Title
Detecting genetic similarity between complex human traits by exploring their common molecular mechanism

Author
Gu, Jialiang

Publication Date
2019

Peer reviewed|Thesis/dissertation

Detecting genetic similarity between complex human traits by exploring their common molecular mechanism

by

Jialiang Gu

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

AND

UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

DocuSigned by: Hao Li
C2CA172BE8684DD...
Hao Li
Chair

DocuSigned by: Ryan Hernandez
B7001B6E0064D8...
Ryan Hernandez

DocuSigned by: Irina Conboy
9D3480CD6C6B43C...
Irina Conboy

Committee Members

Acknowledgement

This project would not have been possible without Prof. Dr. Hao Li, Dr. Jiashun Zheng and Dr. Chris Fuller at the University of California, San Francisco (UCSF) and Caribou Bioscience. The Li lab grew into a multi-facet research group consist of both experimentalists and computational biologists covering three research areas including cellular/molecular mechanism of ageing, genetic determinants of complex human traits and structure, function, evolution of gene regulatory network. Labs like these are the pillar of global success and reputation of UCSF and UC Berkeley, and excellent incubator of research projects similar to the one documented in this manuscript.

Secondly I would like to thank Dr. Ryan Hernandez, Dr. Irina Conboy for serving as my dissertation committee. I truly appreciate their advice and questions which helped me refine research scope and adjust research direction.

Thirdly I would like to thank Dr. Patricia Babbitt and Dr. Adam Abate to be my academic advisors, providing advice ranging from materializable action such as rotation lab selection to philosophical discussion such as the history, evolution and future of life science and its comparison with physical science.

Above all, the success of this project can be attributed to my family who supported my decision in 2014 to resign from the software industry and pursue my true interest. I am extremely grateful for their support bearing the risk with me as their only child.

Finally, this dissertation is part of a story that One's Revenge is Others' Redemption.

Detecting genetic similarity between complex human traits by exploring their common molecular mechanism

Jialiang Gu

Abstract

The rapid accumulation of Genome Wide Association Studies (GWAS) and association studies of intermediate molecular traits provides new opportunities for comparative analysis of the genetic basis of complex human phenotypes. Using a newly developed statistical framework called Sherlock-II that integrates GWAS with eQTL (expression Quantitative Trait Loci) and metabolite-QTL data, we systematically analyzed 445 GWAS datasets, and identified 2114 significant gene-phenotype associations and 469 metabolites-phenotype associations (passing a Q-value cutoff of 1/3). This integrative analysis allows us to translate SNP-phenotype associations into functionally informative gene-phenotype association profiles. Genetic similarity analyses based on these profiles clustered phenotypes into sub-trees that reveal both expected and unexpected relationships. We employed a statistical approach to delineate sets of functionally related genes that contribute to the similarity between their association profiles. This approach suggested common molecular mechanisms that connect the phenotypes in a subtree. For example, we found that fasting insulin, fasting glucose, breast cancer, prostate cancer, and lung cancer clustered into a subtree, and identified cyclic AMP/GMP signaling that connects breast cancer and insulin, NAPDH oxidase/ROS generation that connects the three cancers, and apoptosis that connects all five phenotypes. Our approach can be used to assess genetic similarity and suggest mechanistic connections between phenotypes. It has the potential to improve the diagnosis and treatment of a disease by mapping mechanistic insights from one phenotype onto others based on common molecular underpinnings.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. Importance of disease classification and the need for similarity study

Disease classification is an indispensable component in the health-related system because it influences the curriculum of medical education, biomedical research publication, medical care delivery and reimburse rate. Current disease nomenclature is under periodical updates because it must provide consistent terminology to permit clear communication about diseases that are defined by agreed upon criteria and reflects the scientific advance in our understanding of molecular pathways and environmental factors that contribute to disease origin and pathology[1].

Similarity study is at the core of disease classification because it sets the boundary between disease categories. Traditionally, diseases are classified based on symptom similarity, followed by histology similarity. The recently available genomic data provides a new similarity criterion for disease classification.

1.2. Current state of disease classification and the need for a new similarity metric

Contemporary classification of human disease dates to the late 19th century, and derives from observational correlation between pathological analysis and clinical syndromes. Through the last century, this symptom-based classification has become more objective with the advancement of histology and definitive laboratory tests became an essential part of the overall diagnostic paradigm[2]. Today, International Statistical Classification of Diseases and Related Health Problems(ICD) version 10 has become a wide accepted taxonomy of diseases. Although this

taxonomy has served well for the delivery of medical care, there exist two types of mis-classifications where 1) single phenotype with various symptoms are artificially separated as distinct diseases 2) multiple distinct diseases are classified under an umbrella disease class.

Example for the first type of error include Emery-Dreyfus muscular dystrophy, Charcot-Marie-Tooth axonal neuropathy, lipodystrophy, and premature aging disorders. These disorders can all be traced to mutations in the LMNA gene[1].

