88
CARBOXY_LYASE_ACTIVITY	14	607	-1.96	5.03E-02	NA	1.00	0.32
HYDRO_LYASE_ACTIVITY	27	928	-1.97	**4.85E-02**	NA	0.38	0.70
CARBON_OXYGEN_LYASE_ACTIVITY	31	929	-1.99	**4.63E-02**	NA	0.40	0.69
NUCLEOBASE__NUCLEOSIDE__NUCLEOTIDE_AND_NUCLEIC_ACID_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	10	206	-1.99	**4.62E-02**	NA	-0.48	0.63
EXCITATORY_EXTRACELLULAR_LIGAND_GATED_ION_CHANNEL_ACTIVITY	21	436	-2.01	**4.44E-02**	NA	0.21	0.84
EXTRACELLULAR_LIGAND_GATED_ION_CHANNEL_ACTIVITY	22	436	-2.01	**4.44E-02**	NA	0.21	0.84
HEPARIN_BINDING	23	859	-2.03	**4.22E-02**	NA	-0.09	0.93
HISTONE_DEACETYLASE_COMPLEX	20	902	-2.04	**4.18E-02**	NA	0.30	0.76
NEGATIVE_REGULATION_OF_DNA_REPLICATION	13	214	-2.05	**4.07E-02**	NA	0.59	0.55
T_CELL_PROLIFERATION	19	732	-2.06	**3.98E-02**	NA	0.20	0.84
B_CELL_ACTIVATION	20	844	-2.06	**3.93E-02**	NA	0.13	0.90
PORE_COMPLEX	36	924	-2.10	**3.55E-02**	NA	-0.69	0.49
ESTABLISHMENT_AND_OR_MAINTENANCE_OF_CHROMATIN_ARCHITECTURE	77	1615	-2.11	**3.53E-02**	NA	0.09	0.93
DOUBLE_STRAND_BREAK_REPAIR	23	842	-2.12	**3.41E-02**	NA	1.01	0.31
ENDONUCLEASE_ACTIVITY	25	407	-2.14	**3.26E-02**	NA	-0.53	0.60
ACUTE_INFLAMMATORY_RESPONSE	11	66	-2.15	**3.19E-02**	NA	-0.54	0.59
ACTIVE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	122	4236	-2.15	**3.15E-02**	NA	0.04	0.97
RESPONSE_TO_STEROID_HORMONE_STIMULUS	11	259	-2.18	**2.89E-02**	NA	-0.21	0.83
NUCLEAR_ORGANIZATION_AND_BIOGENESIS	30	433	-2.20	**2.81E-02**	NA	0.21	0.83
HETEROCYCLE_METABOLIC_PROCESS	27	494	-2.20	**2.80E-02**	NA	-0.12	0.91
IMMUNE_EFFECTOR_PROCESS	38	707	-2.20	**2.80E-02**	NA	-0.50	0.62
HEME_BIOSYNTHETIC_PROCESS	10	185	-2.21	**2.71E-02**	NA	0.75	0.45
ANATOMICAL_STRUCTURE_FORMATION	56	4296	-2.23	**2.58E-02**	NA	2.14	0.03
NUCLEAR_MEMBRANE_PART	42	981	-2.25	**2.46E-02**	NA	-1.47	0.14
KINASE_ACTIVATOR_ACTIVITY	12	245	-2.25	**2.44E-02**	NA	-1.82	0.07
COFACTOR_TRANSPORT	11	194	-2.28	**2.25E-02**	NA	3.92	0.00
GLIOGENESIS	11	481	-2.30	**2.14E-02**	NA	0.05	0.96
NUCLEAR_PORE	31	775	-2.33	**1.98E-02**	NA	-1.15	0.25
CATABOLIC_PROCESS	225	7081	-2.34	**1.92E-02**	NA	-0.91	0.36
DNA_DIRECTED_RNA_POLYMERASE_COMPLEX	18	354	-2.34	**1.91E-02**	NA	2.64	0.01
NUCLEAR_DNA_DIRECTED_RNA_POLYMERASE_COMPLEX	18	354	-2.34	**1.91E-02**	NA	2.64	0.01
RNA_POLYMERASE_COMPLEX	18	354	-2.34	**1.91E-02**	NA	2.64	0.01
RNA_SPLICING_FACTOR_ACTIVITY__TRANSESTERIFICATION_MECHANISM	19	194	-2.35	**1.90E-02**	NA	1.58	0.11
HEME_METABOLIC_PROCESS	11	210	-2.35	**1.89E-02**	NA	0.70	0.48
ORGANELLE_MEMBRANE	298	8947	-2.35	**1.88E-02**	NA	-1.51	0.13
NUCLEAR_EXPORT	34	1551	-2.36	**1.84E-02**	NA	-0.73	0.46
POSITIVE_REGULATION_OF_I_KAPPAB_KINASE_NF_KAPPAB_CASCADE	87	1659	-2.37	**1.79E-02**	NA	-0.21	0.84
NUCLEAR_PART	579	16026	-2.39	**1.70E-02**	NA	1.60	0.11
OXIDOREDUCTASE_ACTIVITY__ACTING_ON_NADH_OR_NADPH	25	533	-2.39	**1.67E-02**	NA	-0.62	0.54
EARLY_ENDOSOME	18	883	-2.40	**1.63E-02**	NA	-1.01	0.31
KINESIN_COMPLEX	14	451	-2.40	**1.62E-02**	NA	0.83	0.40
INTEGRIN_BINDING	30	1721	-2.41	**1.58E-02**	NA	2.91	0.00
HYDROLASE_ACTIVITY__ACTING_ON_ACID_ANHYDRIDES__CATALYZING_TRANSMEMBRANE_MOVEMENT_OF_SUBSTANCES	39	1096	-2.42	**1.54E-02**	NA	0.01	0.99
DNA_DAMAGE_RESPONSE__SIGNAL_TRANSDUCTION_BY_P53_CLASS_MEDIATOR	13	379	-2.43	**1.49E-02**	NA	0.92	0.36
DNA_REPLICATION_INITIATION	16	351	-2.44	**1.48E-02**	NA	0.35	0.73
XENOBIOTIC_METABOLIC_PROCESS	11	169	-2.44	**1.45E-02**	NA	-1.32	0.19
NUCLEAR_LUMEN	387	11347	-2.45	**1.45E-02**	NA	1.83	0.07
RESPONSE_TO_STRESS	504	15846	-2.45	**1.41E-02**	NA	0.95	0.34
ANTIGEN_BINDING	22	311	-2.46	**1.39E-02**	NA	1.71	0.09
PRIMARY_ACTIVE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	40	1100	-2.47	**1.33E-02**	NA	-0.01	0.99
NUCLEOPLASM_PART	211	6498	-2.48	**1.32E-02**	NA	1.67	0.10
INFLAMMATORY_RESPONSE	130	2870	-2.50	**1.25E-02**	NA	0.52	0.61
MITOCHONDRIAL_MEMBRANE	85	1060	-2.50	**1.23E-02**	NA	-0.80	0.43
CELLULAR_CATABOLIC_PROCESS	212	6844	-2.52	**1.17E-02**	NA	-0.48	0.63
ANION_CATION_SYMPORTER_ACTIVITY	16	568	-2.56	**1.03E-02**	NA	-0.44	0.66
RESPONSE_TO_XENOBIOTIC_STIMULUS	12	182	-2.58	**9.83E-03**	NA	-1.55	0.12
DNA_DIRECTED_DNA_POLYMERASE_ACTIVITY	14	368	-2.62	**8.86E-03**	NA	-1.08	0.28
ADAPTIVE_IMMUNE_RESPONSE	25	589	-2.62	**8.76E-03**	NA	-0.37	0.71
REGULATION_OF_IMMUNE_EFFECTOR_PROCESS	15	478	-2.62	**8.73E-03**	NA	0.08	0.93
HYDROLASE_ACTIVITY__ACTING_ON_CARBON_NITROGEN__BUT_NOT_PEPTIDE__BONDS__IN_LINEAR_AMIDES	21	179	-2.63	**8.48E-03**	NA	0.62	0.53
ADAPTIVE_IMMUNE_RESPONSE_GO_0002460	24	571	-2.65	**8.17E-03**	NA	-0.24	0.81
REPRODUCTION	264	6264	-2.65	**8.04E-03**	NA	0.62	0.54
MITOCHONDRIAL_ENVELOPE	96	1144	-2.68	**7.37E-03**	NA	-1.15	0.25
RESPONSE_TO_UV	26	509	-2.70	**6.92E-03**	NA	0.82	0.41
MITOCHONDRIAL_MEMBRANE_PART	52	534	-2.84	**4.52E-03**	NA	0.12	0.91
M_PHASE	113	3407	-2.89	**3.88E-03**	NA	0.28	0.78
DNA_DIRECTED_RNA_POLYMERASE_II__HOLOENZYME	67	1915	-2.92	**3.46E-03**	NA	-0.22	0.83
ATPASE_ACTIVITY__COUPLED	93	2967	-2.94	**3.30E-03**	NA	-1.67	0.09
NUCLEUS	1422	44534	-2.96	**3.07E-03**	NA	-1.15	0.25
DNA_DAMAGE_RESPONSE__SIGNAL_TRANSDUCTION_RESULTING_IN_INDUCTION_OF_APOPTOSIS	15	320	-2.97	**3.00E-03**	NA	1.23	0.22
CHROMATIN_REMODELING	25	381	-3.00	**2.74E-03**	NA	-2.05	0.04
COFACTOR_METABOLIC_PROCESS	54	1223	-3.03	**2.42E-03**	NA	-0.23	0.82
DNA_DIRECTED_RNA_POLYMERASE_II__CORE_COMPLEX	13	311	-3.04	**2.33E-03**	NA	2.12	0.03
PRODUCTION_OF_MOLECULAR_MEDIATOR_OF_IMMUNE_RESPONSE	13	207	-3.10	**1.95E-03**	NA	-0.73	0.47
PHOSPHOLIPID_METABOLIC_PROCESS	74	1747	-3.11	**1.87E-03**	NA	-1.39	0.16
RIBOSOME	39	236	-3.11	**1.85E-03**	NA	-0.32	0.75
HELICASE_ACTIVITY	51	1692	-3.13	**1.77E-03**	NA	0.35	0.73
PHOSPHOINOSITIDE_METABOLIC_PROCESS	31	659	-3.13	**1.74E-03**	NA	-0.70	0.48
RIBOSOMAL_SUBUNIT	20	159	-3.14	**1.68E-03**	NA	-1.00	0.32
PROTEIN_TYROSINE_KINASE_ACTIVITY	63	6627	-3.15	**1.61E-03**	NA	-0.65	0.52
ATPASE_ACTIVITY	112	3499	-3.17	**1.53E-03**	NA	-1.05	0.29
MITOCHONDRIAL_RIBOSOME	22	168	-3.19	**1.41E-03**	NA	-0.98	0.33
ORGANELLAR_RIBOSOME	22	168	-3.19	**1.41E-03**	NA	-0.98	0.33
CHROMATIN_MODIFICATION	55	1161	-3.20	**1.37E-03**	NA	0.62	0.54
MITOCHONDRIAL_RESPIRATORY_CHAIN	23	334	-3.21	**1.33E-03**	NA	0.68	0.50
PHOSPHOLIPID_BIOSYNTHETIC_PROCESS	39	944	-3.22	**1.26E-03**	NA	-1.33	0.18
MRNA_SPLICE_SITE_SELECTION	13	157	-3.24	**1.20E-03**	NA	-1.10	0.27
OXIDOREDUCTASE_ACTIVITY__ACTING_ON_THE_CH_NH_GROUP_OF_DONORS	10	139	-3.24	**1.19E-03**	NA	1.43	0.15
ENVELOPE	168	3272	-3.30	**9.63E-04**	NA	-1.72	0.09
ORGANELLE_ENVELOPE	168	3272	-3.30	**9.63E-04**	NA	-1.72	0.09
SPLICEOSOME_ASSEMBLY	21	388	-3.32	**8.88E-04**	NA	-1.39	0.17
MITOCHONDRIAL_RESPIRATORY_CHAIN_COMPLEX_I	14	212	-3.38	**7.19E-04**	NA	1.60	0.11
NADH_DEHYDROGENASE_COMPLEX	14	212	-3.38	**7.19E-04**	NA	1.60	0.11
RESPIRATORY_CHAIN_COMPLEX_I	14	212	-3.38	**7.19E-04**	NA	1.60	0.11
ENDOPLASMIC_RETICULUM_LUMEN	14	726	-3.46	**5.40E-04**	NA	0.98	0.33
GAMETE_GENERATION	114	3200	-3.55	**3.82E-04**	NA	0.23	0.81
AMINO_ACID_AND_DERIVATIVE_METABOLIC_PROCESS	101	3696	-3.55	**3.82E-04**	NA	1.77	0.08
TRANSMEMBRANE_RECEPTOR_PROTEIN_TYROSINE_KINASE_ACTIVITY	43	5380	-3.59	**3.32E-04**	NA	-0.09	0.93
ANION_TRANSPORT	31	906	-3.61	**3.07E-04**	NA	0.40	0.69
BODY_FLUID_SECRETION	10	166	-3.68	**2.36E-04**	NA	1.48	0.14
RNA_EXPORT_FROM_NUCLEUS	21	691	-3.71	**2.11E-04**	NA	-1.54	0.12

NUCLEOPLASM	279	9107	-3.73	**1.88E-04**	NA	1.72	0.08
NEGATIVE_REGULATION_OF_TRANSFERASE_ACTIVITY	35	885	-3.74	**1.81E-04**	NA	-0.97	0.33
RESPONSE_TO_ENDOGENOUS_STIMULUS	198	8792	-3.76	**1.70E-04**	NA	0.90	0.37
DNA_RECOMBINATION	47	1802	-3.77	**1.66E-04**	NA	-0.75	0.45
SEXUAL_REPRODUCTION	139	3643	-3.77	**1.65E-04**	NA	1.01	0.31
STRUCTURAL_CONSTITUENT_OF_RIBOSOME	80	228	-3.78	**1.57E-04**	NA	-0.61	0.54
MRNA_METABOLIC_PROCESS	85	1878	-3.83	**1.29E-04**	NA	-0.69	0.49
LYSOSOME	61	1660	-3.88	**1.06E-04**	NA	0.44	0.66
LYTIC_VACUOLE	61	1660	-3.88	**1.06E-04**	NA	0.44	0.66
TRANSMEMBRANE_RECEPTOR_PROTEIN_KINASE_ACTIVITY	51	5644	-3.92	**8.68E-05**	NA	-0.32	0.75
VACUOLE	69	1806	-3.97	**7.14E-05**	NA	0.49	0.63
AMINO_ACID_METABOLIC_PROCESS	78	3182	-3.98	**6.92E-05**	NA	1.93	0.05
GUANYL_NUCLEOTIDE_BINDING	47	1132	-4.02	**5.74E-05**	NA	-1.90	0.06
CELLULAR_DEFENSE_RESPONSE	56	893	-4.03	**5.55E-05**	NA	-0.59	0.56
CARBOXYLIC_ACID_METABOLIC_PROCESS	178	5801	-4.13	**3.60E-05**	NA	1.68	0.09
MEIOTIC_CELL_CYCLE	35	1367	-4.14	**3.43E-05**	NA	-0.47	0.64
ORGANIC_ACID_METABOLIC_PROCESS	180	5813	-4.15	**3.33E-05**	NA	1.77	0.08
MESODERM_DEVELOPMENT	22	788	-4.18	**2.94E-05**	NA	-1.08	0.28
TRANSLATION_REGULATOR_ACTIVITY	41	808	-4.20	**2.68E-05**	NA	-0.36	0.72
NUCLEAR_MEMBRANE	50	1446	-4.25	**2.17E-05**	NA	-2.50	0.01
ORGANELLE_INNER_MEMBRANE	74	802	-4.28	**1.89E-05**	NA	-1.33	0.18
MITOCHONDRIAL_INNER_MEMBRANE	66	629	-4.29	**1.82E-05**	NA	-0.85	0.40
MEMBRANE_ENCLOSED_LUMEN	458	13079	-4.30	**1.73E-05**	NA	1.70	0.09
ORGANELLE_LUMEN	458	13079	-4.30	**1.73E-05**	NA	1.70	0.09
ATP_DEPENDENT_HELICASE_ACTIVITY	27	544	-4.31	**1.63E-05**	NA	-2.37	0.02
MEIOSIS_I	20	1221	-4.33	**1.52E-05**	NA	-0.95	0.34
GLUTAMINE_FAMILY_AMINO_ACID_METABOLIC_PROCESS	14	422	-4.33	**1.46E-05**	NA	-0.64	0.52
CELLULAR_LIPID_CATABOLIC_PROCESS	35	678	-4.34	**1.43E-05**	NA	-0.23	0.82
MEIOTIC_RECOMBINATION	17	1200	-4.38	**1.19E-05**	NA	-0.84	0.40
TISSUE_MORPHOGENESIS	14	834	-4.39	**1.14E-05**	NA	0.01	0.99
MITOCHONDRIAL_LUMEN	46	808	-4.45	**8.74E-06**	NA	-0.49	0.62
MITOCHONDRIAL_MATRIX	46	808	-4.45	**8.74E-06**	NA	-0.49	0.62
MRNA_PROCESSING_GO_0006397	74	1334	-4.52	**6.27E-06**	NA	-0.96	0.34
NUCLEOBASE__NUCLEOSIDE__NUCLEOTIDE_AND_NUCLEIC_ACID_TRANSPORT	32	978	-4.53	**5.80E-06**	NA	-1.74	0.08
PATTERN_SPECIFICATION_PROCESS	31	1197	-4.56	**5.20E-06**	NA	1.42	0.16
RNA_PROCESSING	173	3918	-4.56	**5.10E-06**	NA	-0.61	0.54
GLYCEROPHOSPHOLIPID_METABOLIC_PROCESS	46	953	-4.61	**3.97E-06**	NA	-1.20	0.23
TRANSLATION_FACTOR_ACTIVITY__NUCLEIC_ACID_BINDING	39	626	-4.65	**3.40E-06**	NA	-1.25	0.21
LIPID_CATABOLIC_PROCESS	38	680	-4.67	**2.98E-06**	NA	-0.33	0.74
GTP_BINDING	46	1041	-4.70	**2.63E-06**	NA	-2.28	0.02
DEFENSE_RESPONSE	269	5328	-4.70	**2.62E-06**	NA	0.67	0.51
MITOCHONDRIAL_PART	141	1946	-4.70	**2.55E-06**	NA	-1.15	0.25
NITROGEN_COMPOUND_CATABOLIC_PROCESS	29	994	-4.73	**2.28E-06**	NA	-0.09	0.93
TUBE_DEVELOPMENT	17	804	-4.74	**2.14E-06**	NA	1.90	0.06
AMINE_CATABOLIC_PROCESS	27	982	-4.75	**2.07E-06**	NA	-0.32	0.75
CELL_FATE_COMMITMENT	13	1155	-4.76	**1.91E-06**	NA	0.78	0.44
TUBE_MORPHOGENESIS	15	776	-4.76	**1.91E-06**	NA	1.91	0.06
GLYCEROPHOSPHOLIPID_BIOSYNTHETIC_PROCESS	30	456	-4.88	**1.04E-06**	NA	-0.88	0.38
TRANSCRIPTION_FACTOR_COMPLEX	89	2405	-4.96	**7.16E-07**	NA	0.55	0.58
DNA_METABOLIC_PROCESS	255	7753	-5.09	**3.59E-07**	NA	-0.63	0.53
ATP_DEPENDENT_RNA_HELICASE_ACTIVITY	17	233	-5.24	**1.58E-07**	NA	-2.90	0.00
DNA_REPLICATION	101	2915	-5.30	**1.18E-07**	NA	0.03	0.98
RNA_SPLICING__VIA_TRANSESTERIFICATION_REACTIONS	36	590	-5.32	**1.03E-07**	NA	-1.44	0.15
RNA_HELICASE_ACTIVITY	24	562	-5.35	**8.67E-08**	NA	-0.35	0.73
DNA_DEPENDENT_DNA_REPLICATION	55	1852	-5.36	**8.52E-08**	NA	1.00	0.32
RESPONSE_TO_DNA_DAMAGE_STIMULUS	160	5590	-5.58	**2.39E-08**	NA	0.05	0.96
PROTEIN_SERINE_THREONINE_PHOSPHATASE_ACTIVITY	24	353	-5.67	**1.39E-08**	NA	0.17	0.87
REGULATION_OF_GENE_EXPRESSION__EPIGENETIC	30	543	-5.69	**1.25E-08**	NA	-0.97	0.33
RNA_DEPENDENT_ATPASE_ACTIVITY	18	239	-5.70	**1.16E-08**	NA	-2.80	0.01
RNA_SPLICING	92	1455	-5.97	**2.31E-09**	NA	-0.91	0.36
AMINO_ACID_CATABOLIC_PROCESS	25	903	-5.99	**2.12E-09**	NA	0.09	0.93
BRUSH_BORDER	10	313	-6.01	**1.90E-09**	NA	-0.29	0.77
GLAND_DEVELOPMENT	13	286	-6.15	**7.95E-10**	NA	0.64	0.52
INORGANIC_ANION_TRANSPORT	18	415	-6.67	**2.55E-11**	NA	0.40	0.69
DNA_REPAIR	123	4001	-6.85	**7.31E-12**	NA	0.35	0.72