Example for the second type of error include hypertrophic cardiomyopathy which can be caused by several different sarcomeric proteins mutations including myosin heavy chain, myosin light chain, tropomyosin, and troponin C [3]. Clearly, searching for one therapy for such a bag of different phenotypes under the same name can hardly lead to success due to the distinct behavior of constituting components.

Given these misclassifications based on symptom and histology, a new similarity metric is needed to capture the real characteristic of human diseases and thus redefine the current disease classification system.

1.3. Example of new similarity metric in Precision Medicine Initiative

Precision Medicine Initiative is a national effort launched in 2015 in the search for prevention and treatment strategies that take individual variability into account. The main driving force for this new wave of precision medicine is large-scale biologic databases (such as the human genome sequence), powerful methods for characterizing patients (such as proteomics, metabolomics, genomics, diverse cellular assays, and even mobile health technology), and computational tools for analyzing large sets of data[4]. Clearly, genomic data has been chosen as

the new evidence in this new trend. Target diseases includes diabetes and cancer where the new metric has proved its efficacy.

For example, multiple lung cancers can be differentiated by their genomic characterization which is traditionally classified by tumor's histological appearance as small-cell lung cancer and non-small-cell lung cancer. Molecular understanding of lung cancer gave birth to two drugs Gefitinib and Erlotinib, which inhibited the function of certain receptor tyrosine kinases, including epidermal growth factor receptor (EGFR). These two drugs exert significant anti-tumor effect in early trials, but only in 10 percent of patients. These 10 percent of patients are found to carry mutations that activates EGFR in their cancer[5, 6].

In this example, lung cancer is classified by genomic mutation instead of histological appearance and lead to improved prediction of drug response. More cancers have been classified by genomic information including transcription profiling and proteomic analysis[7, 8].

Similar to the cancer example, a new wave of association studies have been conducted in the search of genomic characterization of various diseases. Such Genome Wide Association Study provides new evidence for similarity study in the nucleotide resolution.

1.4. Our approach

Genome Wide Association Studies (GWAS) seek to identify the genetic basis of complex phenotypes by associating genetic variants (typically characterized by Single Nucleotide Polymorphisms, or SNPs) to particular traits across different individuals. These studies have been applied to a wide range of phenotypes, including many complex diseases, and have led to the identification of a large number of informative SNPs [9, 10]. These disease associated SNPs brought significant insights into disease mechanisms and informed strategies for improved

intervention and prevention. GWASs have also been successfully implemented to define the relative influence of genotype and environment on phenotype prevalence, assisting in risk prediction[9].

Despite these successes, traditional GWAS analyses based on the association of individual SNPs have only limited power. Many potentially informative SNPs fall below the genome-wide significance threshold, which is set high due to the large number of loci typically tested. Also, for associations that do exceed the genome-wide significance threshold, the functional link between an individual SNP and the phenotype is often poorly understood. To circumvent these problems, integrative analysis methods have emerged that combine GWAS with other quantitative molecular data. These methods attempt to bridge the knowledge gap between GWAS SNPs and phenotypes by uncovering the genes and other intermediate molecular traits (IMTs) through which genetic polymorphisms generate phenotypic differences [11-15]. These approaches utilize the collective power of SNPs and generate mechanistic hypotheses that can be tested and can guide future therapeutic interventions.

As GWAS have been applied to additional human traits, it has become increasingly apparent that seemingly unrelated complex traits may share common genetic origins[16-19]. Phenotypic connections (often in the form of co-occurrence) have long been observed based on epidemiological studies [20]. Recent comparative GWAS analyses have lent genetic support to these earlier observations by searching for pleiotropic SNPs implicated across multiple phenotypes[21-23] . However, genetic similarity between different phenotypes can be obscured at the SNP level. Phenotypes sharing a set of IMTs may not necessarily associate with the same SNPs, since multiple SNPs can converge on the same IMT, and only a small subset of these SNPs will be identified in a particular GWAS. Furthermore, it is quite difficult to infer common

mechanisms based on the shared SNPs, as they often fall into non-coding regions with no obvious functional implication.

We developed an integrative analysis framework in which GWAS-IMT analysis is combined with comparative analysis of multiple phenotypes (Figure 1.1). Such analysis not only enables us to translate SNP-phenotype association to IMT-phenotype association (which yields improved statistical power and mechanistic insight), but also to generate hypotheses regarding common physiological processes underlying different (and sometimes seemingly unrelated) phenotypes based on the similarity of their IMT association profiles.