Appendix A2. Autosomal recessive variants from all individuals

Var	Gene	Case	Control	MAF	Exp Case	Exp Control	P perm
chr17:26684704..26684706	TMEM199	1	0	3.49E-07	0.0	0.0	0.0008
chr15:102211693..102211695	TARSL2	1	0	6.20E-07	0.0	0.0	0.0011
chr1:19166501..19166515	TAS1R2	1	0	4.75E-07	0.0	0.0	0.0013
chr6:109768320..109768322	MICAL1	1	0	4.75E-07	0.0	0.0	0.0013
chr1:16528294..16528295	ARHGEF19	1	0	7.85E-07	0.0	0.0	0.0014
chr19:36351305	KIRREL2	2	0	2.52E-05	0.1	0.1	0.0017
chr10:73082628..73082629	SLC29A3	1	0	7.85E-07	0.0	0.0	0.0018
chr9:5164266	INSL6	1	0	1.17E-06	0.0	0.0	0.002
chr15:75656531,chr15:75658862	MAN2C1	1	0	7.75E-07	0.0	0.0	0.0025
chr7:5920569	OCM	2	0	3.85E-05	0.2	0.2	0.0036
chr2:26644264	CCDC164	1	0	1.90E-06	0.0	0.0	0.0045
chr21:35094909..35094910	ITSN1	1	0	2.18E-06	0.0	0.0	0.005
chr15:50593417	GABPB1	1	0	3.14E-06	0.0	0.0	0.0054
chr6:83938651,chr6:84108085	ME1	1	0	2.21E-06	0.0	0.0	0.0061
chr1:228469903	OBSCN	2	0	6.36E-05	0.3	0.3	0.0063
chr20:3687141	SIGLEC1	3	0	0.000168862	0.9	0.9	0.0073
chr1:156521765	IQGAP3	1	0	3.88E-06	0.0	0.0	0.0084
chr14:24679928	CHMP4A	1	0	3.14E-06	0.0	0.0	0.0084
chr6:160969693,chr6:161006077	LPA	3	0	0.000248854	1.3	1.3	0.0141
chr1:97915614,chr1:97981498	DPYD	1	0	5.99E-06	0.0	0.0	0.0156
chr3:9859364,chr3:9874914	TTLL3	1	0	7.02E-06	0.0	0.0	0.0159
chr1:40947434	ZFP69	1	0	7.60E-06	0.0	0.0	0.017
chr1:36181910	C1orf216	1	0	8.15E-06	0.0	0.0	0.0193
chr2:231988421	HTR2B	1	0	8.15E-06	0.0	0.0	0.0223
chr3:195452000..195452001	MUC20	1	0	9.92E-06	0.1	0.1	0.025
chr5:148989259	ARHGEF37	2	0	0.000108891	0.6	0.6	0.0265
chr12:31815202	METTL20	1	0	1.26E-05	0.1	0.1	0.0293
chr16:48130781,chr16:48174765	ABCC12	2	0	0.000127993	0.6	0.7	0.0312
chr17:80369330	OGFOD3	1	0	1.40E-05	0.1	0.1	0.0321
chr4:89912303	FAM13A	1	0	1.40E-05	0.1	0.1	0.0352
chr14:94931042	SERPINA9	1	0	1.55E-05	0.1	0.1	0.0374
chr2:108863758,chr2:108881338	SULT1C3	1	0	1.55E-05	0.1	0.1	0.0398
chr9:125239501	OR1J1	2	0	0.000163784	0.8	0.8	0.0418
chr17:37034365	LASP1	1	0	1.96E-05	0.1	0.1	0.0481
chr16:48130781,chr16:48141231	ABCC12	2	1	2.18E-05	0.1	0.1	0.0482
chr16:48139232	ABCC12	1	0	2.42E-05	0.1	0.1	0.052
chr2:209302328,chr2:209309610	PTH2R	1	0	1.81E-05	0.1	0.1	0.052
chr14:57947421	C14orf105	8	2	0.001527442	7.7	7.8	0.053
chr1:3544083	TPRG1L	1	0	3.04E-05	0.2	0.2	0.058
chr4:89318021	HERC6	1	0	2.72E-05	0.1	0.1	0.058
chr10:37419292	ANKRD30A	2	0	0.000229841	1.2	1.2	0.063
chr19:35718891,chr19:35719490	FAM187B	1	0	2.81E-05	0.1	0.1	0.0636
chr9:21481483	IFNE	1	0	3.26E-05	0.2	0.2	0.068
chr10:134230570	PWWP2B	1	0	3.97E-05	0.2	0.2	0.073
chr3:186461524	KNG1	6	1	0.000919357	4.7	4.7	0.073
chr2:163134090	IFIH1	1	0	3.73E-05	0.2	0.2	0.074
chr6:161006077	LPA	5	1	0.001166926	5.9	5.9	0.074
chr6:99885246,chr6:99930627	USP45	11	9	3.69E-05	0.2	0.2	0.078
chr3:44943164	TGM4	1	0	3.61E-05	0.2	0.2	0.079

chr19:35718891	FAM187B	4	1	0.001269989	6.4	6.5	0.082
chr13:50205029	ARL11	1	0	4.75E-05	0.2	0.2	0.087
chr14:102729886,chr14:102749873	MOK	1	0	3.97E-05	0.2	0.2	0.087
chr4:76521525	CDKL2	2	0	0.000447983	2.3	2.3	0.088
chr2:209302328	PTH2R	2	0	0.000530657	2.7	2.7	0.104
chr7:143048771	CLCN1	1	0	6.05E-05	0.3	0.3	0.119
chr16:334920	PDIA2	1	0	7.68E-05	0.4	0.4	0.142
chr2:162904013	DPP4	4	2	0.002095506	10.6	10.7	0.17
chr19:4328755	STAP2	1	0	0.000591258	3.0	3.0	0.175
chr11:58378422	ZFP91,ZFP91-CNTF	5	2	0.001262985	6.4	6.4	0.177
chr5:10227711	FAM173B	1	0	0.000110957	0.6	0.6	0.195
chr1:64089432	PGM1	1	0	0.000310521	1.6	1.6	0.22
chr20:31647696	BPIFB3	1	0	0.000151428	0.8	0.8	0.22
chr17:45447802,chr17:45452257	EFCAB13	1	0	0.000186772	0.9	0.9	0.228
chr20:44511257	ZSWIM1	2	1	0.000209421	1.1	1.1	0.23
chr2:106690470	C2orf40	1	0	0.000174018	0.9	0.9	0.236
chr13:100518634	CLYBL	5	3	0.001023648	5.2	5.2	0.25
chr13:53617309	OLFM4	1	0	0.000423327	2.1	2.2	0.26
chr7:20666235	ABCB5	2	1	0.000203761	1.0	1.0	0.261
chr15:22368957	OR4M2	1	0	0.000395445	2.0	2.0	0.27
chr17:60469326	EFCAB3	1	0	0.000187246	0.9	1.0	0.27
chr1:212985592	TATDN3	4	2	0.000872222	4.4	4.4	0.28
chr1:236706300	LGALS8	3	1	0.000503796	2.6	2.6	0.28
chr10:27688101	PTCHD3	1	0	0.00025121	1.3	1.3	0.28
chr4:70512787	UGT2A1	10	7	0.002307938	11.7	11.7	0.28
chr6:132938842	TAAR2	1	0	0.000353549	1.8	1.8	0.28
chr6:26370833	BTN3A2	3	2	0.001146831	5.8	5.8	0.28
chr7:23313823	GPNMB	1	0	0.000342531	1.7	1.7	0.28
chr19:55019261	LAIR2	2	1	0.000907455	4.6	4.6	0.3
chr17:78396004	ENDOV	4	3	0.001242083	6.3	6.3	0.35
chr3:193052769	ATP13A5	4	3	0.00166909	8.5	8.5	0.38
chr13:46287373	SPERT	7	7	0.001850746	9.4	9.4	0.51
chr14:102729886	MOK	3	3	0.00096162	4.9	4.9	0.51
chr2:98779439	VWA3B	6	7	0.002032892	10.3	10.3	0.53
chr20:36869005	KIAA1755	1	1	0.000349858	1.8	1.8	0.58
chr13:25670984	PABPC3	3	3	0.001629115	8.3	8.3	0.59
chr20:18142462	CSRP2BP	4	4	0.002432531	12.3	12.4	0.61
chr6:25969631	TRIM38	2	2	0.000447983	2.3	2.3	0.61
chr12:13243647	GSG1	1	1	0.000357263	1.8	1.8	0.62
chr16:1825982	EME2	1	1	0.000300197	1.5	1.5	0.63
chr19:2936535	ZNF77	4	5	0.001629115	8.3	8.3	0.65
chr4:15482360	CC2D2A	3	4	0.002015175	10.2	10.2	0.65
chr17:7290695	TNK1	4	5	0.002041784	10.4	10.4	0.66
chr19:57646393	ZIM3	1	1	0.000349858	1.8	1.8	0.66
chr5:68616079	CCDC125	1	1	0.000403315	2.0	2.1	0.68
chr12:1022569,chr12:1023218	RAD52	1	2	0.000223404	1.1	1.1	0.69
chr16:84495318	ATP2C2	1	2	0.000158783	0.8	0.8	0.69
chr9:125140206	PTGS1	1	1	0.000553581	2.8	2.8	0.69
chr6:36274148	PNPLA1	1	1	0.000151428	0.8	0.8	0.7
chr7:75401263	CCL26	1	1	0.00032811	1.7	1.7	0.7
chr1:12919041	PRAMEF2	2	3	0.000447983	2.3	2.3	0.71

chr19:51585978	KLK14	1	2	0.000372301	1.9	1.9	0.71
chr6:25778182	SLC17A4	1	3	0.000360992	1.8	1.8	0.71
chr11:4598956	C11orf40	3	4	0.001979974	10.0	10.1	0.72
chr12:113403675	OAS3	1	2	0.000376112	1.9	1.9	0.72
chr15:78807407	AGPHD1	1	1	0.000226864	1.2	1.2	0.72
chr16:48130781	ABCC12	3	4	0.000765236	3.9	3.9	0.72
chr6:30071363	TRIM31	1	1	0.000163784	0.8	0.8	0.72
chr11:62996038	SLC22A25	1	1	0.000179252	0.9	0.9	0.73
chr2:28081439	RBKS	1	1	0.000134943	0.7	0.7	0.73
chr6:111587310	KIAA1919	1	2	0.000539768	2.7	2.7	0.73
chr6:167711559	UNC93A	1	1	0.000331684	1.7	1.7	0.78
chr8:143763531	PSCA	2	4	0.00123515	6.3	6.3	0.78
chr17:45438788,chr17:45447802	EFCAB13	1	1	0.00010895	0.6	0.6	0.82
chr2:98165911	ANKRD36B	3	5	0.001775299	9.0	9.0	0.82
chr3:130187662	COL6A5	1	2	7.68E-05	0.4	0.4	0.82
chr12:1023218	RAD52	1	2	0.000123748	0.6	0.6	0.86
chr17:72588438	C17orf77	1	2	7.85E-05	0.4	0.4	0.86
chr22:21369548	P2RX6	1	1	6.36E-05	0.3	0.3	0.86
chr1:79101171	IFI44L	1	1	7.50E-05	0.4	0.4	0.88
chr18:51880889	STARD6	1	4	0.000701275	3.6	3.6	0.88
chr18:658170	C18orf56	1	3	0.000883891	4.5	4.5	0.88
chr22:18912677	PRODH	1	3	0.000973871	4.9	5.0	0.89
chr12:1022569	RAD52	0	1	0.000403315	2.0	2.1	0.9
chr2:108863758	SULT1C3	1	4	0.000685722	3.5	3.5	0.91
chr6:53133964	ELOVL5	1	5	0.000866413	4.4	4.4	0.91
chr1:109823457	PSRC1	0	1	0.000360992	1.8	1.8	0.92
chr17:39551099	KRT31	0	1	0.000722287	3.7	3.7	0.92
chr10:52005095	ASAH2	1	1	4.48E-05	0.2	0.2	0.93
chr2:163124596,chr2:163136505	IFIH1	1	1	2.06E-05	0.1	0.1	0.94
chr5:160113099	ATP10B	0	1	0.000307059	1.6	1.6	0.94
chr16:48130781,chr16:48139232	ABCC12	0	1	0.000136163	0.7	0.7	0.95
chr2:242128114	ANO7	0	1	0.000229841	1.2	1.2	0.95
chr3:126135332	CCDC37	0	1	9.12E-05	0.5	0.5	0.95
chr4:20728910	PACRGL	0	2	0.000267054	1.4	1.4	0.95
chr5:94749787	FAM81B	0	1	0.000680573	3.5	3.5	0.95
chr19:21910125	ZNF100	0	1	0.000163784	0.8	0.8	0.96
chr9:21304913	IFNA5	0	1	7.34E-05	0.4	0.4	0.96
chr16:48174765	ABCC12	0	1	2.14E-05	0.1	0.1	0.98
chr17:7760574	LSMD1	0	1	0.000108891	0.6	0.6	0.98
chr22:29656431	RHBDD3	0	3	0.000548956	2.8	2.8	0.98
chr3:56628033,chr3:56655621	CCDC66	0	1	2.07E-05	0.1	0.1	0.98
chr3:9874914	TTLL3	1	4	0.001269989	6.4	6.5	0.98
chr4:111431444	ENPEP	0	1	0.000168862	0.9	0.9	0.98
chr7:156468559	RNF32	0	3	0.000776168	3.9	3.9	0.98
chr1:89655720	GBP4	0	1	3.04E-05	0.2	0.2	0.99
chr11:67235051	TMEM134	0	1	6.20E-05	0.3	0.3	0.99
chr12:13240162,chr12:13243647	GSG1	0	1	8.00E-05	0.4	0.4	0.99
chr12:7475081	ACSM4	0	5	0.001370125	6.9	7.0	0.99
chr17:45438788,chr17:45452257	EFCAB13	0	1	9.85E-05	0.5	0.5	0.99
chr17:56348226	MPO	0	1	5.45E-05	0.3	0.3	0.99
chr2:28532947	BRE	0	1	2.93E-05	0.1	0.1	0.99

chr22:45684998	UPK3A	0	1	5.60E-05	0.3	0.3	0.99
chr1:152280023,chr1:152285861	FLG	0	4	4.48E-05	0.2	0.2	1
chr1:156035856	RAB25	0	1	2.80E-06	0.0	0.0	1
chr1:27688743	MAP3K6	0	1	2.83E-05	0.1	0.1	1
chr1:27700455	FCN3	0	1	4.69E-06	0.0	0.0	1
chr10:50901917	C10orf53	0	1	1.55E-05	0.1	0.1	1
chr11:102584135	MMP8	0	5	0.000660182	3.3	3.4	1
chr11:119054559..119054564	NLRX1	0	1	6.20E-07	0.0	0.0	1
chr11:5462702	OR51I1	0	3	0.000189949	1.0	1.0	1
chr11:59245593	OR4D10	0	1	0.000106847	0.5	0.5	1
chr11:59480952	OR10V1	0	1	5.60E-05	0.3	0.3	1
chr12:18234352	RERGL	0	1	7.06E-06	0.0	0.0	1
chr15:22369175..22369177	OR4M2	0	1	3.49E-07	0.0	0.0	1
chr15:73029908..73029919, chr15:73029915..73029917	BBS4	0	1	1.55E-07	0.0	0.0	1
chr15:86800251,chr15:87217666	AGBL1	0	1	3.41E-06	0.0	0.0	1
chr15:90274712..90274714	WDR93	0	1	6.20E-07	0.0	0.0	1
chr16:1967935	HS3ST6	0	1	4.75E-07	0.0	0.0	1
chr16:48204130,chr16:48261815	ABCC11	0	1	2.60E-05	0.1	0.1	1
chr17:4348380	SPNS3	0	1	2.93E-05	0.1	0.1	1
chr17:56272501	EPX	0	1	4.89E-05	0.2	0.2	1
chr17:65887957..65887960	BPTF	0	1	3.49E-07	0.0	0.0	1
chr19:17434468,chr19:17434475	ANO8	1	1	4.01E-06	0.0	0.0	1
chr19:35718891,chr19:35718956	FAM187B	0	1	1.05E-05	0.1	0.1	1
chr19:41594954	CYP2A13	0	1	4.75E-05	0.2	0.2	1
chr2:160714958	LY75-CD302,LY75	0	1	1.79E-05	0.1	0.1	1
chr2:196602786	DNAH7	0	1	9.31E-06	0.0	0.0	1
chr2:202245415	TRAK2	0	1	3.50E-06	0.0	0.0	1
chr2:29117790..29117791,chr2:29129	WDR43	0	1	4.75E-07	0.0	0.0	1
chr20:2397966..2397976	TGM6	0	1	1.90E-06	0.0	0.0	1
chr22:27019197	CRYBA4	0	1	1.12E-05	0.1	0.1	1
chr22:32643460	SLC5A4	0	2	0.000130407	0.7	0.7	1
chr3:132337477,chr3:132363709	ACAD11	0	1	2.11E-06	0.0	0.0	1
chr3:52721297..52721298	GNL3	0	1	6.20E-07	0.0	0.0	1
chr4:106888602	NPNT	0	1	6.20E-07	0.0	0.0	1
chr4:76805745	PPEF2	0	4	0.000558221	2.8	2.8	1
chr6:167711559,chr6:167717457	UNC93A	0	2	5.02E-05	0.3	0.3	1
chr6:25769227,chr6:25778182	SLC17A4	0	1	3.74E-06	0.0	0.0	1
chr6:30115320	TRIM40	0	1	1.12E-05	0.1	0.1	1
chr6:31237115..31237116	HLA-C	0	1	3.49E-07	0.0	0.0	1
chr7:23871861	STK31	0	1	2.18E-06	0.0	0.0	1
chr8:52284560	PXDNL	0	2	7.85E-05	0.4	0.4	1
chr9:139259592	CARD9	0	1	1.33E-05	0.1	0.1	1
chr9:35906438	HRCT1	0	1	1.26E-05	0.1	0.1	1
Total		**242**	**254**		**338.9**	**339.8**	

Appendix A3. Autosomal recessive variants from homgenous Swedish subset of individuals

Var	Gene	Case	Control	MAF	Exp Case	Exp Control	P perm
chr11:118851208..118851209	FOXR1	1	0	6.01E-07	0.0	0.0	0.0007
chr10:73082628..73082629	SLC29A3	1	0	1.18E-06	0.0	0.0	0.0019
chr15:102211693..102211695	TARSL2	1	0	1.18E-06	0.0	0.0	0.0019
chr15:75656531,chr15:75658862	MAN2C1	1	0	1.08E-06	0.0	0.0	0.0020
chr21:35094909..35094910	ITSN1	1	0	2.40E-06	0.0	0.0	0.0032
chr14:57947421	C14orf105	6	0	0.001477457	4.5	5.0	0.0037
chr2:26644264	CCDC164	1	0	3.46E-06	0.0	0.0	0.0042
chr6:83938651,chr6:84108085	ME1	1	0	3.56E-06	0.0	0.0	0.0051
chr1:156521765	IQGAP3	1	0	4.71E-06	0.0	0.0	0.0064
chr14:24679928	CHMP4A	1	0	6.94E-06	0.0	0.0	0.0100
chr1:97915614,chr1:97981498	DPYD	1	0	7.78E-06	0.0	0.0	0.0121
chr12:31815202	METTL20	1	0	1.27E-05	0.0	0.0	0.0200
chr2:209302328,chr2:209309610	PTH2R	1	0	1.47E-05	0.0	0.0	0.0202
chr16:48130781,chr16:48174765	ABCC12	2	0	1.63E-04	0.5	0.6	0.0220
chr19:35718891,chr19:35719490	FAM187B	1	0	1.51E-05	0.0	0.1	0.0220
chr9:21481483	IFNE	1	0	2.02E-05	0.1	0.1	0.0292
chr14:94931042	SERPINA9	1	0	2.31E-05	0.1	0.1	0.0325
chr17:80369330	OGFOD3	1	0	2.16E-05	0.1	0.1	0.0327
chr2:108863758,chr2:108881338	SULT1C3	1	0	2.51E-05	0.1	0.1	0.0347
chr3:44943164	TGM4	1	0	2.78E-05	0.1	0.1	0.0373
chr1:3544083	TPRG1L	1	0	3.11E-05	0.1	0.1	0.0433
chr6:99885246,chr6:99930627	USP45	5	4	3.19E-05	0.1	0.1	0.0437
chr7:143048771	CLCN1	1	0	3.47E-05	0.1	0.1	0.0479
chr19:36351305	KIRREL2	1	0	3.29E-05	0.1	0.1	0.05
chr13:50205029	ARL11	1	0	5.08E-05	0.2	0.2	0.06
chr7:5920569	OCM	1	0	5.08E-05	0.2	0.2	0.07
chr6:161006077	LPA	2	0	1.19E-03	3.7	4.0	0.07
chr19:35718891	FAM187B	2	0	0.001059372	3.3	3.6	0.08
chr2:162904013	DPP4	3	1	2.46E-03	7.5	8.3	0.09
chr3:186461524	KNG1	2	0	0.000951297	2.9	3.2	0.10
chr5:148989259	ARHGEF37	1	0	0.000101494	0.3	0.3	0.12
chr20:31647696	BPIFB3	1	0	1.14E-04	0.4	0.4	0.13
chr6:25969631	TRIM38	1	0	9.53E-05	0.3	0.3	0.13
chr20:44511257	ZSWIM1	1	0	0.000128013	0.4	0.4	0.14
chr19:55019261	LAIR2	1	0	0.000960882	2.9	3.3	0.15
chr6:30071363	TRIM31	1	0	0.00017356	0.5	0.6	0.16
chr5:10227711	FAM173B	1	0	0.000121095	0.4	0.4	0.16
chr10:37419292	ANKRD30A	1	0	2.85E-04	0.9	1.0	0.20
chr6:160969693,chr6:161006077	LPA	1	0	0.000176779	0.5	0.6	0.20
chr19:4328755	STAP2	1	0	5.92E-04	1.8	2.0	0.22
chr4:70512787	UGT2A1	6	3	0.002413963	7.4	8.2	0.22
chr17:45447802,chr17:45452257	EFCAB13	1	0	2.45E-04	0.8	0.8	0.22
chr6:132938842	TAAR2	1	0	0.000345919	1.1	1.2	0.23
chr10:27688101	PTCHD3	1	0	0.000285407	0.9	1.0	0.23
chr13:46287373	SPERT	4	2	0.001599008	4.9	5.4	0.23
chr20:3687141	SIGLEC1	1	0	2.26E-04	0.7	0.8	0.23
chr9:125239501	OR1J1	1	0	1.74E-04	0.5	0.6	0.24
chr1:64089432	PGM1	1	0	0.000306737	0.9	1.0	0.25
chr15:22368957	OR4M2	1	0	0.000399752	1.2	1.4	0.25
chr19:57646393	ZIM3	1	0	0.000444312	1.4	1.5	0.26
chr13:25670984	PABPC3	2	1	0.001856524	5.7	6.3	0.26
chr13:53617309	OLFM4	1	0	2.55E-04	0.8	0.9	0.26
chr1:236706300	LGALS8	2	1	0.000540497	1.7	1.8	0.28

chr2:209302328	PTH2R	1	0	5.62E-04	1.7	1.9	0.28
chr8:143763531	PSCA	2	1	0.001216119	3.7	4.1	0.30
chr1:212985592	TATDN3	2	1	1.16E-03	3.6	3.9	0.31
chr4:76521525	CDKL2	1	0	3.40E-04	1.0	1.2	0.31
chr17:78396004	ENDOV	2	1	0.001152127	3.5	3.9	0.33
chr17:67149973	ABCA10	4	4	2.35E-03	7.2	8.0	0.50
chr2:98165911	ANKRD36B	3	3	2.01E-03	6.2	6.8	0.51
chr6:26370833	BTN3A2	1	1	1.29E-03	4.0	4.4	0.51
chr13:100518634	CLYBL	1	1	0.0008948	2.7	3.0	0.53
chr3:193052769	ATP13A5	2	2	1.55E-03	4.8	5.2	0.58
chr19:2936535	ZNF77	3	4	0.00202026	6.2	6.8	0.60
chr11:58378422	ZFP91,ZFP91-CNTF	1	1	1.15E-03	3.5	3.9	0.61
chr14:102729886	MOK	2	2	0.000913442	2.8	3.1	0.63
chr15:31294714	TRPM1	3	4	0.002475262	7.6	8.4	0.64
chr12:13243647	GSG1	1	1	0.000412244	1.3	1.4	0.68
chr17:7290695	TNK1	1	2	0.002190914	6.7	7.4	0.69
chr11:4598956	C11orf40	1	2	0.0021476	6.6	7.3	0.70
chr15:78807407	AGPHD1	1	1	0.000245051	0.8	0.8	0.70
chr20:36869005	KIAA1755	1	1	0.000345919	1.1	1.2	0.71
chr18:51880889	STARD6	1	3	7.87E-04	2.4	2.7	0.73
chr7:75401263	CCL26	1	1	3.63E-04	1.1	1.2	0.73
chr1:12919041	PRAMEF2	1	2	0.000431343	1.3	1.5	0.75
chr2:98779439	VWA3B	3	5	0.002264056	6.9	7.7	0.75
chr6:111587310	KIAA1919	1	2	6.30E-04	1.9	2.1	0.75
chr18:658170	C18orf56	1	3	1.15E-03	3.5	3.9	0.77
chr2:108863758	SULT1C3	1	3	7.27E-04	2.2	2.5	0.77
chr11:62996038	SLC22A25	1	1	1.95E-04	0.6	0.7	0.79
chr16:48130781	ABCC12	1	3	0.001019404	3.1	3.4	0.79
chr6:36274148	PNPLA1	1	1	1.74E-04	0.5	0.6	0.80
chr17:39551099	KRT31	0	1	0.000761114	2.3	2.6	0.84
chr5:94749787	FAM81B	0	1	0.00088555	2.7	3.0	0.85
chr22:18912677	PRODH	0	1	1.06E-03	3.3	3.6	0.86
chr4:15482360	CC2D2A	1	3	2.05E-03	6.3	6.9	0.90
chr12:113403675	OAS3	0	1	0.000424929	1.3	1.4	0.91
chr1:109823457	PSRC1	0	1	4.38E-04	1.3	1.5	0.92
chr1:79101171	IFI44L	1	1	7.53E-05	0.2	0.3	0.92
chr19:51585978	KLK14	0	1	0.000437801	1.3	1.5	0.92
chr3:9874914	TTLL3	1	4	0.001237836	3.8	4.2	0.92
chr12:7475081	ACSM4	0	2	0.001292963	4.0	4.4	0.93
chr4:76805745	PPEF2	0	1	0.000512054	1.6	1.7	0.93
chr16:1825982	EME2	0	1	2.60E-04	0.8	0.9	0.94
chr5:160113099	ATP10B	0	1	0.000181824	0.6	0.6	0.94
chr9:125140206	PTGS1	0	1	0.000444312	1.4	1.5	0.94
chr2:242128114	ANO7	0	1	1.86E-04	0.6	0.6	0.95
chr4:20728910	PACRGL	0	1	0.000207766	0.6	0.7	0.95
chr16:48130781,chr16:48139232	ABCC12	0	1	0.000158354	0.5	0.5	0.96
chr16:48130781,chr16:48141231	ABCC12	1	1	1.98E-05	0.1	0.1	0.97
chr19:21910125	ZNF100	0	1	1.62E-04	0.5	0.5	0.97
chr2:163124596,chr2:163136505	IFIH1	1	1	2.74E-05	0.1	0.1	0.97
chr17:7760574	LSMD1	0	1	1.62E-04	0.5	0.5	0.98
chr4:111431444	ENPEP	0	1	1.78E-04	0.5	0.6	0.98
chr6:53133964	ELOVL5	0	4	1.01E-03	3.1	3.4	0.98
chr7:156468559	RNF32	0	3	0.000932276	2.9	3.2	0.98
chr12:1022569,chr12:1023218	RAD52	0	1	2.15E-04	0.7	0.7	0.99
chr17:45438788,chr17:45452257	EFCAB13	0	1	0.000113	0.3	0.4	0.99

chr19:17434468,chr19:17434475	ANO8	1	1	6.05E-06	0.0	0.0	0.99
chr3:130187662	COL6A5	0	1	8.36E-05	0.3	0.3	0.99
chr8:52284560	PXDNL	0	1	8.94E-05	0.3	0.3	0.99
chr1:152280023,chr1:152285861	FLG	0	2	4.87E-05	0.1	0.2	1
chr1:156035856	RAB25	0	1	2.40E-06	0.0	0.0	1
chr1:27688743	MAP3K6	0	1	2.46E-05	0.1	0.1	1
chr1:89655720	GBP4	0	1	2.94E-05	0.1	0.1	1
chr10:52005095	ASAH2	0	1	2.94E-05	0.1	0.1	1
chr11:102584135	MMP8	0	5	0.00070243	2.2	2.4	1
chr11:5462702	OR51I1	0	2	0.000149923	0.5	0.5	1
chr12:1023218	RAD52	0	2	1.18E-04	0.4	0.4	1
chr12:18234352	RERGL	0	1	1.27E-05	0.0	0.0	1
chr15:44859633..44859637	SPG11	0	1	6.01E-07	0.0	0.0	1
chr15:73029908..73029919, chr15:73029915..73029917	BBS4	0	1	2.88E-07	0.0	0.0	1
chr15:86800251,chr15:87217666	AGBL1	0	1	2.31E-06	0.0	0.0	1
chr16:48204130,chr16:48261815	ABCC11	0	1	3.19E-05	0.1	0.1	1
chr17:4348380	SPNS3	0	1	4.44E-05	0.1	0.2	1
chr17:56272501	EPX	0	1	4.24E-05	0.1	0.1	1
chr17:56348226	MPO	0	1	7.27E-05	0.2	0.2	1
chr17:65887957..65887960	BPTF	0	1	6.01E-07	0.0	0.0	1
chr17:72588438	C17orf77	0	1	6.25E-05	0.2	0.2	1
chr19:35718891,chr19:35718956	FAM187B	0	1	1.01E-05	0.0	0.0	1
chr2:196602786	DNAH7	0	1	1.50E-05	0.0	0.1	1
chr2:202245415	TRAK2	0	1	5.40E-06	0.0	0.0	1
chr2:29117790..29117791,chr2:29129325	WDR43	0	1	8.65E-07	0.0	0.0	1
chr22:27019197	CRYBA4	0	1	1.38E-05	0.0	0.0	1
chr22:32643460	SLC5A4	0	2	0.00017356	0.5	0.6	1
chr3:126135332	CCDC37	0	1	5.53E-05	0.2	0.2	1
chr3:132337477,chr3:132363709	ACAD11	0	1	3.75E-06	0.0	0.0	1
chr3:56628033,chr3:56655621	CCDC66	0	1	2.31E-05	0.1	0.1	1
chr4:106888602	NPNT	0	1	8.65E-07	0.0	0.0	1
chr6:167711559,chr6:167717457	UNC93A	0	2	5.67E-05	0.2	0.2	1
chr6:25778182	SLC17A4	0	3	0.000393578	1.2	1.3	1
chr7:23871861	STK31	0	1	4.71E-06	0.0	0.0	1
chr9:21304913	IFNA5	0	1	4.65E-05	0.1	0.2	1
Total		**129**	**151**		**207.2**	**228.5**	

Appendix A4. Autosomal recessive variants from subset of Finnish individuals

Var	Gene	Case	Control	MAF	Exp Case	Exp Control	P perm
chr11:123813863..123813864	OR6T1	1	0	5.86E-06	0.0	0.0	0.0006
chr9:95076685..95076686	NOL8	1	0	5.86E-06	0.0	0.0	0.0006
chr14:68256161..68256162	ZFYVE26	1	0	5.86E-06	0.0	0.0	0.0008
chr7:130038730..130038732	CEP41	1	0	5.86E-06	0.0	0.0	0.0008
chr14:23523813..23523814	CDH24	1	0	5.86E-06	0.0	0.0	0.0008
chr1:152191442..152191462	HRNR	1	0	5.86E-06	0.0	0.0	0.0009
chr11:4566459..4566460	OR52M1	1	0	5.86E-06	0.0	0.0	0.0009
chr3:121341076..121341093	FBXO40	1	0	5.86E-06	0.0	0.0	0.0009
chr9:5164266	INSL6	1	0	5.86E-06	0.0	0.0	0.0012
chr19:57723399..57723400	ZNF264	1	0	5.86E-06	0.0	0.0	0.0014
chr19:47425241..47425243	ARHGAP35	1	0	1.32E-05	0.0	0.0	0.0022
chr16:48130781,chr16:48141231	ABCC12	1	0	5.28E-05	0.0	0.0	0.0152
chr1:228469903	OBSCN	2	0	0.000916049	0.5	0.3	0.0176
chr17:45447802,chr17:45468858	EFCAB13	1	0	7.62E-05	0.0	0.0	0.0203
chr6:160969693,chr6:161006077	LPA	2	0	0.00109633	0.6	0.3	0.0288
chr17:72588438	C17orf77	1	0	0.000146567	0.1	0.0	0.0351
chr6:53133964	ELOVL5	1	0	0.0002477	0.1	0.1	0.051
chr3:186461524	KNG1	1	0	0.000211057	0.1	0.1	0.060
chr2:231988421	HTR2B	1	0	0.000329778	0.2	0.1	0.064
chr1:212985592	TATDN3	1	0	0.000287275	0.2	0.1	0.076
chr17:37034365	LASP1	1	0	0.000329778	0.2	0.1	0.077
chr11:58378422	ZFP91,ZFP91-CNTF	2	0	0.002345097	1.3	0.7	0.106
chr12:1023218	RAD52	1	0	0.000423582	0.2	0.1	0.111
chr9:139333054..139333058	INPP5E	2	0	0.002463812	1.4	0.7	0.111
chr10:134230570	PWWP2B	1	0	0.000586274	0.3	0.2	0.142
chr13:25670984	PABPC3	1	0	0.000916049	0.5	0.3	0.205
chr4:89318021	HERC6	1	0	0.001068486	0.6	0.3	0.224
chr5:68616079	CCDC125	1	0	0.001068486	0.6	0.3	0.229
chr4:70512787	UGT2A1	2	1	0.001899521	1.0	0.5	0.32
chr13:46287373	SPERT	1	0	0.001899521	1.0	0.5	0.33
chr17:60469326	EFCAB3	1	0	0.00200652	1.1	0.6	0.35
chr6:161006077	LPA	1	0	0.001694327	0.9	0.5	0.35
chr3:193052769	ATP13A5	1	0	0.002116442	1.2	0.6	0.36
chr2:162904013	DPP4	1	0	0.002229303	1.2	0.6	0.38
chr11:47306630	MADD	1	1	0.002229303	1.2	0.6	0.83
chr14:102729886	MOK	1	1	0.001500857	0.8	0.4	0.86
chr4:76805745	PPEF2	0	1	0.001319113	0.7	0.4	0.98
chr4:15482360	CC2D2A	0	1	0.001149095	0.6	0.3	0.99
chr1:152280023,chr1:152285861	FLG	0	1	2.20E-05	0.0	0.0	1
chr1:168200807..168200809	SFT2D2	0	1	5.86E-06	0.0	0.0	1
chr12:106715314..106715316	TCP11L2	0	1	5.86E-06	0.0	0.0	1
chr12:113403675	OAS3	0	1	0.000287275	0.2	0.1	1
chr16:84495318	ATP2C2	0	1	0.000287275	0.2	0.1	1
chr17:55334440..55334442	MSI2	0	1	5.86E-06	0.0	0.0	1
chr20:44511257	ZSWIM1	0	1	0.000287275	0.2	0.1	1
chr22:29656431	RHBDD3	0	1	9.38E-05	0.1	0.0	1
chr22:45684998	UPK3A	0	1	0.001149095	0.6	0.3	1
chr4:20728910	PACRGL	0	1	0.000709391	0.4	0.2	1
Total		**41**	**15**		**18.4**	**9.3**	

Appendix A5. Genic results with details if occurs within known CNV region (red indicates nominally significant)