The first step in this integrative analysis is to associate the IMTs with phenotypes using all the SNPs present in GWAS results. We developed a new computational algorithm called Sherlock-II to integrate GWAS data with eQTL and metabolite-QTL data, translating SNP-phenotype associations into gene/metabolite-phenotype associations. This produces a gene-phenotype association matrix and a metabolite-phenotype association matrix and defines a gene association profile for each phenotype (one column of the gene-phenotype association matrix).

We then analyze the relation between phenotypes and infer possible common underlying physiological processes (PP) by two different approaches. In the first approach, we group the phenotypes based on the similarity between their gene association profiles and then identify groups of functionally related genes (such as those defined by gene ontology (GO) terms or pathways) that contribute to the similarity. This generate hypotheses on PPs (green boxes in Figure 1.1) that connect multiple phenotypes (orange boxes in Figure 1.1).

In the second approach, we apply two-way bi-clustering to the gene-phenotype association matrix to identified sub-clusters in which a group of genes are associated with a subset of

phenotypes. We hypothesize that such a group of genes participate in the same PPs that influence the subset of phenotypes. We conducted literature searches on each group of genes to identify similar mechanisms through which they may affect the phenotypes, such as common pathways. This often leads to important clues into the PPs involved.

Our analysis predicts new connections between PPs and phenotypes (Figure 1.1) supported by multiple genes and thus generates etiological hypotheses with genetic support. Identifying different diseases associated with common PPs also provide an opportunity to transfer existing intervention from one disease to another.

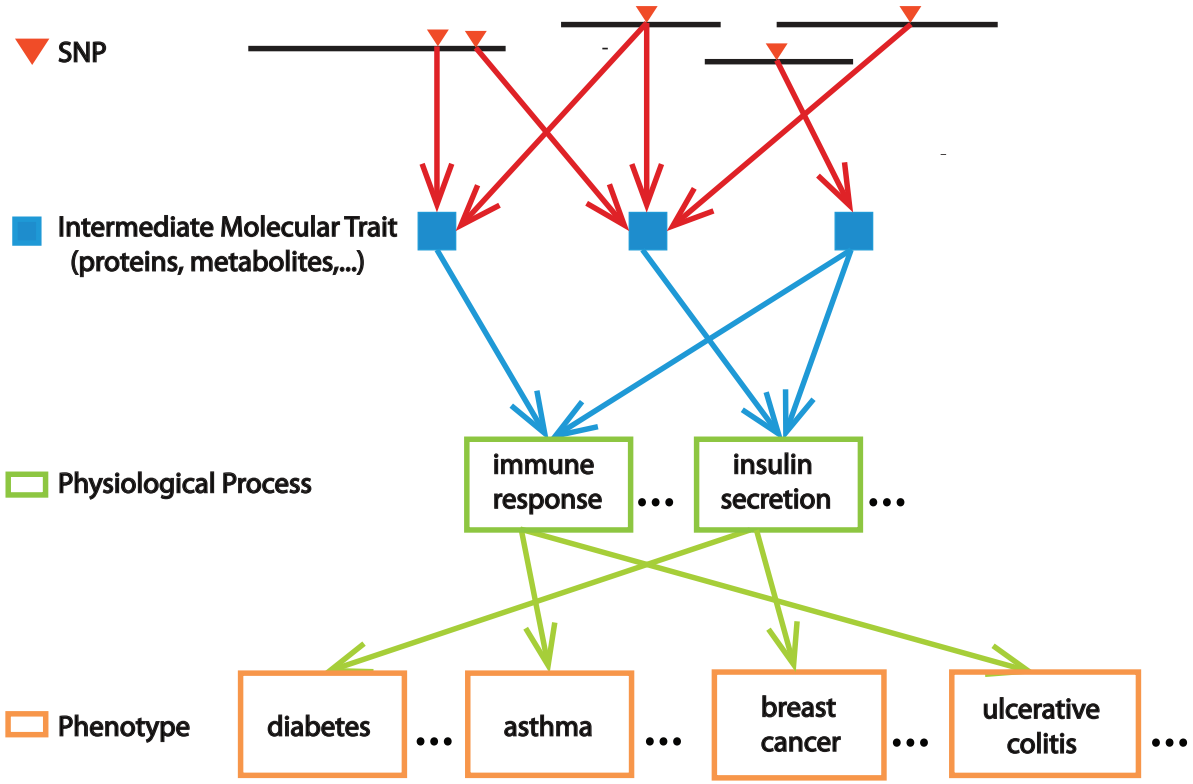


Figure 1. 1 A schematic of the integrative analysis framework.

The underlying model is that SNPs (red triangles) influence phenotypes (orange boxes) by influencing intermediate molecular traits (IMTs, blue squares) which in turn contribute to various physiological processes (green boxes), and each phenotype is controlled by a combination of physiological processes. In the first step of the analysis, data from GWASs and association studies of IMTs (such as eQTL and metabolite-QTL) are integrated to infer IMT-phenotype associations. A comparative analysis of the IMT association profiles for different phenotypes then identifies genetic similarity between phenotypes, supported by shared IMTs. These IMTs may point to a set of common physiological processes (blue arrows pointing to the same green boxes), suggesting shared physiological processes underlying different phenotypes (green arrows initiating from the same green box).