Gene	chr	start	end	gene length	case	control	all	p-value	Psych CNV
NADK	1	1682670	1711508	3653	4:7458	3:7189	7:14647	1.000	
PER3	1	7844762	7905237	6203	4:7458	2:7189	6:14647	0.654	
RERE	1	8412463	8877699	8194	0:7458	2:7189	2:14647	0.087	
VPS13D	1	12290112	12572098	16328	2:7458	0:7189	2:14647	0.103	
PRAMEF1	1	12851545	12856777	2160	0:7458	3:7189	3:14647	0.032	
LOC649330	1	12907235	12908237	1002	1:7458	2:7189	3:14647	1.000	
HNRNPCL1	1	12907260	12908578	1108	1:7458	2:7189	3:14647	1.000	
RSC1A1	1	15986363	15988217	1854	1:7458	4:7189	5:14647	0.130	
SPATA21	1	16725137	16763919	2015	3:7458	0:7189	3:14647	0.055	
NECAP2	1	16767166	16786584	2081	1:7458	1:7189	2:14647	1.000	
HSPG2	1	22148736	22263750	14288	3:7458	6:7189	9:14647	0.273	
CELA3B	1	22303417	22315847	1003	6:7458	9:7189	15:14647	0.322	
SLC9A1	1	27425299	27481621	4742	1:7458	1:7189	2:14647	1.000	
FAM76A	1	28052489	28089423	3459	1:7458	1:7189	2:14647	1.000	
THEMIS2	1	28199054	28213193	2720	2:7458	0:7189	2:14647	0.109	
SMPDL3B	1	28261503	28285663	1873	3:7458	0:7189	3:14647	0.055	
EYA3	1	28296854	28415148	6026	1:7458	1:7189	2:14647	1.000	
ATPIF1	1	28562601	28564616	1925	2:7458	0:7189	2:14647	0.113	
SESN2	1	28585962	28609002	3528	2:7458	1:7189	3:14647	0.837	
HCRTR1	1	32083300	32092919	2001	2:7458	1:7189	3:14647	1.000	
PEF1	1	32095462	32110838	1979	3:7458	0:7189	3:14647	0.052	
GRIK3	1	37261127	37499844	9219	2:7458	0:7189	2:14647	0.105	
TESK2	1	45809554	45956840	3076	1:7458	1:7189	2:14647	1.000	
FGGY	1	59762624	60228402	2086	5:7458	3:7189	8:14647	0.717	
CLCA4	1	87012758	87046432	3208	1:7458	1:7189	2:14647	1.000	
BRDT	1	92414927	92479985	3703	0:7458	2:7189	2:14647	0.081	
VAV3	1	108113781	108507545	4769	1:7458	2:7189	3:14647	0.502	
GNAT2	1	110145888	110155705	1365	3:7458	3:7189	6:14647	1.000	
OVGP1	1	111956936	111970399	2242	2:7458	1:7189	3:14647	0.677	
LRIG2	1	113615830	113667342	4015	2:7458	0:7189	2:14647	0.111	
DENND2C	1	115127195	115212732	4514	4:7458	0:7189	4:14647	0.023	
AMPD1	1	115215719	115238239	2406	2:7458	1:7189	3:14647	0.809	
NRAS	1	115247084	115259515	4454	1:7458	1:7189	2:14647	1.000	
PDE4DIP	1	144851426	144995022	8363	3:7458	6:7189	9:14647	0.271	
PEX11B	1	145516164	145523732	1829	1:7458	2:7189	3:14647	0.761	
ITGA10	1	145524989	145543868	5170	2:7458	2:7189	4:14647	1.000	
ANKRD35	1	145549208	145568526	3363	2:7458	1:7189	3:14647	1.000	
PIAS3	1	145575987	145586546	2902	1:7458	1:7189	2:14647	1.000	
POLR3C	1	145592604	145610884	1851	0:7458	2:7189	2:14647	0.081	
CD160	1	145695797	145715565	1559	2:7458	2:7189	4:14647	1.000	
PRKAB2	1	146626684	146644129	5423	4:7458	4:7189	8:14647	1.000	
CHD1L	1	146714290	146767447	3014	5:7458	3:7189	8:14647	1.000	
BCL9	1	147013181	147098015	6276	4:7458	3:7189	7:14647	1.000	
GJA8	1	147374945	147381395	1366	5:7458	4:7189	9:14647	1.000	
IGSF8	1	160061128	160068439	2304	5:7458	0:7189	5:14647	0.013	
ATP1A2	1	160085519	160113374	5457	2:7458	0:7189	2:14647	0.107	
F11R	1	160965000	160991133	4830	2:7458	1:7189	3:14647	0.692	
USF1	1	161009040	161015757	1797	2:7458	0:7189	2:14647	0.115	
FMO4	1	171283485	171311223	2139	1:7458	1:7189	2:14647	1.000	
RGL1	1	183605207	183897666	5113	0:7458	2:7189	2:14647	0.084	
RNF2	1	185014550	185071740	3551	0:7458	2:7189	2:14647	0.081	
DENND1B	1	197473878	197744623	8377	1:7458	4:7189	5:14647	0.141	
USH2A	1	215796235	216596738	18883	2:7458	7:7189	9:14647	0.073	
SPATA17	1	217804694	218040484	1235	0:7458	2:7189	2:14647	0.069	
RNU5F-1	1	220046618	220292777	72	3:7458	1:7189	4:14647	1.000	
DNAH14	1	225117355	225586996	13788	3:7458	0:7189	3:14647	0.046	
EPHX1	1	226013001	226033262	1831	1:7458	1:7189	2:14647	1.000	
PRSS38	1	228003417	228034171	1243	0:7458	2:7189	2:14647	0.070	
COG2	1	230778201	230829731	2960	1:7458	1:7189	2:14647	1.000	
DISC1	1	231762560	232177019	7156	1:7458	2:7189	3:14647	0.515	
COA6	1	234509213	234519795	698	1:7458	1:7189	2:14647	1.000	
RBM34	1	235294497	235324571	1853	1:7458	1:7189	2:14647	1.000	
RYR2	1	237205701	237997288	16365	2:7458	2:7189	4:14647	0.955	
PLD5	1	242251688	242687998	3305	0:7458	2:7189	2:14647	0.080	
ZNF670-ZNF(	1	247108848	247242115	1119	2:7458	2:7189	4:14647	1.000	
OR11L1	1	248004229	248005198	969	1:7458	1:7189	2:14647	1.000	
TRIM58	1	248020500	248043438	5156	2:7458	2:7189	4:14647	1.000	
OR2W3	1	248058888	248059833	945	0:7458	2:7189	2:14647	0.072	
OR2L13	1	248100492	248264224	1884	1:7458	2:7189	3:14647	0.732	
OR2L3	1	248223983	248224922	939	1:7458	2:7189	3:14647	0.756	
OR2M7	1	248486931	248487870	939	1:7458	1:7189	2:14647	1.000	
FAM110C	2	38813	46588	3963	3:7458	1:7189	4:14647	0.501	
SH3YL1	2	218135	264068	1788	1:7458	2:7189	3:14647	0.641	
SNTG2	2	946553	1371384	1887	0:7458	2:7189	2:14647	0.075	
ALLC	2	3705785	3750260	1444	2:7458	2:7189	4:14647	1.000	
C2orf43	2	20884817	21022827	2856	1:7458	1:7189	2:14647	1.000	
PLB1	2	28718937	28866653	5147	8:7458	6:7189	14:14647	0.739	
EHD3	2	31456879	31491260	3906	1:7458	2:7189	3:14647	0.505	
BIRC6	2	32582095	32843965	15702	2:7458	2:7189	4:14647	1.000	
TTC27	2	32853086	33046118	2918	3:7458	1:7189	4:14647	0.396	
CRIM1	2	36583369	36778278	5611	0:7458	2:7189	2:14647	0.085	
PKDCC	2	42275160	42285668	2501	0:7458	2:7189	2:14647	0.080	
THADA	2	43457974	43823113	6326	2:7458	6:7189	8:14647	0.118	
PLEKHH2	2	43864438	43995126	6954	2:7458	5:7189	7:14647	0.210	
LOC728819	2	43902291	43903461	1170	1:7458	2:7189	3:14647	1.000	
PPM1B	2	44395999	44461742	3850	2:7458	0:7189	2:14647	0.110	
SLC3A1	2	44502596	44547962	2320	3:7458	2:7189	5:14647	0.804	
STON1	2	48757307	48825654	5633	0:7458	2:7189	2:14647	0.086	
STON1-GTF2(	2	48796158	48906748	3830	4:7458	1:7189	5:14647	0.212	
GTF2A1L	2	48844936	48906749	1682	2:7458	0:7189	2:14647	0.119	
LHCGR	2	48913912	48982880	3093	3:7458	0:7189	3:14647	0.052	
FSHR	2	49189295	49381666	2774	3:7458	3:7189	6:14647	1.000	
NRXN1	2	50145642	51259674	9560	3:7458	1:7189	4:14647	0.340	NRXN
ASB3	2	53897116	54013985	2448	3:7458	4:7189	7:14647	0.603	
ACYP2	2	54342409	54532435	1222	0:7458	2:7189	2:14647	0.068	
NAT8	2	73867849	73869537	1073	4:7458	1:7189	5:14647	0.296	
MRPL19	2	75873908	75889334	7827	1:7458	1:7189	2:14647	1.000	
GCFC2	2	75889831	75938111	4436	1:7458	2:7189	3:14647	0.624	
FAHD2B	2	97749322	97760582	1265	3:7458	5:7189	8:14647	0.504	
ANKRD36B	2	98121260	98206428	5986	2:7458	2:7189	4:14647	1.000	
AFF3	2	100163715	100759037	8137	0:7458	2:7189	2:14647	0.087	
ST6GAL2	2	107418055	107503563	6857	1:7458	1:7189	2:14647	1.000	
SLC5A7	2	108602994	108630443	5145	1:7458	1:7189	2:14647	1.000	
SULT1C3	2	108863650	108881807	915	2:7458	0:7189	2:14647	0.127	
RANBP2	2	109335936	109402267	11711	2:7458	2:7189	4:14647	1.000	

CCDC138	2	109403218	109492847	2196	2:7458	1:7189	3:14647	1.000	
EDAR	2	109510926	109605828	4214	1:7458	1:7189	2:14647	1.000	
NPHP1	2	110880913	110962639	2750	16:7458	25:7189	41:14647	0.130	
MERTK	2	112656190	112786945	3626	2:7458	5:7189	7:14647	0.194	
GTDC1	2	144703580	145052060	2694	2:7458	0:7189	2:14647	0.115	
MBD5	2	148778579	149271044	6016	1:7458	1:7189	2:14647	0.974	
LY75	2	160659867	160761267	6930	2:7458	3:7189	5:14647	0.591	
SCN7A	2	167260082	167343481	7280	2:7458	3:7189	5:14647	0.591	
ABCB11	2	169779448	169887833	4775	4:7458	1:7189	5:14647	0.203	
LRP2	2	169983618	170219122	15735	1:7458	1:7189	2:14647	1.000	
MYO3B	2	171034654	171511674	6334	1:7458	1:7189	2:14647	0.964	
C2orf47	2	200820039	200828847	1511	1:7458	2:7189	3:14647	0.471	
MDH1B	2	207602488	207630050	2330	2:7458	0:7189	2:14647	0.115	
C2orf80	2	209030070	209054773	1214	4:7458	8:7189	12:14647	0.125	
PIKFYVE	2	209130990	209223475	9901	3:7458	3:7189	6:14647	0.963	
MOGAT1	2	223536456	223574649	1096	4:7458	1:7189	5:14647	0.326	
DIS3L2	2	232826292	233201908	4002	3:7458	0:7189	3:14647	0.050	
ALPP	2	233243347	233247599	2867	4:7458	1:7189	5:14647	0.220	
ALPPL2	2	233271551	233275424	2494	6:7458	1:7189	7:14647	0.072	
CHRNG	2	233404436	233411038	2187	2:7458	2:7189	4:14647	1.000	
UGT1A9	2	234580543	234681951	2376	1:7458	8:7189	9:14647	0.010	
UGT1A7	2	234590583	234681945	2333	1:7458	8:7189	9:14647	0.010	
UGT1A6	2	234601511	234681951	2484	1:7458	6:7189	7:14647	0.043	
UGT1A5	2	234621637	234681945	2345	1:7458	6:7189	7:14647	0.043	
UGT1A4	2	234627437	234681945	2374	1:7458	3:7189	4:14647	0.409	
UGT1A3	2	234637772	234681945	2345	1:7458	3:7189	4:14647	0.411	
HJURP	2	234745485	234763212	3050	3:7458	0:7189	3:14647	0.049	
TRPM8	2	234826042	234928166	5621	2:7458	0:7189	2:14647	0.106	
AGAP1	2	236402732	237040444	10850	1:7458	1:7189	2:14647	1.000	
ASB18	2	237103514	237172988	1401	2:7458	0:7189	2:14647	0.118	
IQCA1	2	237232789	237416178	3343	1:7458	1:7189	2:14647	1.000	
CNTN6	3	1134628	1445278	3530	2:7458	3:7189	5:14647	0.579	
CNTN4	3	2280512	3099645	5206	2:7458	1:7189	3:14647	0.698	
RAD18	3	8918879	9005159	5739	1:7458	2:7189	3:14647	0.546	
TBC1D5	3	17198653	17782399	7854	3:7458	2:7189	5:14647	0.725	
EFHB	3	19920965	19975706	2855	4:7458	1:7189	5:14647	0.218	
ZCWPW2	3	28431986	28566632	1659	0:7458	3:7189	3:14647	0.027	
TRANK1	3	36868307	36986548	10484	11:7458	9:7189	20:14647	0.757	
SPINK8	3	48348335	48369831	440	1:7458	1:7189	2:14647	1.000	
FBXW12	3	48413708	48436190	1647	1:7458	2:7189	3:14647	0.719	
PLXNB1	3	48445260	48470872	7310	0:7458	3:7189	3:14647	0.036	
ATRIP	3	48488496	48507054	2660	0:7458	2:7189	2:14647	0.083	
SHISA5	3	48509196	48541661	2166	0:7458	2:7189	2:14647	0.082	
C3orf67	3	58727736	59035715	2551	0:7458	2:7189	2:14647	0.078	
FHIT	3	59735035	61237133	1090	2:7458	3:7189	5:14647	0.453	
MIR548G	3	99273152	99717059	83	3:7458	2:7189	5:14647	0.641	
CMSS1	3	99536677	99897476	1230	2:7458	2:7189	4:14647	1.000	
FILIP1L	3	99566771	99833349	4211	2:7458	2:7189	4:14647	1.000	
TBC1D23	3	99979660	100044096	3886	1:7458	1:7189	2:14647	1.000	
NIT2	3	100053561	100074478	1271	1:7458	2:7189	3:14647	1.000	
TOMM70A	3	100082302	100120242	4381	2:7458	0:7189	2:14647	0.106	
LNP1	3	100120036	100175170	2217	1:7458	1:7189	2:14647	1.000	
GPR128	3	100328432	100414323	3140	1:7458	1:7189	2:14647	1.000	
TFG	3	100428174	100467811	2073	2:7458	0:7189	2:14647	0.112	
ABI3BP	3	100468178	100712334	4473	2:7458	1:7189	3:14647	0.904	
IMPG2	3	100941389	101039419	8352	1:7458	1:7189	2:14647	1.000	
SENP7	3	101043117	101232085	4933	2:7458	0:7189	2:14647	0.105	
TRMT10C	3	101280679	101285290	1845	2:7458	0:7189	2:14647	0.113	
CEP97	3	101443493	101486181	4411	2:7458	1:7189	3:14647	0.906	
NXPE3	3	101498285	101547075	8913	2:7458	0:7189	2:14647	0.103	
ZPLD1	3	102153858	102198685	3619	1:7458	1:7189	2:14647	1.000	
CBLB	3	105377108	105587887	3976	2:7458	0:7189	2:14647	0.108	
SIDT1	3	113251217	113348422	5002	2:7458	2:7189	4:14647	0.949	
HGD	3	120347014	120401418	2010	3:7458	1:7189	4:14647	1.000	
GTF2E1	3	120461557	120501916	3033	1:7458	1:7189	2:14647	1.000	
MYLK	3	123331142	123603149	7836	0:7458	2:7189	2:14647	0.087	
CCDC14	3	123632273	123680255	4182	1:7458	2:7189	3:14647	0.528	
KALRN	3	123813557	124440036	10807	2:7458	0:7189	2:14647	0.104	
NEK11	3	130745693	131069309	2953	1:7458	2:7189	3:14647	0.634	
LOC339874	3	131043935	131100319	1978	1:7458	1:7189	2:14647	1.000	
SLC9A9	3	142984063	143567373	3627	1:7458	1:7189	2:14647	1.000	
LRRIQ4	3	169539709	169555560	1824	2:7458	0:7189	2:14647	0.118	
SLC2A2	3	170714136	170744768	3439	1:7458	1:7189	2:14647	1.000	
LPP	3	187930720	188608460	18278	1:7458	1:7189	2:14647	1.000	
CCDC50	3	191046873	191116459	8949	2:7458	1:7189	3:14647	0.647	
ZDHHC19	3	195924322	195938300	1337	3:7458	1:7189	4:14647	1.000	3q29
SLC51A	3	195943382	195960301	1455	3:7458	0:7189	3:14647	0.051	3q29
PCYT1A	3	195965252	196014584	1583	3:7458	1:7189	4:14647	0.935	3q29
TCTEX1D2	3	196018089	196045165	679	2:7458	1:7189	3:14647	1.000	3q29
TM4SF19-TC	3	196042955	196065291	1008	2:7458	2:7189	4:14647	1.000	3q29
TM4SF19	3	196050416	196065291	1059	2:7458	1:7189	3:14647	1.000	3q29
RNF168	3	196195656	196230639	5344	4:7458	0:7189	4:14647	0.022	3q29
PAK2	3	196466727	196559518	6139	3:7458	0:7189	3:14647	0.046	3q29
PIGZ	3	196673213	196695704	2701	3:7458	0:7189	3:14647	0.048	3q29
MFI2	3	196728611	196756687	3962	3:7458	1:7189	4:14647	0.747	3q29
DLG1	3	196769430	197025447	5034	3:7458	0:7189	3:14647	0.046	3q29
LRCH3	3	197518144	197598456	2258	1:7458	2:7189	3:14647	0.558	
IQCG	3	197615945	197686886	2258	1:7458	3:7189	4:14647	0.269	
ZNF718	4	53276	156490	2121	4:7458	2:7189	6:14647	0.571	
ABCA11P	4	419223	467998	1895	2:7458	1:7189	3:14647	0.742	
ZNF721	4	433778	493442	5168	2:7458	1:7189	3:14647	0.648	
TRMT44	4	8442531	8478282	2846	2:7458	2:7189	4:14647	1.000	
GPR78	4	8582216	8591750	4245	2:7458	1:7189	3:14647	0.797	
CPZ	4	8594386	8621488	2540	3:7458	3:7189	6:14647	1.000	
HS3ST1	4	11399987	11430537	1965	1:7458	1:7189	2:14647	1.000	
TLR1	4	38797875	38806412	2851	1:7458	3:7189	4:14647	0.251	
FAM114A1	4	38869353	38947365	4138	3:7458	1:7189	4:14647	0.374	
TMEM156	4	38968440	39034041	1859	2:7458	2:7189	4:14647	1.000	
N4BP2	4	40058523	40159872	9666	1:7458	2:7189	3:14647	0.550	
ATP8A1	4	42410391	42659122	8268	2:7458	2:7189	4:14647	0.939	
FRYL	4	48499379	48782316	11706	1:7458	2:7189	3:14647	0.523	
NMU	4	56461397	56502465	816	1:7458	1:7189	2:14647	1.000	
EXOC1	4	56719815	56771244	3582	2:7458	2:7189	4:14647	1.000	
CEP135	4	56814973	56899529	5735	1:7458	2:7189	3:14647	0.555	
TECRL	4	65144176	65275178	2814	1:7458	2:7189	3:14647	0.476	
CABS1	4	71200670	71202833	1544	0:7458	2:7189	2:14647	0.076	
SMR3A	4	71226492	71232823	598	1:7458	1:7189	2:14647	1.000	

PAQR3	4	79839093	79860582	3739	3:7458	0:7189	3:14647	0.050
C4orf22	4	81256873	81884910	946	1:7458	2:7189	3:14647	0.423
MMRN1	4	90816051	90875780	4969	3:7458	1:7189	4:14647	0.372
STPG2	4	98480024	99064391	1654	5:7458	3:7189	8:14647	0.697
RAP1GDS1	4	99182526	99365012	3761	4:7458	1:7189	5:14647	0.208
EMCN	4	101316497	101439250	4037	2:7458	0:7189	2:14647	0.107
BANK1	4	102711763	102995969	3445	2:7458	0:7189	2:14647	0.108
TBCK	4	106967232	107237423	3541	1:7458	2:7189	3:14647	0.560
COL25A1	4	109731876	110223799	5702	1:7458	2:7189	3:14647	0.507
PDE5A	4	120415549	120549981	6989	1:7458	1:7189	2:14647	1.000
KIAA1109	4	123091757	123283914	15575	5:7458	3:7189	8:14647	0.535
ADAD1	4	123300120	123350947	2016	2:7458	0:7189	2:14647	0.113
NUDT6	4	123813798	123843761	1169	1:7458	1:7189	2:14647	1.000
SPATA5	4	123844224	124240604	8140	1:7458	3:7189	4:14647	0.282
INTU	4	128554086	128637934	3275	4:7458	2:7189	6:14647	0.613
NPY2R	4	156129780	156138228	3626	1:7458	1:7189	2:14647	1.000
MARCH1	4	164445449	165304407	5502	3:7458	0:7189	3:14647	0.049
DCTD	4	183811243	183838630	2042	0:7458	3:7189	3:14647	0.031
WWC2	4	184020462	184241929	8808	1:7458	2:7189	3:14647	0.566
CCDC110	4	186366337	186392913	2896	2:7458	1:7189	3:14647	1.000
PDLIM3	4	186421814	186456712	2852	1:7458	2:7189	3:14647	0.574
TRIML2	4	189012426	189026408	1300	2:7458	1:7189	3:14647	0.791
TRIML1	4	189060597	189068649	1645	2:7458	0:7189	2:14647	0.118
CDH18	5	19473154	19988353	3087	0:7458	2:7189	2:14647	0.082
RANBP3L	5	36249103	36302011	2618	1:7458	1:7189	2:14647	1.000
MRPS30	5	44809026	44815618	1672	3:7458	1:7189	4:14647	0.433
ITGA1	5	52084135	52249485	4796	2:7458	1:7189	3:14647	0.626
SKIV2L2	5	54603575	54721409	4204	0:7458	2:7189	2:14647	0.083
RNF180	5	63461670	63668696	4944	3:7458	0:7189	3:14647	0.049
MAST4	5	65892175	66465423	10724	1:7458	3:7189	4:14647	0.276
RAD17	5	68666531	68710628	3164	3:7458	1:7189	4:14647	0.373
MARVELD2	5	68710938	68737890	2160	3:7458	0:7189	3:14647	0.055
IQGAP2	5	75699148	76003957	5769	4:7458	5:7189	9:14647	0.661
XRCC4	5	82373316	82649579	1698	1:7458	2:7189	3:14647	0.457
SLCO4C1	5	101569691	101632253	5332	1:7458	2:7189	3:14647	0.560
MAN2A1	5	109025155	109203429	5128	1:7458	1:7189	2:14647	1.000
TSLP	5	110405777	110413722	3834	1:7458	1:7189	2:14647	1.000
FAM170A	5	118965253	118971517	1375	2:7458	1:7189	3:14647	1.000
PCDHA5	5	140201360	140391929	5218	6:7458	2:7189	8:14647	0.252
PCDHA6	5	140207649	140391929	5287	6:7458	2:7189	8:14647	0.250
PCDHA7	5	140213968	140391929	5221	6:7458	2:7189	8:14647	0.252
PCDHA8	5	140220906	140391929	5260	6:7458	2:7189	8:14647	0.251
PCDHA9	5	140227356	140233744	6388	6:7458	2:7189	8:14647	0.235
PCDHA10	5	140235633	140391929	5254	5:7458	2:7189	7:14647	0.415
SPINK5	5	147443534	147516925	3729	2:7458	3:7189	5:14647	0.561
GALNT10	5	153570294	153800543	5955	0:7458	2:7189	2:14647	0.085
RANBP17	5	170288895	170727019	4566	6:7458	2:7189	8:14647	0.196
RGS14	5	176784843	176799599	2406	0:7458	3:7189	3:14647	0.034
F12	5	176829138	176836577	2051	1:7458	1:7189	2:14647	1.000
GCNT2	6	10555948	10629601	4672	2:7458	2:7189	4:14647	1.000
GPLD1	6	24426061	24489850	5832	3:7458	1:7189	4:14647	0.369
FAM65B	6	24804512	24911195	5471	1:7458	1:7189	2:14647	1.000
BTN3A3	6	26440699	26453643	3002	1:7458	5:7189	6:14647	0.094
BTN2A1	6	26458152	26469866	3150	1:7458	4:7189	5:14647	0.256
TRIM40	6	30104509	30116512	1917	2:7458	0:7189	2:14647	0.116
TRIM15	6	30130982	30140473	2224	3:7458	0:7189	3:14647	0.055
APOM	6	31623670	31625987	761	1:7458	1:7189	2:14647	1.000
LY6G5B	6	31638727	31640227	774	1:7458	1:7189	2:14647	1.000
LY6G6F	6	31674683	31678372	898	1:7458	1:7189	2:14647	1.000
MSH5	6	31707724	31730455	2938	4:7458	0:7189	4:14647	0.025
VWA7	6	31733370	31745108	2914	1:7458	1:7189	2:14647	1.000
NEU1	6	31826828	31830709	2071	2:7458	0:7189	2:14647	0.112
DNAH8	6	38683116	38998574	14639	7:7458	7:7189	14:14647	0.914
SUPT3H	6	44794466	45345788	4509	3:7458	3:7189	6:14647	0.912
GSTA2	6	52614884	52628361	1309	1:7458	2:7189	3:14647	0.489
COL21A1	6	55921387	56112378	4173	3:7458	0:7189	3:14647	0.050
RNU6-71	6	56137688	56615138	82	6:7458	3:7189	9:14647	1.000
KIAA1586	6	56911383	56920023	2896	1:7458	3:7189	4:14647	0.356
BAI3	6	69345631	70099403	5637	2:7458	0:7189	2:14647	0.107
PHIP	6	79644135	79788011	11966	2:7458	0:7189	2:14647	0.103
LCA5	6	80194707	80247147	4719	1:7458	1:7189	2:14647	1.000
MDN1	6	90353230	90529442	17400	3:7458	2:7189	5:14647	0.818
CASP8AP2	6	90539618	90584155	6763	3:7458	0:7189	3:14647	0.047
GJA10	6	90604187	90605819	1632	2:7458	1:7189	3:14647	1.000
ASCC3	6	100956607	101329224	7592	6:7458	4:7189	10:14647	0.612
PDSS2	6	107473760	107780779	3552	3:7458	2:7189	5:14647	0.800
LAMA2	6	129204285	129837710	9692	3:7458	3:7189	6:14647	1.000
AIG1	6	143382022	143661441	1385	1:7458	1:7189	2:14647	1.000
UTRN	6	144612872	145174170	12431	4:7458	2:7189	6:14647	0.455
SHPRH	6	146205944	146285233	7333	4:7458	1:7189	5:14647	0.194
SMOC2	6	168841830	169068674	3150	1:7458	3:7189	4:14647	0.258
MAD1L1	7	1855427	2272583	2738	2:7458	4:7189	6:14647	0.412
NUDT1	7	2281856	2290780	786	1:7458	4:7189	5:14647	0.284
IQCE	7	2598631	2654368	6844	2:7458	3:7189	5:14647	0.655
RADIL	7	4838739	4923335	3670	2:7458	1:7189	3:14647	0.696
MMD2	7	4945619	4998844	2416	3:7458	1:7189	4:14647	0.391
RBAK-LOC38	7	5085451	5112854	858	0:7458	3:7189	3:14647	0.023
RBAK	7	5085552	5109119	6551	0:7458	3:7189	3:14647	0.036
RNF216	7	5659671	5821361	5850	1:7458	1:7189	2:14647	1.000
PMS2	7	6012869	6048737	2836	4:7458	1:7189	5:14647	0.218
AIMP2	7	6048881	6063465	1219	1:7458	2:7189	3:14647	1.000
EIF2AK1	7	6061877	6098860	4465	1:7458	2:7189	3:14647	1.000
COL28A1	7	7398243	7575460	3515	2:7458	4:7189	6:14647	0.349
SCIN	7	12610202	12693228	3246	3:7458	1:7189	4:14647	0.375
DGKB	7	14184673	14881075	6684	2:7458	0:7189	2:14647	0.107
AGMO	7	15239942	15601640	2475	1:7458	3:7189	4:14647	0.268
AGR3	7	16899029	16921613	750	1:7458	1:7189	2:14647	0.774
ABCB5	7	20655244	20796637	5401	4:7458	4:7189	8:14647	0.901
SP4	7	21467688	21554151	5800	4:7458	0:7189	4:14647	0.022
DNAH11	7	21582832	21941457	14188	6:7458	4:7189	10:14647	0.594
CDCA7L	7	21940516	21985542	2962	0:7458	2:7189	2:14647	0.083
HIBADH	7	27565058	27702620	1998	0:7458	2:7189	2:14647	0.074
ABCA13	7	48211056	48687091	17184	5:7458	8:7189	13:14647	0.352
ZPBP	7	49977023	50132860	1226	1:7458	1:7189	2:14647	1.000
ZNF273	7	64363619	64391955	4414	1:7458	1:7189	2:14647	1.000
ZNF117	7	64434829	64451414	6404	6:7458	1:7189	7:14647	0.064
ZNF92	7	64838767	64865998	3114	18:7458	16:7189	34:14647	1.000

AUTS2	7	69063905	70257885	6417	2:7458	1:7189	3:14647	0.614
POM121	7	72349935	72421979	7011	0:7458	2:7189	2:14647	0.088
HIP1	7	75162618	75368283	8033	2:7458	0:7189	2:14647	0.104
CCL24	7	75441113	75443033	360	2:7458	0:7189	2:14647	0.184
POR	7	75544419	75616173	2499	2:7458	2:7189	4:14647	1.000
SRCRB4D	7	76018645	76039012	2806	2:7458	0:7189	2:14647	0.109
ZP3	7	76026840	76071388	1477	4:7458	0:7189	4:14647	0.030
UPK3B	7	76139744	76157199	1609	2:7458	1:7189	3:14647	1.000
CCDC146	7	76751933	76924521	3333	1:7458	1:7189	2:14647	1.000
RSBN1L	7	77325742	77409120	3220	2:7458	0:7189	2:14647	0.111
MAGI2	7	77646373	79082890	6880	4:7458	0:7189	4:14647	0.022
CD36	7	80267831	80308593	4727	4:7458	1:7189	5:14647	0.242
PILRB	7	99933687	99965454	3683	1:7458	1:7189	2:14647	1.000
PILRA	7	99971067	99997722	1316	2:7458	0:7189	2:14647	0.123
ZCWPW1	7	99998494	100026431	2466	1:7458	1:7189	2:14647	1.000
FBXL13	7	102453307	102715015	3116	3:7458	1:7189	4:14647	0.379
COG5	7	106848297	107204959	5666	4:7458	1:7189	5:14647	0.241
GPR22	7	107110501	107116125	2966	3:7458	0:7189	3:14647	0.053
TAS2R16	7	122634758	122635754	996	1:7458	2:7189	3:14647	0.475
GRM8	7	126078651	126892428	3860	2:7458	1:7189	3:14647	0.678
LRGUK	7	133812104	133948933	2700	1:7458	2:7189	3:14647	0.526
AKR1B1	7	134127106	134143888	1398	3:7458	1:7189	4:14647	0.432
AGK	7	141251077	141354209	2925	3:7458	1:7189	4:14647	0.392
PRSS37	7	141536077	141541221	1197	0:7458	2:7189	2:14647	0.071
KEL	7	142638200	142659503	2547	7:7458	7:7189	14:14647	1.000
GALNTL5	7	151653463	151717019	1984	3:7458	0:7189	3:14647	0.059
DPP6	7	153584418	154685995	4561	2:7458	1:7189	3:14647	0.638
PTPRN2	7	157331749	158380482	4827	1:7458	2:7189	3:14647	0.518
VIPR2	7	158820865	158937649	3944	2:7458	1:7189	3:14647	0.635
C8orf42	8	439789	495781	3423	0:7458	2:7189	2:14647	0.081
ARHGEF10	8	1772148	1906807	5591	5:7458	2:7189	7:14647	0.301
MYOM2	8	1993157	2093380	5011	4:7458	5:7189	9:14647	0.634
SGK223	8	8175257	8239257	4627	3:7458	1:7189	4:14647	0.352
CLDN23	8	8559665	8561617	1952	2:7458	0:7189	2:14647	0.113
MFHAS1	8	8641998	8751131	5255	1:7458	2:7189	3:14647	1.000
ERI1	8	8860313	8890849	4615	3:7458	0:7189	3:14647	0.048
TNKS	8	9413444	9639856	9599	2:7458	0:7189	2:14647	0.103
MSRA	8	9953065	10286401	1708	2:7458	1:7189	3:14647	0.962
RP1L1	8	10463859	10512617	7977	6:7458	4:7189	10:14647	0.727
PINX1	8	10622883	10697299	1047	2:7458	3:7189	5:14647	0.495
FDFT1	8	11660189	11696818	2106	3:7458	1:7189	4:14647	0.389
CTSB	8	11700033	11725646	3944	2:7458	0:7189	2:14647	0.108
DEFB135	8	11839829	11842099	234	2:7458	1:7189	3:14647	0.982
MSR1	8	15965386	16050300	3761	9:7458	6:7189	15:14647	0.556
FGL1	8	17721899	17752913	1464	2:7458	0:7189	2:14647	0.121
LOXL2	8	23154409	23261722	3810	2:7458	1:7189	3:14647	0.637
FUT10	8	33228343	33330664	3547	1:7458	1:7189	2:14647	0.927
KIAA0146	8	48173488	48648563	3361	2:7458	1:7189	3:14647	0.643
NKAIN3	8	63161500	63903628	1566	1:7458	1:7189	2:14647	1.000
SLC10A5	8	82605890	82607207	1317	0:7458	2:7189	2:14647	0.075
VPS13B	8	100025493	100889814	14100	6:7458	2:7189	8:14647	0.173
KCNV1	8	110979232	110986959	2929	2:7458	0:7189	2:14647	0.112
CSMD3	8	113235158	114449242	13124	0:7458	3:7189	3:14647	0.037
TRAPPC9	8	140742585	141468678	4474	2:7458	3:7189	5:14647	0.575
PYCRL	8	144686082	144691784	2678	1:7458	2:7189	3:14647	0.538
PPP1R16A	8	145722108	145727504	2274	0:7458	2:7189	2:14647	0.082
GPT	8	145729464	145732555	1886	1:7458	2:7189	3:14647	0.680
RECQL4	8	145736666	145743210	3816	6:7458	3:7189	9:14647	0.522
DOCK8	9	214864	465259	7452	2:7458	4:7189	6:14647	0.361
SLC1A1	9	4490426	4587469	3739	0:7458	2:7189	2:14647	0.085
AK3	9	4709556	4741309	4325	1:7458	1:7189	2:14647	1.000
JAK2	9	4985244	5128183	5285	2:7458	2:7189	4:14647	1.000
INSL6	9	5163862	5185618	708	1:7458	1:7189	2:14647	1.000
INSL4	9	5231418	5233967	615	1:7458	1:7189	2:14647	1.000
RLN2	9	5299867	5304580	889	2:7458	0:7189	2:14647	0.138
KDM4C	9	6757640	7175648	4670	2:7458	2:7189	4:14647	1.000
MPDZ	9	13105702	13279563	7609	2:7458	3:7189	5:14647	0.591
CCDC171	9	15553096	15971897	4359	5:7458	3:7189	8:14647	0.586
IFNA7	9	21201467	21202204	737	3:7458	2:7189	5:14647	1.000
MIR31HG	9	21454266	21559697	2148	3:7458	2:7189	5:14647	1.000
IFNE	9	21480838	21482312	1474	2:7458	1:7189	3:14647	1.000
LINGO2	9	27948083	28719303	3236	11:7458	14:7189	25:14647	0.360
RECK	9	36036909	36124452	4412	1:7458	3:7189	4:14647	0.265
CNTNAP3	9	39072763	39288300	5229	10:7458	7:7189	17:14647	0.556
ZNF658	9	40771401	40792112	4025	3:7458	2:7189	5:14647	0.939
TMEM2	9	74298281	74383800	6523	1:7458	1:7189	2:14647	1.000
ECM2	9	95257593	95298374	3290	1:7458	2:7189	3:14647	0.584
ERCC6L2	9	98637899	98731122	4647	0:7458	4:7189	4:14647	0.014
AKAP2	9	112810877	112934791	7002	0:7458	3:7189	3:14647	0.036
ZNF883	9	115759399	115774472	2379	1:7458	3:7189	4:14647	0.260
TRIM32	9	119449580	119463579	3719	2:7458	1:7189	3:14647	1.000
OR1B1	9	125390857	125391814	957	0:7458	2:7189	2:14647	0.070
OR1L1	9	125423994	125424927	933	0:7458	2:7189	2:14647	0.069
NAIF1	9	130823511	130829599	3398	1:7458	1:7189	2:14647	1.000
ARRDC1	9	140500095	140509812	1598	1:7458	2:7189	3:14647	0.563
CACNA1B	9	140772240	141019076	9790	3:7458	0:7189	3:14647	0.047
PFKP	10	3110818	3178997	2781	0:7458	6:7189	6:14647	0.002
DHTKD1	10	12110933	12165224	5196	2:7458	3:7189	5:14647	0.681
UCMA	10	13263766	13276328	823	1:7458	2:7189	3:14647	0.355
DCLRE1C	10	14948870	14996094	4107	2:7458	1:7189	3:14647	0.640
MYO3A	10	26223001	26501465	5784	3:7458	1:7189	4:14647	0.352
ANKRD26	10	27293044	27389427	6779	3:7458	2:7189	5:14647	0.758
MARCH8	10	45952816	46090354	2589	4:7458	1:7189	5:14647	0.223
ZNF488	10	48355088	48373866	3496	0:7458	3:7189	3:14647	0.034
RBP3	10	48381486	48390991	4276	1:7458	1:7189	2:14647	1.000
GDF2	10	48413091	48416853	1936	1:7458	1:7189	2:14647	1.000
GDF10	10	48425787	48439138	2647	1:7458	1:7189	2:14647	1.000
FRMPD2	10	49364601	49482941	5024	0:7458	3:7189	3:14647	0.036
ARHGAP22	10	49654067	49864310	3297	0:7458	2:7189	2:14647	0.085
WDFY4	10	49893517	50191001	10013	0:7458	2:7189	2:14647	0.089
LRRC18	10	50117528	50122280	1619	0:7458	2:7189	2:14647	0.080
VSTM4	10	50222332	50323559	6354	1:7458	1:7189	2:14647	1.000
ERCC6	10	50664490	50747147	7006	0:7458	4:7189	4:14647	0.016
PGBD3	10	50723150	50732359	2298	0:7458	2:7189	2:14647	0.083
SLC18A3	10	50818346	50820766	2420	0:7458	3:7189	3:14647	0.033
CHAT	10	50822082	50873150	2528	0:7458	3:7189	3:14647	0.033
AGAP7	10	51464161	51486327	2420	1:7458	1:7189	2:14647	1.000

NCOA4	10	51565107	51590734	3738	0:7458	2:7189	2:14647	0.086
FAM21A	10	51827683	51893269	4672	0:7458	2:7189	2:14647	0.087
NRBF2	10	64893006	64914786	1896	3:7458	0:7189	3:14647	0.055
CTNNA3	10	67679724	69455949	3248	14:7458	15:7189	29:14647	0.641
UNC5B	10	72972291	73062635	6890	3:7458	0:7189	3:14647	0.047
SLC29A3	10	73079009	73123147	2273	1:7458	1:7189	2:14647	1.000
CDH23	10	73156690	73575704	11124	2:7458	7:7189	9:14647	0.073
CYP2C18	10	96443250	96495947	2545	1:7458	4:7189	5:14647	0.132
CYP2C19	10	96522462	96612671	1473	10:7458	6:7189	16:14647	0.565
ATRNL1	10	116853123	117708496	8732	0:7458	2:7189	2:14647	0.087
DHX32	10	127524908	127569884	3070	2:7458	2:7189	4:14647	1.000
FANK1	10	127585107	127698161	1296	2:7458	1:7189	3:14647	1.000
DOCK1	10	128594022	129250780	6751	2:7458	1:7189	3:14647	0.611
LRRC27	10	134145740	134195010	8205	1:7458	2:7189	3:14647	0.566
OR52M1	11	4566420	4567374	954	4:7458	3:7189	7:14647	1.000
SBF2	11	9800213	10315754	7439	0:7458	2:7189	2:14647	0.086
MICALCL	11	12308446	12380691	3044	0:7458	3:7189	3:14647	0.033
PARVA	11	12399025	12556903	8719	1:7458	1:7189	2:14647	1.000
ANO3	11	26353677	26684836	5936	0:7458	2:7189	2:14647	0.086
BBOX1	11	27062508	27149354	1886	4:7458	1:7189	5:14647	0.291
DCDC5	11	30885149	31014233	4759	4:7458	2:7189	6:14647	0.490
CCDC73	11	32623625	32816187	4015	4:7458	0:7189	4:14647	0.024
PAMR1	11	35453375	35547176	2785	4:7458	2:7189	6:14647	0.541
OR4A16	11	55110676	55111663	987	0:7458	3:7189	3:14647	0.024
OR5AK2	11	56756388	56757318	930	2:7458	0:7189	2:14647	0.148
MS4A10	11	60552820	60568778	2296	9:7458	0:7189	9:14647	0.001
CCDC86	11	60609428	60618561	2015	8:7458	1:7189	9:14647	0.048
EML3	11	62369690	62380237	3255	3:7458	0:7189	3:14647	0.050
ROM1	11	62380212	62382592	1878	2:7458	0:7189	2:14647	0.113
C11orf48	11	62430288	62439241	1320	1:7458	1:7189	2:14647	1.000
C11orf83	11	62439125	62441162	1950	1:7458	1:7189	2:14647	1.000
LRRN4CL	11	62453873	62457371	2585	1:7458	1:7189	2:14647	1.000
BSCL2	11	62457733	62477091	2102	2:7458	1:7189	3:14647	0.918
HNRNPUL2	11	62480096	62494857	5147	1:7458	1:7189	2:14647	1.000
TTC9C	11	62495951	62506108	1139	1:7458	3:7189	4:14647	1.000
SLC22A25	11	62931295	62997124	1644	2:7458	2:7189	4:14647	1.000
ATG2A	11	64662003	64684722	6373	2:7458	0:7189	2:14647	0.104
C11orf85	11	64705928	64727609	835	2:7458	0:7189	2:14647	0.130
ARL2-SNX15	11	64781584	64808044	2463	2:7458	1:7189	3:14647	1.000
NAALADL1	11	64812294	64826009	2687	3:7458	0:7189	3:14647	0.051
ZFPL1	11	64851693	64855874	1386	2:7458	0:7189	2:14647	0.117
VPS51	11	64863682	64879180	2510	1:7458	1:7189	2:14647	1.000
SPDYC	11	64937706	64940688	882	2:7458	0:7189	2:14647	0.128
CAPN1	11	64948685	64979477	3062	2:7458	2:7189	4:14647	1.000
POLA2	11	65029431	65065088	2477	1:7458	1:7189	2:14647	1.000
TIGD3	11	65122281	65125082	1966	3:7458	1:7189	4:14647	0.410
SLC25A45	11	65142662	65150142	2286	1:7458	1:7189	2:14647	1.000
LTBP3	11	65306029	65325699	4702	2:7458	4:7189	6:14647	1.000
SSSCA1	11	65337942	65339239	654	2:7458	4:7189	6:14647	1.000
EHBP1L1	11	65343508	65360116	5180	2:7458	6:7189	8:14647	0.129
PCNXL3	11	65383782	65404910	6566	2:7458	1:7189	3:14647	1.000
AP5B1	11	65541368	65548062	6595	1:7458	2:7189	3:14647	1.000
SNX32	11	65601409	65621172	1705	1:7458	1:7189	2:14647	1.000
EFEMP2	11	65633911	65640405	2068	2:7458	0:7189	2:14647	0.111
CTSW	11	65647283	65651212	1296	2:7458	0:7189	2:14647	0.119
SART1	11	65729159	65747607	3601	2:7458	0:7189	2:14647	0.107
SYTL2	11	85405264	85437511	4804	5:7458	1:7189	6:14647	0.113
NAALAD2	11	89867817	89925779	3196	2:7458	5:7189	7:14647	0.186
CCDC67	11	93063882	93171636	2666	3:7458	2:7189	5:14647	0.825
DYNC2H1	11	102980159	103350591	13699	4:7458	1:7189	5:14647	0.186
BCO2	11	112046207	112089649	2909	2:7458	2:7189	4:14647	1.000
CCDC15	11	124824016	124911385	3893	1:7458	1:7189	2:14647	1.000
SPATA19	11	133710516	133715392	866	2:7458	1:7189	3:14647	1.000
JAM3	11	133938819	134021652	3668	1:7458	2:7189	3:14647	1.000
NCAPD3	11	134022336	134094426	5605	1:7458	2:7189	3:14647	1.000
ACAD8	11	134123433	134135746	2201	1:7458	2:7189	3:14647	1.000
GLB1L3	11	134146274	134189458	2944	4:7458	2:7189	6:14647	0.488
GLB1L2	11	134201767	134246218	3155	5:7458	3:7189	8:14647	0.595
B3GAT1	11	134248397	134281812	3656	2:7458	1:7189	3:14647	1.000
RAD52	12	1021254	1058863	2673	4:7458	1:7189	5:14647	0.220
CACNA2D4	12	1901122	2027870	5343	2:7458	0:7189	2:14647	0.108
NCAPD2	12	6603297	6641132	4806	1:7458	1:7189	2:14647	1.000
GAPDH	12	6644408	6647537	1407	2:7458	0:7189	2:14647	0.118
IFFO1	12	6648693	6665249	2798	2:7458	0:7189	2:14647	0.109
CHD4	12	6679247	6716551	6497	1:7458	1:7189	2:14647	1.000
LPAR5	12	6728000	6745297	2855	2:7458	0:7189	2:14647	0.109
ACRBP	12	6747241	6756580	1886	2:7458	1:7189	3:14647	0.892
FOXJ2	12	8185358	8208118	5482	2:7458	0:7189	2:14647	0.106
SLCO1B7	12	21168629	21243040	1923	1:7458	1:7189	2:14647	1.000
ETNK1	12	22778075	22843608	7101	0:7458	2:7189	2:14647	0.087
IFLTD1	12	25649250	25801496	2489	2:7458	0:7189	2:14647	0.117
ITPR2	12	26488284	26986131	12568	3:7458	2:7189	5:14647	0.743
FGFR1OP2	12	27091304	27119581	3030	2:7458	0:7189	2:14647	0.109
MED21	12	27175454	27183606	2955	1:7458	1:7189	2:14647	1.000
STK38L	12	27397077	27478890	5096	2:7458	1:7189	3:14647	1.000
ARNTL2	12	27485786	27578746	7411	5:7458	0:7189	5:14647	0.010
PPFIBP1	12	27677044	27848497	6095	5:7458	3:7189	8:14647	0.656
TMTC1	12	29653745	29937692	8832	1:7458	2:7189	3:14647	0.520
ALG10	12	34175215	34181236	3127	1:7458	1:7189	2:14647	1.000
ADAMTS20	12	43748011	43945724	5733	2:7458	2:7189	4:14647	1.000
RDH16	12	57345215	57351418	1723	8:7458	11:7189	19:14647	0.580
SLC16A7	12	60083117	60183635	12052	1:7458	1:7189	2:14647	1.000
E2F7	12	77415025	77459360	5741	2:7458	0:7189	2:14647	0.104
PAWR	12	79985744	80084790	1951	2:7458	1:7189	3:14647	0.646
PPP1R12A	12	80167342	80328978	5726	2:7458	0:7189	2:14647	0.104
OTOGL	12	80603232	80772870	8083	4:7458	5:7189	9:14647	0.740
MYF5	12	81110707	81113447	1520	2:7458	0:7189	2:14647	0.116
ACSS3	12	81471808	81649582	3033	1:7458	4:7189	5:14647	1.000
TMTC2	12	83080933	83528067	4842	1:7458	2:7189	3:14647	1.000
SLC6A15	12	85253266	85306608	4832	2:7458	0:7189	2:14647	0.105
MGAT4C	12	86373036	87232681	1934	1:7458	3:7189	4:14647	0.224
DDX51	12	132621139	132628880	4717	2:7458	1:7189	3:14647	0.893
NOC4L	12	132628992	132636986	1632	1:7458	2:7189	3:14647	1.000
P2RX2	12	133195402	133198972	1908	1:7458	1:7189	2:14647	1.000
POLE	12	133200347	133263945	7840	3:7458	5:7189	8:14647	0.487
ANKLE2	12	133302253	133338451	4458	4:7458	1:7189	5:14647	0.205
GOLGA3	12	133345494	133405288	9252	2:7458	0:7189	2:14647	0.103

CHFR	12	133416937	133464204	3281	1:7458	1:7189	2:14647	1.000	
ZNF10	12	133707213	133736049	4403	4:7458	2:7189	6:14647	0.510	
ZNF268	12	133757994	133783697	5861	4:7458	0:7189	4:14647	0.022	
ZMYM5	13	20397623	20437776	3268	1:7458	1:7189	2:14647	1.000	
ZMYM2	13	20532809	20665968	10246	4:7458	0:7189	4:14647	0.022	
CRYL1	13	20977805	21100012	1501	2:7458	1:7189	3:14647	0.699	
SGCG	13	23755059	23899304	1655	1:7458	1:7189	2:14647	1.000	
SACS	13	23902964	24007841	15639	1:7458	3:7189	4:14647	0.701	
MIPEP	13	24304327	24463587	2429	2:7458	1:7189	3:14647	0.914	
C1QTNF9B-A	13	24463027	24466242	804	2:7458	0:7189	2:14647	0.132	
C1QTNF9B	13	24465427	24471125	1002	2:7458	0:7189	2:14647	0.125	
MTRF1	13	41790515	41837713	2174	3:7458	2:7189	5:14647	0.810	
SUCLA2	13	48516790	48575462	2164	2:7458	0:7189	2:14647	0.116	
DLEU2	13	50556687	50699677	2768	2:7458	0:7189	2:14647	0.109	
TRIM13	13	50571142	50592603	7290	2:7458	0:7189	2:14647	0.104	
DLEU7	13	51286758	51417885	1180	2:7458	0:7189	2:14647	0.121	
DLEU7-AS1	13	51381990	51424041	1807	2:7458	0:7189	2:14647	0.114	
RNASEH2B	13	51483813	51544596	1669	3:7458	0:7189	3:14647	0.056	
DIAPH3	13	60239722	60738119	4796	1:7458	5:7189	6:14647	0.070	
PCCA	13	100741268	101182691	2564	4:7458	0:7189	4:14647	0.027	
FGF14	13	102373204	103054124	2831	2:7458	0:7189	2:14647	0.111	
OR4K15	14	20443677	20444724	1047	1:7458	1:7189	2:14647	1.000	
OR4L1	14	20528203	20529142	939	1:7458	1:7189	2:14647	1.000	
OR4N5	14	20611894	20612821	927	1:7458	1:7189	2:14647	1.000	
OR11H6	14	20691868	20692861	993	1:7458	2:7189	3:14647	0.664	
DHRS4	14	24422964	24438488	1278	3:7458	0:7189	3:14647	0.060	
DHRS4L2	14	24439082	24475617	1575	2:7458	1:7189	3:14647	1.000	
SOS2	14	50583845	50698099	5313	2:7458	0:7189	2:14647	0.106	
ATP5S	14	50779046	50792946	1881	2:7458	0:7189	2:14647	0.114	
CDKL1	14	50796719	50862617	1177	6:7458	1:7189	7:14647	0.105	
NUMB	14	73741917	73925286	3644	0:7458	2:7189	2:14647	0.081	
EIF2B2	14	75469611	75476294	1541	1:7458	2:7189	3:14647	0.605	
MLH3	14	75480466	75518235	7896	1:7458	1:7189	2:14647	1.000	
NRXN3	14	78870092	80330760	6130	2:7458	1:7189	3:14647	0.621	
GALC	14	88399357	88459615	3883	7:7458	13:7189	20:14647	0.120	
RIN3	14	92980124	93155334	3847	1:7458	2:7189	3:14647	0.490	
TUBGCP5	15	22833394	22873891	3848	8:7458	11:7189	19:14647	1.000	
CYFIP1	15	22892683	23003603	4429	12:7458	8:7189	20:14647	1.000	
FAN1	15	31196075	31235310	4890	7:7458	2:7189	9:14647	0.332	15q13.3
TRPM1	15	31293263	31453476	6004	7:7458	3:7189	10:14647	0.236	15q13.3
OTUD7A	15	31775328	31947542	3042	5:7458	3:7189	8:14647	1.000	15q13.3
CHRNA7	15	32322685	32462384	3433	6:7458	3:7189	9:14647	0.535	15q13.3
STRC	15	43891760	43910998	5515	3:7458	4:7189	7:14647	0.716	
CATSPER2	15	43922771	43941039	1935	11:7458	15:7189	26:14647	0.358	
SPG11	15	44854893	44955876	7788	6:7458	2:7189	8:14647	0.181	
TRIM69	15	45028559	45060025	1901	0:7458	3:7189	3:14647	0.029	
CEP152	15	49030134	49103343	5637	4:7458	3:7189	7:14647	0.902	
SHC4	15	49115933	49255641	4556	2:7458	2:7189	4:14647	1.000	
FAM227B	15	49620591	49913118	1994	2:7458	5:7189	7:14647	0.164	
USP50	15	50792758	50838902	1392	1:7458	1:7189	2:14647	1.000	
TRPM7	15	50849351	50979012	10403	1:7458	4:7189	5:14647	0.166	
WDR72	15	53805937	54051859	7309	1:7458	1:7189	2:14647	1.000	
UNC13C	15	54305100	54920806	8140	2:7458	5:7189	7:14647	0.210	
NEDD4	15	56119121	56209329	7017	4:7458	0:7189	4:14647	0.022	
GCOM1	15	57884101	58009755	4662	1:7458	1:7189	2:14647	1.000	
TLN2	15	62939509	63136829	11649	0:7458	2:7189	2:14647	0.089	
MEGF11	15	66187633	66546075	5914	1:7458	3:7189	4:14647	0.265	
AAGAB	15	67493366	67547074	2845	3:7458	5:7189	8:14647	0.366	
IQCH	15	67547137	67794142	4239	2:7458	8:7189	10:14647	0.033	
FAM154B	15	82555151	82577267	3130	1:7458	2:7189	3:14647	0.587	
FSD2	15	83428023	83474806	2493	3:7458	1:7189	4:14647	0.386	
AKAP13	15	85923846	86292589	13329	1:7458	3:7189	4:14647	0.278	
BLM	15	91260578	91358686	4528	4:7458	3:7189	7:14647	0.824	
TTC23	15	99676527	99789815	3808	4:7458	2:7189	6:14647	0.525	
ADAMTS17	15	100511642	100882183	6331	0:7458	3:7189	3:14647	0.036	
TARSL2	15	102193954	102264645	3294	3:7458	2:7189	5:14647	0.945	
WDR90	16	699362	717829	5541	4:7458	3:7189	7:14647	1.000	
FBXL16	16	742499	755825	3543	1:7458	2:7189	3:14647	0.577	
CCDC78	16	772581	776473	1611	1:7458	2:7189	3:14647	0.717	
MSLN	16	810764	818865	2169	0:7458	2:7189	2:14647	0.081	
RPUSD1	16	834973	838383	2050	0:7458	2:7189	2:14647	0.081	
CHTF18	16	838621	848074	3090	0:7458	3:7189	3:14647	0.033	
GNG13	16	848040	850733	984	0:7458	2:7189	2:14647	0.072	
LMF1	16	903634	1020984	2912	2:7458	3:7189	5:14647	0.562	
RPL3L	16	1994579	2004679	1529	3:7458	1:7189	4:14647	0.425	
NDUFB10	16	2009516	2011976	697	1:7458	1:7189	2:14647	1.000	
TBL3	16	2022063	2028751	2594	2:7458	1:7189	3:14647	0.866	
NOXO1	16	2028917	2031550	1636	2:7458	0:7189	2:14647	0.115	
NTHL1	16	2089815	2097867	1067	1:7458	1:7189	2:14647	1.000	
PKD1	16	2138710	2185899	14138	1:7458	1:7189	2:14647	1.000	
SRRM2	16	2802329	2821413	9353	2:7458	0:7189	2:14647	0.103	
TCEB2	16	2821414	2827297	990	3:7458	0:7189	3:14647	0.067	
PRSS21	16	2867163	2871723	1168	2:7458	0:7189	2:14647	0.122	
ZG16B	16	2880172	2882285	828	2:7458	0:7189	2:14647	0.131	
FLYWCH1	16	2961979	3001209	4992	0:7458	2:7189	2:14647	0.086	
PKMYT1	16	3022791	3030540	2209	1:7458	2:7189	3:14647	1.000	
CLDN9	16	3062456	3064506	2050	1:7458	1:7189	2:14647	1.000	
THOC6	16	3074031	3077756	1420	1:7458	2:7189	3:14647	1.000	
CCDC64B	16	3077867	3085542	1811	2:7458	0:7189	2:14647	0.114	
MMP25	16	3096681	3110724	3551	1:7458	1:7189	2:14647	1.000	
IL32	16	3115312	3119668	1157	2:7458	0:7189	2:14647	0.122	
ZSCAN10	16	3138894	3142861	2463	2:7458	0:7189	2:14647	0.110	
OR1F1	16	3254246	3255185	939	2:7458	0:7189	2:14647	0.129	
ZNF200	16	3272324	3285186	3387	2:7458	0:7189	2:14647	0.109	
MEFV	16	3292027	3306627	3499	2:7458	0:7189	2:14647	0.109	
NMRAL1	16	4511694	4524896	1362	1:7458	3:7189	4:14647	1.000	
HMOX2	16	4526340	4560348	1849	2:7458	0:7189	2:14647	0.115	
ABAT	16	8806825	8878432	5586	2:7458	2:7189	4:14647	1.000	
TMEM186	16	8889036	8891505	1448	0:7458	2:7189	2:14647	0.076	
ATF7IP2	16	10479911	10577495	3665	2:7458	1:7189	3:14647	0.620	
EMP2	16	10622278	10674539	5181	1:7458	1:7189	2:14647	1.000	
TEKT5	16	10721360	10788802	1609	2:7458	1:7189	3:14647	0.783	
SNX29	16	12070601	12668146	8164	2:7458	2:7189	4:14647	1.000	
CPPED1	16	12753655	12897744	6143	0:7458	2:7189	2:14647	0.085	
PDXDC1	16	15068832	15131552	4011	4:7458	2:7189	6:14647	0.468	
NDE1	16	15737123	15820208	3934	1:7458	2:7189	3:14647	1.000	
FOPNL	16	15959576	15982447	2250	2:7458	1:7189	3:14647	1.000	

ABCC1	16	16043433	16236930	6563	2:7458	3:7189	5:14647	0.619
ABCC6	16	16243421	16317328	5117	3:7458	2:7189	5:14647	0.911
VWA3A	16	22103862	22168287	4600	1:7458	1:7189	2:14647	1.000
EEF2K	16	22217591	22300066	7412	1:7458	2:7189	3:14647	0.644
POLR3E	16	22308695	22346424	3731	1:7458	3:7189	4:14647	0.287
APOBR	16	28505969	28510291	3792	2:7458	1:7189	3:14647	1.000
SULT1A2	16	28603263	28607801	1324	4:7458	0:7189	4:14647	0.027
SULT1A1	16	28616907	28620649	1595	2:7458	1:7189	3:14647	1.000
ATP2A1	16	28889808	28915830	3536	8:7458	2:7189	10:14647	0.068
LOC1002890	16	28890277	28891242	584	5:7458	1:7189	6:14647	1.000
RABEP2	16	28915741	28936532	2280	5:7458	2:7189	7:14647	0.591
CD19	16	28943259	28950668	1958	4:7458	2:7189	6:14647	1.000
HIRIP3	16	30003641	30007417	3048	1:7458	2:7189	3:14647	1.000
INO80E	16	30007530	30017112	1185	1:7458	1:7189	2:14647	1.000
FAM57B	16	30035747	30042186	1919	2:7458	0:7189	2:14647	0.113
TBX6	16	30097114	30103205	1802	2:7458	0:7189	2:14647	0.114
GDPD3	16	30116130	30124878	1098	3:7458	0:7189	3:14647	0.064
ZNF48	16	30406432	30411429	3340	1:7458	1:7189	2:14647	1.000
ITGAL	16	30483982	30534506	5209	1:7458	2:7189	3:14647	1.000
ZNF688	16	30581018	30583728	1827	2:7458	0:7189	2:14647	0.113
ZNF785	16	30591993	30597092	3265	3:7458	0:7189	3:14647	0.050
ZNF689	16	30614699	30621682	2708	2:7458	0:7189	2:14647	0.109
PHKG2	16	30759619	30772497	5510	1:7458	1:7189	2:14647	1.000
C16orf93	16	30768743	30773565	1545	1:7458	1:7189	2:14647	1.000
BCL7C	16	30899115	30905399	858	1:7458	1:7189	2:14647	1.000
SETD1A	16	30968614	30995981	6447	3:7458	0:7189	3:14647	0.047
ZNF668	16	31072168	31084823	2865	2:7458	0:7189	2:14647	0.108
OGFOD1	16	56485423	56511407	3020	1:7458	2:7189	3:14647	0.633
BBS2	16	56518258	56554008	2814	2:7458	0:7189	2:14647	0.112
GPR56	16	57673206	57698944	4043	1:7458	2:7189	3:14647	0.864
COG8	16	69362523	69373526	2495	2:7458	0:7189	2:14647	0.110
TMED6	16	69377148	69385712	940	1:7458	1:7189	2:14647	1.000
NFAT5	16	69599868	69738569	13363	2:7458	0:7189	2:14647	0.102
NOB1	16	69775773	69788829	1716	1:7458	1:7189	2:14647	1.000
WWP2	16	69796186	69975644	4590	2:7458	0:7189	2:14647	0.106
PDPR	16	70147528	70195184	7999	8:7458	8:7189	16:14647	1.000
EXOSC6	16	70284133	70285833	1700	2:7458	0:7189	2:14647	0.117
HYDIN	16	70841286	71264625	15746	3:7458	0:7189	3:14647	0.046
ADAMTS18	16	77316024	77469011	5913	4:7458	4:7189	8:14647	1.000
WWOX	16	78133326	79246564	2465	2:7458	3:7189	5:14647	0.558
PKD1L2	16	81134483	81253975	7623	1:7458	3:7189	4:14647	0.294
PLCG2	16	81812898	81991899	4272	2:7458	1:7189	3:14647	0.627
CDH13	16	82660398	83830215	4127	1:7458	2:7189	3:14647	0.513
MLYCD	16	83932729	83949787	2195	1:7458	1:7189	2:14647	1.000
OSGIN1	16	83986826	83999937	1940	0:7458	2:7189	2:14647	0.081
NECAB2	16	84002236	84036379	1606	1:7458	2:7189	3:14647	0.751
SLC38A8	16	84043388	84075762	1308	0:7458	2:7189	2:14647	0.077
DNAAF1	16	84178864	84211524	2436	0:7458	2:7189	2:14647	0.083
TAF1C	16	84211452	84220676	3946	2:7458	1:7189	3:14647	1.000
ADAD2	16	84224722	84230772	2304	2:7458	2:7189	4:14647	1.000
KCNG4	16	84254740	84273356	2763	3:7458	1:7189	4:14647	1.000
WFDC1	16	84328400	84363450	1378	0:7458	2:7189	2:14647	0.078
ATP2C2	16	84402132	84497793	3385	4:7458	4:7189	8:14647	0.963
KIAA1609	16	84509965	84538288	5031	4:7458	1:7189	5:14647	0.248
CHMP1A	16	89710838	89724129	2527	1:7458	1:7189	2:14647	1.000
CDK10	16	89753075	89762772	1798	2:7458	1:7189	3:14647	0.837
SPATA2L	16	89762764	89768121	2339	1:7458	1:7189	2:14647	1.000
VPS9D1	16	89773540	89787394	2749	1:7458	1:7189	2:14647	1.000
ZNF276	16	89787951	89807332	4620	3:7458	2:7189	5:14647	1.000
FANCA	16	89803958	89883065	5460	5:7458	4:7189	9:14647	1.000
SPIRE2	16	89894906	89937727	3244	1:7458	2:7189	3:14647	0.535
RPH3AL	17	62179	202633	2754	2:7458	1:7189	3:14647	1.000
C17orf97	17	260117	264457	1839	3:7458	0:7189	3:14647	0.050
GEMIN4	17	647660	655501	3741	2:7458	3:7189	5:14647	1.000
GLOD4	17	662548	685571	1837	3:7458	1:7189	4:14647	0.769
RNMTL1	17	685512	695741	1801	4:7458	0:7189	4:14647	0.025
SHPK	17	3511555	3539616	3838	6:7458	3:7189	9:14647	0.394
CTNS	17	3539761	3566397	4431	5:7458	3:7189	8:14647	0.695
P2RX5-TAX1I	17	3566186	3599698	6285	2:7458	0:7189	2:14647	0.105
DNAH2	17	7623038	7737058	13505	0:7458	5:7189	5:14647	0.007
KDM6B	17	7743234	7758118	6704	1:7458	1:7189	2:14647	1.000
TMEM88	17	7758383	7759417	874	0:7458	2:7189	2:14647	0.071
LSMD1	17	7760002	7761172	999	1:7458	1:7189	2:14647	1.000
ALOXE3	17	7999217	8022234	3397	2:7458	2:7189	4:14647	1.000
PER1	17	8043787	8055753	4709	3:7458	0:7189	3:14647	0.049
TMEM107	17	8076296	8079714	1775	1:7458	1:7189	2:14647	1.000
MYH13	17	10204182	10276322	5992	2:7458	3:7189	5:14647	0.585
MYH8	17	10293641	10325267	6038	4:7458	0:7189	4:14647	0.023
TEKT3	17	15207128	15244958	1784	2:7458	4:7189	6:14647	1.000
TVP23C	17	15440293	15466945	4227	2:7458	2:7189	4:14647	1.000
TRIM16L	17	18625401	18639431	1982	2:7458	0:7189	2:14647	0.114
ALDH3A2	17	19552063	19580904	3823	0:7458	2:7189	2:14647	0.085
ALDH3A1	17	19641297	19648772	1928	0:7458	3:7189	3:14647	0.031
LGALS9	17	25958173	25976586	1764	1:7458	1:7189	2:14647	1.000
EFCAB5	17	28268622	28435470	5132	4:7458	1:7189	5:14647	0.201
ACBD4	17	43209966	43221543	2030	2:7458	1:7189	3:14647	0.774
HEXIM1	17	43224683	43229468	4785	1:7458	1:7189	2:14647	1.000
SPATA32	17	43331759	43339479	1284	1:7458	1:7189	2:14647	1.000
MAP3K14-AS	17	43339557	43345997	1260	1:7458	1:7189	2:14647	1.000
ARHGAP27	17	43471267	43503012	3650	2:7458	1:7189	3:14647	0.807
MAPT	17	43971747	44105699	6815	1:7458	2:7189	3:14647	0.662
EFCAB13	17	45401326	45518677	3950	1:7458	2:7189	3:14647	0.506
MSI2	17	55334373	55710544	2143	2:7458	0:7189	2:14647	0.117
BCAS3	17	58755171	59470199	3609	1:7458	1:7189	2:14647	0.947
TEX2	17	62224794	62340653	5090	0:7458	2:7189	2:14647	0.084
SLC16A5	17	73083821	73102255	2017	2:7458	0:7189	2:14647	0.118
CLUL1	18	596997	650293	1938	1:7458	4:7189	5:14647	1.000
TYMS	18	657603	673499	1583	1:7458	3:7189	4:14647	1.000
ENOSF1	18	670323	712662	5388	1:7458	6:7189	7:14647	0.042
MYOM1	18	3066804	3220106	5847	4:7458	1:7189	5:14647	0.198
TGIF1	18	3451590	3458406	2061	2:7458	2:7189	4:14647	0.972
RBBP8	18	20513294	20606449	3271	2:7458	0:7189	2:14647	0.113
DSC3	18	28570051	28622781	6978	1:7458	1:7189	2:14647	1.000
DSC2	18	28645941	28682388	5245	2:7458	1:7189	3:14647	0.616
DSC1	18	28709213	28742819	4256	1:7458	2:7189	3:14647	1.000
MEP1B	18	29769986	29800366	2311	1:7458	3:7189	4:14647	0.255
CCDC178	18	30517365	31020045	3355	4:7458	0:7189	4:14647	0.024

RIT2	18	40323191	40695657	1093	2:7458	0:7189	2:14647	0.140	
KATNAL2	18	44526786	44628614	2833	2:7458	1:7189	3:14647	0.649	
ME2	18	48405431	48476162	4622	3:7458	1:7189	4:14647	0.361	
PIGN	18	59711457	59854289	4875	1:7458	2:7189	3:14647	0.510	
KIAA1468	18	59854523	59974355	5454	0:7458	2:7189	2:14647	0.085	
CDH19	18	64168423	64271375	6241	5:7458	6:7189	11:14647	0.829	
POLRMT	19	617222	633568	3800	0:7458	5:7189	5:14647	0.006	
RNF126	19	647525	663233	1650	1:7458	1:7189	2:14647	1.000	
R3HDM4	19	896502	913225	1809	0:7458	2:7189	2:14647	0.081	
KISS1R	19	917341	921015	1625	0:7458	2:7189	2:14647	0.080	
GRIN3B	19	1000436	1009723	3254	1:7458	2:7189	3:14647	0.774	
CNN2	19	1026297	1039064	2458	1:7458	2:7189	3:14647	0.784	
ABCA7	19	1040101	1065570	6818	1:7458	7:7189	8:14647	0.020	
HMHA1	19	1067164	1086627	4273	0:7458	3:7189	3:14647	0.037	
POLR2E	19	1086577	1095391	2866	0:7458	3:7189	3:14647	0.036	
MIDN	19	1248551	1259142	3793	2:7458	2:7189	4:14647	1.000	
RPS15	19	1438362	1440492	531	0:7458	2:7189	2:14647	0.061	
APC2	19	1450147	1473243	10151	0:7458	2:7189	2:14647	0.089	
PCSK4	19	1481426	1490407	2661	2:7458	2:7189	4:14647	1.000	
ADAMTSL5	19	1505016	1513188	2857	1:7458	1:7189	2:14647	1.000	
BTBD2	19	1985446	2015702	2618	2:7458	2:7189	4:14647	1.000	
IZUMO4	19	2096867	2099583	1015	1:7458	4:7189	5:14647	1.000	
AMH	19	2249112	2252072	2018	4:7458	3:7189	7:14647	1.000	
TMPRSS9	19	2389783	2426086	3180	1:7458	2:7189	3:14647	1.000	
TIMM13	19	2425621	2427875	1956	1:7458	1:7189	2:14647	1.000	
ZNF554	19	2819871	2836733	3698	1:7458	1:7189	2:14647	1.000	
ZNF555	19	2841432	2860472	8547	3:7458	1:7189	4:14647	0.605	
TJP3	19	3708334	3750811	3123	0:7458	2:7189	2:14647	0.080	
MATK	19	3777966	3786415	2138	2:7458	1:7189	3:14647	0.703	
FBN3	19	8130286	8212385	8967	2:7458	4:7189	6:14647	0.365	
ATG4D	19	10654646	10664094	1896	1:7458	1:7189	2:14647	1.000	
SLC44A2	19	10736170	10755235	3434	2:7458	0:7189	2:14647	0.107	
QTRT1	19	10812111	10824043	1334	2:7458	0:7189	2:14647	0.119	
ZNF442	19	12460184	12476475	2825	2:7458	2:7189	4:14647	1.000	
CC2D1A	19	14016955	14041693	3640	1:7458	1:7189	2:14647	1.000	
PODNL1	19	14041999	14049289	3331	2:7458	0:7189	2:14647	0.108	
DCAF15	19	14063318	14072256	2258	1:7458	1:7189	2:14647	1.000	
RFX1	19	14072341	14117134	4377	2:7458	0:7189	2:14647	0.106	
RLN3	19	14139016	14141783	452	3:7458	1:7189	4:14647	0.706	
IFI30	19	18284578	18288927	1032	0:7458	2:7189	2:14647	0.073	
ZNF429	19	21688436	21721079	2361	0:7458	4:7189	4:14647	0.014	
ZNF100	19	21906842	21950430	4470	2:7458	3:7189	5:14647	1.000	
ZNF43	19	21987750	22034870	5771	1:7458	3:7189	4:14647	0.293	
ZNF681	19	23921996	23941693	6497	6:7458	2:7189	8:14647	0.179	
ZNF790-AS1	19	37288451	37319013	1054	2:7458	0:7189	2:14647	0.125	
ZNF790	19	37309223	37328929	2451	2:7458	0:7189	2:14647	0.111	
ZNF345	19	37341259	37370477	3126	3:7458	0:7189	3:14647	0.051	
CYP2B6	19	41497203	41524301	3052	3:7458	4:7189	7:14647	0.681	
CYP2A13	19	41594355	41602100	1760	3:7458	0:7189	3:14647	0.057	
CCDC61	19	46498718	46521874	1823	1:7458	1:7189	2:14647	1.000	
PGLYRP1	19	46522411	46526556	952	2:7458	1:7189	3:14647	0.795	
DHX34	19	47852537	47885961	4372	5:7458	0:7189	5:14647	0.011	
TULP2	19	49384221	49401996	1760	2:7458	0:7189	2:14647	0.120	
NUCB1	19	49403306	49426540	2579	1:7458	1:7189	2:14647	1.000	
CEACAM18	19	51979837	51986611	1197	3:7458	0:7189	3:14647	0.055	
SIGLEC12	19	51994612	52005043	2126	6:7458	0:7189	6:14647	0.007	
ZNF175	19	52074530	52092991	3785	3:7458	1:7189	4:14647	0.406	
ZNF578	19	52956828	53020131	6741	3:7458	1:7189	4:14647	0.348	
ZNF28	19	53300660	53324922	4712	2:7458	4:7189	6:14647	0.345	
ZNF468	19	53341784	53360902	4031	7:7458	7:7189	14:14647	0.914	
ZNF761	19	53935226	53961515	4212	4:7458	5:7189	9:14647	0.669	
ZNF813	19	53970988	53997546	4188	3:7458	3:7189	6:14647	1.000	
LILRB5	19	54754269	54761167	2267	1:7458	2:7189	3:14647	0.626	
LILRA3	19	54799854	54804265	1603	1:7458	1:7189	2:14647	1.000	
LILRA4	19	54844691	54850421	1708	0:7458	4:7189	4:14647	0.012	
LILRA2	19	55085256	55099027	1755	2:7458	1:7189	3:14647	1.000	
LILRB1	19	55128628	55149004	2968	0:7458	2:7189	2:14647	0.083	
NLRP2	19	55477710	55512510	3575	2:7458	1:7189	3:14647	0.705	
BRSK1	19	55795533	55823903	3081	1:7458	1:7189	2:14647	1.000	
TMEM150B	19	55824226	55836708	875	1:7458	1:7189	2:14647	1.000	
COX6B2	19	55861069	55866182	1663	1:7458	1:7189	2:14647	1.000	
ZNF628	19	55987698	55995854	3847	1:7458	1:7189	2:14647	1.000	
ZBTB45	19	59024896	59030921	2352	1:7458	1:7189	2:14647	1.000	
LOC1001316	19	59070552	59086164	2606	2:7458	1:7189	3:14647	0.779	
MZF1	19	59073283	59084942	2921	2:7458	1:7189	3:14647	0.766	
SLC23A2	20	4833001	4990939	6962	0:7458	3:7189	3:14647	0.035	
TMX4	20	7961715	8000393	2365	1:7458	1:7189	2:14647	1.000	
PLCB1	20	8113295	8865547	6823	1:7458	1:7189	2:14647	0.994	
TASP1	20	13370035	13619583	2351	1:7458	2:7189	3:14647	0.491	
OVOL2	20	18004795	18038521	1555	0:7458	2:7189	2:14647	0.075	
CST1	20	23728189	23731574	760	2:7458	0:7189	2:14647	0.147	
CST2	20	23804403	23807312	694	1:7458	1:7189	2:14647	1.000	
BPIFA3	20	31805134	31815559	1109	2:7458	1:7189	3:14647	0.783	
BPIFA1	20	31823801	31831115	1078	1:7458	1:7189	2:14647	1.000	
KRTAP19-6	21	31913853	31914181	328	2:7458	0:7189	2:14647	0.272	
TIAM1	21	32490735	32931290	7198	1:7458	1:7189	2:14647	1.000	
HLCS	21	38123188	38338956	6953	1:7458	1:7189	2:14647	1.000	
KRTAP10-10	21	46057272	46058372	1100	2:7458	0:7189	2:14647	0.127	
PRMT2	21	48055506	48075276	4438	3:7458	1:7189	4:14647	0.362	
XKR3	22	17264305	17302584	1681	1:7458	1:7189	2:14647	1.000	
DGCR6	22	18893735	18899601	1214	4:7458	1:7189	5:14647	1.000	22q11.2
DGCR14	22	19117791	19132190	5385	5:7458	0:7189	5:14647	0.010	22q11.2
GSC2	22	19136503	19137796	618	4:7458	1:7189	5:14647	1.000	22q11.2
SLC25A1	22	19163087	19166018	1662	4:7458	1:7189	5:14647	1.000	22q11.2
CLTCL1	22	19166986	19279239	5514	6:7458	1:7189	7:14647	1.000	22q11.2
GNB1L	22	19775933	19842462	1519	7:7458	2:7189	9:14647	1.000	22q11.2
C22orf29	22	19833660	19842371	6635	7:7458	1:7189	8:14647	0.287	22q11.2
COMT	22	19929262	19957498	2304	6:7458	1:7189	7:14647	1.000	22q11.2
TRMT2A	22	20099388	20104818	2986	7:7458	3:7189	10:14647	1.000	22q11.2
ZNF74	22	20748404	20762753	4040	5:7458	2:7189	7:14647	1.000	22q11.2
PI4KA	22	21061978	21213100	6752	7:7458	1:7189	8:14647	0.049	22q11.2
SNAP29	22	21213291	21245501	4261	5:7458	2:7189	7:14647	1.000	22q11.2
AIFM3	22	21321446	21335649	2429	5:7458	5:7189	10:14647	1.000	22q11.2
LZTR1	22	21336557	21353326	4315	10:7458	4:7189	14:14647	0.178	22q11.2
THAP7	22	21354060	21356404	1242	6:7458	1:7189	7:14647	0.299	22q11.2
P2RX6	22	21369463	21382302	2722	5:7458	2:7189	7:14647	1.000	22q11.2
SLC35E4	22	31031792	31043862	2542	2:7458	0:7189	2:14647	0.113	

OSBP2	22	31090792	31303811	4340	1:7458	2:7189	3:14647	0.657
SFI1	22	31892124	32014537	4391	4:7458	7:7189	11:14647	0.296
PISD	22	32014476	32026810	2679	0:7458	3:7189	3:14647	0.033
MPST	22	37415682	37425863	1535	1:7458	1:7189	2:14647	1.000
TMPRSS6	22	37461478	37499693	3191	2:7458	0:7189	2:14647	0.108
SSTR3	22	37602244	37608353	2123	2:7458	0:7189	2:14647	0.112
RAC2	22	37621309	37640305	1468	1:7458	1:7189	2:14647	1.000
CYTH4	22	37678494	37711389	3138	1:7458	1:7189	2:14647	1.000
APOBEC3A_E	22	39353526	39388783	1103	0:7458	2:7189	2:14647	0.073
APOBEC3D	22	39417117	39429256	2503	0:7458	2:7189	2:14647	0.081
EFCAB6	22	43924623	44208217	4929	2:7458	3:7189	5:14647	0.574
CELSR1	22	46756730	46933067	11389	1:7458	2:7189	3:14647	0.523
KANSL1	17_ctg5_hap1	596171	760700	5354	3:7458	1:7189	4:14647	0.378
NOTCH4	6_cox_hap2	3633297	3662520	6751	1:7458	1:7189	2:14647	1.000

Appendix B: Main C++ code for two-hit test

```
#include "genic.h"
#include <iostream>
#include <functional>
#include <cmath>
#include <cstdlib>
#include <cstddef>
#include <iomanip>
#include <list>
#include <algorithm>
#include "util.h"


extern GStore g;
extern Pseq::Util::Options args;
using namespace std;

double Pseq::Assoc::Aux_two_hit::prev = PLINKSeq::DEFAULT_PREV();
std::map< std::string, int > Pseq::Assoc::Aux_two_hit::func_inc;
std::map< std::string, int > Pseq::Assoc::Aux_two_hit::func_exc;

int Pseq::Assoc::Aux_two_hit::ncases = 0;
int Pseq::Assoc::Aux_two_hit::ncontrols = 0;
int Pseq::Assoc::Aux_two_hit::nmissing = 0;

bool Pseq::Assoc::Aux_two_hit::singles = false;
bool Pseq::Assoc::Aux_two_hit::mhit = false;

void Pseq::Assoc::Aux_two_hit::initialize(){

  prev = PLINKSeq::DEFAULT_PREV();
  singles = false;
  mhit = false;

  if ( args.has( "prev" ) )
    prev =  Helper::str2dbl(args.as_string( "prev" ));

  if ( args.has( "singles" ) )
    singles = true;

  if ( args.has( "mhit" ) )
    mhit = true;


  if( args.has( "func-inc" ) ){
    std::vector<std::string> inc = args.as_string_vector( "func-inc" );
    for( int i = 0; i < inc.size(); i++ )
      func_inc[inc[i]] = i;
  }

  if( args.has( "func-exc" ) ){
    std::vector<std::string> inc = args.as_string_vector( "func-exc" );
```

```
    for( int i = 0; i < inc.size(); i++ )
      func_exc[inc[i]] = i;
  }
}


double Pseq::Assoc::stat_two_hit( const VariantGroup & vars,
                          Aux_prelim * pre,
                          Aux_two_hit * aux,
                          std::map<std::string,
                          std::string> * output,
                          bool original )
{
  Out & pout = Out::stream( "twohit.vars" );

  if( aux->ncases == 0 ) {

    const int n = vars.n_individuals();

    for (int i = 0; i < n; i++)
      {
      if ( vars.ind(i)->affected() == CASE )
        aux->ncases++;
      if ( vars.ind(i)->affected() == CONTROL )
        aux->ncontrols++;
      if ( vars.ind(i)->missing() )
        aux->nmissing++;
      }
  }

  std::vector<double> hets;
  std::vector<std::string> ann;
  std::map<std::string, int> valid_chet;

  int mat[3][3];
  int ahom = 0; int uhom = 0;
  int achet = 0; int uchet = 0;
  int ahet = 0; int uhet = 0;

  int ahomi = 0; int uhomi = 0;
  int acheti = 0; int ucheti = 0;
  int aheti = 0; int uheti = 0;

  const int n = vars.n_individuals();

  for (int i = 0; i < n; i++)
    {
      // permuted individual index i->j
      int j = g.perm.pos(i);

      ahomi = uhomi = acheti = ucheti = aheti = uheti = 0;

      hets.clear();
      ann.clear();
```

```
    for( int j = 0; j < 3; j++ )
      for( int k = 0; k < 3; k++ )
        mat[j][k] = 0;

    for (int v = 0; v < vars.size(); v++)
    {

      const Genotype & gt = vars.geno(v,i);
      double d = gt.null() ? 0 : gt.minor_allele_count(pre->refmin.find(v)
               == pre->refmin.end());

      std::vector<int> pl;
      std::vector<int> ad;
      double ab  = -1;

      if ( vars.geno(v,i).meta.has_field( PLINKSeq::DEFAULT_AD() ) )
        {
          ad = vars.geno(v,i).meta.get_int( PLINKSeq::DEFAULT_AD() );
          ab = (ad[0] * 1.0) / (ad[0] + ad[1]);
        }
      else
        if( d != 0 )
          plog.warn( "NO ALLELIC DEPTH INFORMATION FOR VARIANT LEVEL QC");

      // count non-ref homozygotes
      std::vector< std::string > annot;

      if( vars(v).meta.has_field( PLINKSeq::DEFAULT_TRANS() )
          && vars(v).meta.has_field( PLINKSeq::DEFAULT_FUNC() )){

        std::vector<std::string> func =
          vars(v).meta.get_string( PLINKSeq::DEFAULT_FUNC() );
        std::vector<std::string> transcript =
          vars(v).meta.get_string( PLINKSeq::DEFAULT_TRANS() );

        //  fast string tokenizer:
        int n = 0;
        Helper::char_tok tok_trans( transcript[0] , &n , ',' );

        n = 0;
        Helper::char_tok tok_func( func[0] , &n , ',' );
        if( tok_trans.size() == 0 )
            annot.push_back(func[0]);
        else{
          for( int ti = 0; ti < tok_trans.size(); ti++)
          if( vars.name().compare(tok_trans[ti]) == 0 )
                annot.push_back(tok_func[ti]);
        }
      }

      std::string annot1 = "";
      int cnt = 0;

// only include defined annotations unless none given then include all
      bool pass = false;
```

```
if (aux->func_inc.size() == 0)
   pass = true;


 for( int ti = 0; ti < annot.size(); ti++){
   if( cnt > 0 )
     annot1 += "/";
   annot1 += annot[ti];
   cnt++;
   if( aux->func_inc.count(annot[ti]) > 0
     || aux->func_inc.size() == 0 )
     pass = true;

   if( aux->func_exc.count(annot[ti]) > 0 ){
     pass = false;
     break;
   }
 }

 if( annot1.length() == 0 )
   annot1 = ".";
 if( d == 2 && (ab < PLINKSeq::DEFAULT_AB_HOMMAX()
     || ab == -1) && pass )
   {
     if( vars.ind( j )->affected() == CASE){
     if(original)
       pout << "HOM\t" \
            << vars.name() << "\t" \
            << g.locdb.alias( vars.name() , false ) << "\t" \
            << vars(v) << "\t" \
            << vars.ind( j )->id() \
            << "\tcase\t" \
            << annot1 << "\n";
     ahomi++;
     }
     if ( vars.ind(j)->affected() == CONTROL ){
     if(original)
       pout << "HOM\t" \
            << vars.name() << "\t" \
                << g.locdb.alias( vars.name() , false ) << "\t" \
            << vars(v) << "\t" \
            << vars.ind( j )->id() \
            << "\tcontrol\t" \
            << annot1 << "\n";
     uhomi++;
     }
   }
 // found het, store for later
 if( d == 1 && ((ab > PLINKSeq::DEFAULT_AB_HETMIN()
     && ab < PLINKSeq::DEFAULT_AB_HETMAX()) || ab == -1) && pass){
   hets.push_back(v);
   ann.push_back(annot1);
   if ( vars.ind(j)->affected() == CASE )
     aheti++;
   if ( vars.ind(j)->affected() == CONTROL )
```

```
          uheti++;
        }
      }

      if ( hets.size() > 0 ) {
        for (int z = 0; z < hets.size()-1; z++){
          for (int k = z+1; k < hets.size(); k++){

            std:string pair = Helper::int2str(vars(hets[z]).position()) + "," +
            Helper::int2str(vars(hets[k]).position());

          // clear matrix

          //perform population check only when pair hasn't been seen before
          //      bool chet = false;
          if (  valid_chet.count(pair) == 0 ){
            //clear matrix
            for (int u = 0; u < 3; u++ )
            for (int v = 0; v < 3; v++ )
              mat[u][v] = 0;


            // fill in genotype matrix to test two hits
            for (int l = 0; l < n; l++){
            int var1 = (int) vars(hets[z], l).null() ? 0 : vars(hets[z],
l).minor_allele_count(pre->refmin.find(hets[z]) == pre->refmin.end());
            int var2 = (int) vars(hets[k], l).null() ? 0 : vars(hets[k],
l).minor_allele_count(pre->refmin.find(hets[k]) == pre->refmin.end());
            mat[var1][var2]++;
            }

            // test two hit
            int rest = mat[0][1] + mat[0][2] + mat[1][0] + mat[1][2] + mat[2][0] +
            mat[2][1] + mat[2][2];

            if( ((mat[0][1] + mat[0][2]) > 0 && (mat[1][0] + mat[2][0]) > 0 && ( mat[1][2]
            + mat[2][1] + mat[2][2] == 0 ) ) || (mat[1][1] == 1 && rest == 0 && aux-
            >singles) || aux->mhit )
                  valid_chet[pair] = 1;
            else
                  valid_chet[pair] = 0;

            //std::cout << mat[0][0] << " " << mat[0][1] << " " << mat[0][2] << std::endl;
            //std::cout << mat[1][0] << " " << mat[1][1] << " " << mat[1][2] << std::endl;
            //std::cout << mat[2][0] << " " << mat[2][1] << " " << mat[2][2] << std::endl;
          }

          if ( valid_chet[pair] == 1 ){ // chet || ( valid_chet.count(pair) > 0 &&
          valid_chet[pair] == 1 ) ){

            //sort two annotations for consistancy
            std::vector<std::string> vtmp;
            if( ann[z].length() > 0 )
                  vtmp.push_back(ann[z]);
            else
```

```
            vtmp.push_back(".");
        if( ann[k].length() > 0 )
            vtmp.push_back(ann[k]);
        else
            vtmp.push_back(".");

        sort(vtmp.begin(), vtmp.end());


        if ( vars.ind(j)->affected() == CASE )
        {
          if ( original )
            pout << "CHET\t"         \
               << vars.name() << "\t" \
               << g.locdb.alias( vars.name() , false ) << "\t" \
               << vars(hets[z]) << "," << vars(hets[k]) << "\t" \
               << vars.ind( j )->id() \
               << "\tcase\t" \
               << vtmp[0] << "," << vtmp[1] << "\n";
          acheti++;
        }
        if ( vars.ind(j)->affected() == CONTROL )
        {
          if( original )
            pout << "CHET\t" \
                 << vars.name() << "\t" \
                      << g.locdb.alias( vars.name() , false ) << "\t" \
               << vars(hets[z]) << "," << vars(hets[k]) << "\t" \
               << vars.ind( j )->id() \
               << "\tcontrol\t" \
               << vtmp[0] << "," << vtmp[1] << "\n";
          ucheti++;
        }
        }
      }
    }
   }

   // count total number of cases and controls
   // if case
   if( ahomi > 0 || acheti > 0 )
   ahom++;
   if( aheti > 0 )
   ahet++;

   //control
   if( uhomi > 0 || ucheti > 0 )
     uhom++;
   if( uheti >0 )
   uhet++;
  }


// skip chrX and Y
```

```
//  haploid ??? chrX ??

double total_hets = ahet + uhet;
double a = 2 * aux->ncontrols;
double arec = ahom;
double urec = uhom;
double f = (( 2 * urec ) + uhet) / a;

//double f = (uhet + (2*urec))/(2*aux->ncontrols);

double pAA = f * f;
double pAA_1 = urec / aux->ncontrols;
if ( pAA < pAA_1) {
  pAA = pAA_1;
}
if ( pAA == 0 ) {
  pAA = 0.0001;
}

double l0 = (arec * log(pAA)) + (aux->ncases - arec) * log( 1.0 - pAA );
l0 += (urec * log(pAA)) + (aux->ncontrols - urec) * log( 1.0 - pAA );
double l1 = l0;

double denom1 = 0;
double this_l1 = 0;
double x = 0;
double pAA_nodis = 0;
double grrmax = 1.1;


for (double gr = 1.1; gr <= 50; gr += 0.1) {

  denom1 = gr * pAA + ( 1 - pAA );

  this_l1 = arec * log(gr*pAA/denom1) + (aux->ncases-arec)*log(1.0 - (gr*pAA/denom1));
  x = aux->prev/denom1;
  pAA_nodis = (1.0 - gr*x)*pAA / ( 1.0 - aux->prev );
  this_l1 += urec*log(pAA_nodis) + (aux->ncontrols-urec)*log(1.0 - (pAA_nodis));

  if (this_l1 > l1) {
    l1 = this_l1;
    grrmax = gr;
  }
}

double diff = (l1-l0) / log(10);
double diff2 = l1-l0;

double chisqVal = 2 * diff2;

double pvalue = Statistics::chi2_prob(chisqVal, 1);
aux->returned_pvalue = pvalue < 0 ? 1.00 : pvalue ;


// set the statistic on the Aux_two_hit struct.
```

```
  // gseq will read this value each permutation (this also includes adaptive permutation)

  if (original)
    {
      (*output)["TWO-HIT"] = "P=" + Helper::dbl2str( pvalue )
      + ";AF=" + Helper::dbl2str(f)
      + ";CASES=" + Helper::int2str(arec) + "," + Helper::int2str(ahet) + "," +
Helper::int2str(aux->ncases)
      + ";CONTROLS=" + Helper::int2str(urec) + "," + Helper::int2str(uhet) + "," +
Helper::int2str(aux->ncontrols);
    }
  return chisqVal;
}
```

References

1. Sturtevant AH. A Third Group of Linked Genes in Drosophila Ampelophila. Science. 1913;37(965):990-2.
2. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American journal of human genetics. 1980;32(3):314-31.
3. Macdonald ME, Ambrose CM, Duyao MP, Myers RH, Lin C, Srinidhi L, et al. A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntingtons-Disease Chromosomes. Cell. 1993;72(6):971-83.
4. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.
5. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291(5507):1304-51.
6. Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, et al. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature. 2000;407(6803):513-6.
7. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 2001;409(6822):928-33.
8. A haplotype map of the human genome. Nature. 2005;437(7063):1299-320.
9. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73.
10. Pericak-Vance MA, Bebout JL, Gaskell PC, Jr., Yamaoka LH, Hung WY, Alberts MJ, et al. Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. American journal of human genetics. 1991;48(6):1034-50.
11. Jeunemaitre X, Soubrier F, Kotelevtsev YV, Lifton RP, Williams CS, Charru A, et al. Molecular basis of human hypertension: role of angiotensinogen. Cell. 1992;71(1):169-80.
12. Edwards AO, Ritter R, 3rd, Abel KJ, Manning A, Panhuysen C, Farrer LA. Complement factor H polymorphism and age-related macular degeneration. Science. 2005;308(5720):421-4.
13. Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS. Replication of linkage studies of complex traits: an examination of variation in location estimates. American journal of human genetics. 1999;65(3):876-84.
14. Cordell HJ. Sample size requirements to control for stochastic variation in magnitude and location of allele-sharing linkage statistics in affected sibling pairs. Annals of human genetics. 2001;65(Pt 5):491-502.
15. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996;273(5281):1516-7.
16. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. 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. 2009;106(23):9362-7.
17. van Os J, Kapur S. Schizophrenia. Lancet. 2009;374(9690):635-45.
18. Lichtenstein P, Bjork C, Hultman CM, Scolnick E, Sklar P, Sullivan PF. Recurrence risks for schizophrenia in a Swedish national cohort. Psychological medicine. 2006;36(10):1417-25.
19. McGrath J, Saha S, Chant D, Welham J. Schizophrenia: a concise overview of incidence, prevalence, and mortality. Epidemiologic reviews. 2008;30:67-76.
20. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Archives of general psychiatry. 2003;60(12):1187-92.
21. Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet. 2009;373(9659):234-9.
22. Krabbendam L, van Os J. Schizophrenia and urbanicity: a major environmental influence--conditional on genetic risk. Schizophrenia bulletin. 2005;31(4):795-9.
23. Moore TH, Zammit S, Lingford-Hughes A, Barnes TR, Jones PB, Burke M, et al. Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. Lancet. 2007;370(9584):319-28.
24. Bassett AS, Chow EW, AbdelMalik P, Gheorghiu M, Husted J, Weksberg R. The schizophrenia phenotype in 22q11 deletion syndrome. The American journal of psychiatry. 2003;160(9):1580-6.
25. Blackwood DH, Fordyce A, Walker MT, St Clair DM, Porteous DJ, Muir WJ. Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. American journal of human genetics. 2001;69(2):428-33.
26. Pickard BS, Malloy MP, Christoforou A, Thomson PA, Evans KL, Morris SW, et al. Cytogenetic and genetic evidence supports a role for the kainate-type glutamate receptor gene, GRIK4, in schizophrenia and bipolar disorder. Molecular psychiatry. 2006;11(9):847-57.
27. Thomson PA, Parla JS, McRae AF, Kramer M, Ramakrishnan K, Yao J, et al. 708 Common and 2010 rare DISC1 locus variants identified in 1542 subjects: analysis for association with psychiatric disorder and cognitive traits. Molecular psychiatry. 2013.
28. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science. 2008;320(5875):539-43.
29. Kirov G, Gumus D, Chen W, Norton N, Georgieva L, Sari M, et al. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. Hum Mol Genet. 2008;17(3):458-65.
30. Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature. 2010;464(7289):713-20.
31. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number

variation detection in whole-genome SNP genotyping data. Genome Res. 2007;17(11):1665-74.
32. Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet. 2008;40(10):1253-60.
33. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature. 2008;455(7210):237-41.
34. Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, et al. Large recurrent microdeletions associated with schizophrenia. Nature. 2008;455(7210):232-6.
35. Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol Psychiatry. 2012;17(2):142-53.
36. Bayes A, van de Lagemaat LN, Collins MO, Croning MD, Whittle IR, Choudhary JS, et al. Characterization of the proteome, diseases and evolution of the human postsynaptic density. Nature neuroscience. 2011;14(1):19-21.
37. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nature reviews Genetics. 2012;13(8):537-51.
38. Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Molecular psychiatry. 2012;17(2):142-53.
39. Holmans PA, Riley B, Pulver AE, Owen MJ, Wildenauer DB, Gejman PV, et al. Genomewide linkage scan of schizophrenia in a large multicenter pedigree sample using single nucleotide polymorphisms. Molecular psychiatry. 2009;14(8):786-95.
40. Huang Q, Shete S, Amos CI. Ignoring linkage disequilibrium among tightly linked markers induces false-positive evidence of linkage for affected sib pair analysis. American journal of human genetics. 2004;75(6):1106-12.
41. John S, Shephard N, Liu G, Zeggini E, Cao M, Chen W, et al. Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites. American journal of human genetics. 2004;75(1):54-64.
42. Levinson DF, Evgrafov OV, Knowles JA, Potash JB, Weissman MM, Scheftner WA, et al. Genetics of recurrent early-onset major depression (GenRED): significant linkage on chromosome 15q25-q26 after fine mapping with single nucleotide polymorphism markers. The American journal of psychiatry. 2007;164(2):259-64.
43. Li D, Collier DA, He L. Meta-analysis shows strong positive association of the neuregulin 1 (NRG1) gene with schizophrenia. Hum Mol Genet. 2006;15(12):1995-2002.
44. Straub RE, Jiang Y, MacLean CJ, Ma Y, Webb BT, Myakishev MV, et al. Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. Am J Hum Genet. 2002;71(2):337-48.
45. Ng MY, Levinson DF, Faraone SV, Suarez BK, DeLisi LE, Arinami T, et al. Meta-analysis of 32 genome-wide linkage studies of schizophrenia. Molecular psychiatry. 2009;14(8):774-85.

46. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature genetics. 1995;11(3):241-7.
47. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics. 2006;38(8):904-9.
48. O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nature genetics. 2008;40(9):1053-5.
49. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, et al. Common variants conferring risk of schizophrenia. Nature. 2009;460(7256):744-7.
50. Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature. 2009;460(7256):753-7.
51. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460(7256):748-52.
52. Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, et al. Genome-wide association study identifies five new schizophrenia loci. Nature genetics. 2011;43(10):969-76.
53. McClellan J, King MC. Genetic heterogeneity in human disease. Cell. 2010;141(2):210-7.
54. Ruderfer DM, Korn J, Purcell SM. Family-based genetic risk prediction of multifactorial disease. Genome medicine. 2010;2(1):2.
55. Wray NR, Purcell SM, Visscher PM. Synthetic associations created by rare variants do not explain most GWAS results. PLoS biology. 2011;9(1):e1000579.
56. Anderson CA, Soranzo N, Zeggini E, Barrett JC. Synthetic associations are unlikely to account for many common disease genome-wide association signals. PLoS biology. 2011;9(1):e1000580.
57. Goldstein DB. The importance of synthetic associations will only be resolved empirically. PLoS biology. 2011;9(1):e1001008.
58. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nature genetics. 2010;42(7):565-9.
59. Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, Voight BF, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nature genetics. 2012;44(5):483-9.
60. Lee SH, DeCandia TR, Ripke S, Yang J, Sullivan PF, Goddard ME, et al. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nature genetics. 2012;44(3):247-50.
61. Visscher PM, Goddard ME, Derks EM, Wray NR. Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. Molecular psychiatry. 2012;17(5):474-85.
62. Jouan L, Girard SL, Dobrzeniecka S, Ambalavanan A, Krebs MO, Joober R, et al. Investigation of rare variants in LRP1, KPNA1, ALS2CL and ZNF480 genes in

schizophrenia patients reflects genetic heterogeneity of the disease. Behavioral and brain functions : BBF. 2013;9:9.
63. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. 2012;485(7397):242-5.
64. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012;485(7397):237-41.
65. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012;485(7397):246-50.
66. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. Neuron. 2012;74(2):285-99.
67. Holmans P. Statistical methods for pathway analysis of genome-wide data for association with complex genetic traits. Advances in genetics. 2010;72:141-79.
68. Lee PH, O'Dushlaine C, Thomas B, Purcell SM. INRICH: interval-based enrichment analysis for genome-wide association studies. Bioinformatics. 2012;28(13):1797-9.
69. O'Dushlaine C, Kenny E, Heron EA, Segurado R, Gill M, Morris DW, et al. The SNP ratio test: pathway analysis of genome-wide association datasets. Bioinformatics. 2009;25(20):2762-3.
70. Holmans P, Green EK, Pahwa JS, Ferreira MA, Purcell SM, Sklar P, et al. Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. American journal of human genetics. 2009;85(1):13-24.
71. O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D, et al. Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. Molecular psychiatry. 2011;16(3):286-92.
72. Lee YH, Kim JH, Song GG. Pathway analysis of a genome-wide association study in schizophrenia. Gene. 2013;525(1):107-15.
73. Jia P, Wang L, Meltzer HY, Zhao Z. Common variants conferring risk of schizophrenia: a pathway analysis of GWAS data. Schizophrenia research. 2010;122(1-3):38-42.
74. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. PLoS medicine. 2005;2(5):e141.
75. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of de novo copy number mutations with autism. Science. 2007;316(5823):445-9.
76. Kirov G. The role of copy number variation in schizophrenia. Expert review of neurotherapeutics. 2010;10(1):25-32.
77. Kirov G, Grozeva D, Norton N, Ivanov D, Mantripragada KK, Holmans P, et al. Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. Hum Mol Genet. 2009;18(8):1497-503.
78. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, et al. Microduplications of 16p11.2 are associated with schizophrenia. Nature genetics. 2009;41(11):1223-7.

79. Sebat J, Levy DL, McCarthy SE. Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. Trends in genetics : TIG. 2009;25(12):528-35.
80. Murphy KC. Schizophrenia and velo-cardio-facial syndrome. Lancet. 2002;359(9304):426-30.
81. Sullivan PF, Lin D, Tzeng JY, van den Oord E, Perkins D, Stroup TS, et al. Genomewide association for schizophrenia in the CATIE study: results of stage 1. Mol Psychiatry. 2008;13(6):570-84.
82. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007;81(3):559-75.
83. Raychaudhuri S, Korn JM, McCarroll SA, Altshuler D, Sklar P, Purcell S, et al. Accurately assessing the risk of schizophrenia conferred by rare copy-number variation affecting genes with brain function. PLoS genetics. 2010;6(9).
84. Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, et al. The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic Acids Res. 2005;33(Database issue):D284-8.
85. Zhang D, Cheng L, Qian Y, Alliey-Rodriguez N, Kelsoe JR, Greenwood T, et al. Singleton deletions throughout the genome increase risk of bipolar disorder. Molecular psychiatry. 2009;14(4):376-80.
86. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2008;36(Database issue):D13-21.
87. Ruderfer DM, Kirov G, Chambert K, Moran JL, Owen MJ, O'Donovan MC, et al. A family-based study of common polygenic variation and risk of schizophrenia. Molecular psychiatry. 2011;16(9):887-8.
88. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet. 2009;84(2):210-23.
89. The International HapMap Project. Nature. 2003;426(6968):789-96.
90. Tikhonov AN. Numerical methods for the solution of ill-posed problems. Dordrecht ; Boston: Kluwer Academic Publishers; 1995. ix, 253 p. p.
91. Keller MC, Simonson MA, Ripke S, Neale BM, Gejman PV, Howrigan DP, et al. Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. PLoS genetics. 2012;8(4):e1002656.
92. Need AC, McEvoy JP, Gennarelli M, Heinzen EL, Ge D, Maia JM, et al. Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. American journal of human genetics. 2012;91(2):303-12.
93. Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S, et al. Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nature genetics. 2011;43(9):864-8.
94. Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y, et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nature genetics. 2012.

95. Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, MacArthur DG, et al. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron. 2013;77(2):235-42.
96. Aleman A, Kahn RS, Selten JP. Sex differences in the risk of schizophrenia: evidence from meta-analysis. Archives of general psychiatry. 2003;60(6):565-71.
97. Kuperberg G, Heckers S. Schizophrenia and cognitive function. Current opinion in neurobiology. 2000;10(2):205-10.
98. Dalman C, Broms J, Cullberg J, Allebeck P. Young cases of schizophrenia identified in a national inpatient register--are the diagnoses valid? Social psychiatry and psychiatric epidemiology. 2002;37(11):527-31.
99. Ekholm B, Ekholm A, Adolfsson R, Vares M, Osby U, Sedvall GC, et al. Evaluation of diagnostic procedures in Swedish patients with schizophrenia and related psychoses. Nordic journal of psychiatry. 2005;59(6):457-64.
100. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research. 2010;20(9):1297-303.
101. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics. 2011;43(5):491-8.
102. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.
103. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. American journal of human genetics. 2012;91(2):275-92.
104. Daw SC, Taylor C, Kraman M, Call K, Mao J, Schuffenhauer S, et al. A common region of 10p deleted in DiGeorge and velocardiofacial syndromes. Nature genetics. 1996;13(4):458-60.
105. Jablensky A. Epidemiology of schizophrenia: the global burden of disease and disability. European archives of psychiatry and clinical neuroscience. 2000;250(6):274-85.
106. Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell. 1991;65(5):905-14.
107. Hagerman R, Au J, Hagerman P. FMR1 premutation and full mutation molecular mechanisms related to autism. Journal of neurodevelopmental disorders. 2011;3(3):211-24.
108. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nature genetics. 2004;36(9):949-51.
109. Antunez C, Boada M, Gonzalez-Perez A, Gayan J, Ramirez-Lorca R, Marin J, et al. The membrane-spanning 4-domains, subfamily A (MS4A) gene cluster contains a common variant associated with Alzheimer's disease. Genome medicine. 2011;3(5):33.
110. Bassett AS, Chow EW. Schizophrenia and 22q11.2 deletion syndrome. Current psychiatry reports. 2008;10(2):148-57.

111. Wang H, Duan S, Du J, Li X, Xu Y, Zhang Z, et al. Transmission disequilibrium test provides evidence of association between promoter polymorphisms in 22q11 gene DGCR14 and schizophrenia. J Neural Transm. 2006;113(10):1551-61.