2. Aims and Strategy

In this work, an attempt is made to demystify genetic mechanism of complex phenotypes. We approach this in three steps.

First, IMT-phenotype associations are identified as genetic evidence to detect phenotypic similarity. Multiple GWAS SNPs are assembled together if they are identified by omics dataset to influence the same IMT. For example, multiple GWAS SNPs influencing the same gene's expression level are assembled together, indicating such gene's association with target phenotype. Association strength is characterized by combining the GWAS SNP's summary statistics in a way similar to Fisher's combined probability test.

Second, PP-phenotype association are identified by grouping IMTs based on existing knowledge. Multiple IMTs are integrate together if they are classified into the same functional group based on existing knowledge. For example, multiple genes belonging to the same GO pathway are integrated together, indicating this GO pathway is contributing to the genetic underpinning of target phenotype. PP-phenotype association strength is characterized by an approach previously developed to identify a subset of functionally related genes that drive the similarity between two gene expression profiles[24]

Finally, we expand the PP-phenotype association by removing the involvement of existing IMT-interaction knowledge. Specifically, IMTs are not integrated based on knowledge such as GO terms but are integrated in an unsupervised manner. We employ algorithm previously built for the analysis of large-scale gene expression data to cluster phenotypes and IMTs at the same time. As a result, similarity between multiple phenotypes are captured by a subset of IMT where the IMT-phenotype association are statistically significant.

3. Result

3.1. Sherlock-II: a new computational algorithm to infer IMT-phenotype association

In order to investigate the underlying biology of multiple phenotypes simultaneously, we built a computational tool called Sherlock-II (a new generation of our previous published tool Sherlock[13]) to infer IMT-phenotype associations from GWAS and functional datasets in a statistically robust manner. Similar to Sherlock, Sherlock-II detects gene-phenotype association by comparing the GWAS SNP with the eQTL SNP of a gene, with the assumption that if the gene is causal to the phenotype, SNPs that influence the expression of the gene should also influence the phenotype, thus the eQTL profile of the gene should have significant overlap with the GWAS profile of the phenotype. The statistical framework used by Sherlock-II is also directly applicable to metabolite-QTL data [25], allowing us to infer associations between metabolite levels and phenotypes.

3.1.1. Metabolite phenotype associations

Table 3. 1 metabolite-phenotype association passing Q-value cutoff 0.05.

phenotype	metabolite	p-value	Q-value
ALZHEIMER	palmitoyldihydrosphingomyelin(d18:0/16:0)	7.38E-05	0.0445
AUTISM	pregnenolonesulfate	7.04E-05	0.0454
	X-15503	5.51E-05	0.0355
	homocitrulline	1.47E-06	0.000932
	alpha-glutamylglycine	0.000134	0.0426
	16a-hydroxyDHEA3-sulfate	0.000131	0.028
	X-11308	1.80E-05	0.0104
	dehydroisoandrosteronesulfate(DHEA-S)	5.25E-05	0.0169
	1-palmitoyl-2-linoleoyl-GPC(16:0/18:2)	5.11E-05	0.0169
	X-11850	7.55E-05	0.0427
BIPOLAR DISORDER	ergothioneine	8.01E-05	0.0484
BIPOLAR DISORDER AND SCHIZOPHRENIA	X-11441	4.34E-05	0.0223
CHRONIC KIDNEY DISEASE	methioninesulfone	1.50E-05	0.00609
	homocitrulline	4.26E-05	0.00915
	guanidinoacetate	1.89E-05	0.00609
	xanthine	0.000178	0.041
	homoarginine	0.000135	0.041
HIGH-DENSITY LIPOPROTEIN CHOLESTEROL	2-arachidonoyl-GPE(20:4)	8.41E-05	0.0498
PANCREATIC CANCER	X-17299	8.33E-05	0.0468
SCHIZOPHRENIA	methioninesulfone	2.15E-06	0.00139
	gamma-glutamylglutamate	3.97E-07	0.000246
WEIGHT	trigonelline(N'-methylnicotinate)	9.26E-05	0.045

Application of Sherlock-II to integrate GWAS with metabolite-QTL data leads to the identification of 469significant metabolites-phenotype associations. For example, we have replicated the association between methionine-sulfone and chronic kidney disease found in a previous study[25] with more supporting SNPs. Interestingly, metabolite-phenotype associations seem to be enriched for neuropsychiatric/neurodegenerative diseases when a stringent statistical threshold is applied (Q-value<0.05, Fisher's exact test p-value= 0.0264, see Table 3.1). For examples, we have identified a strong association between gamma-glutamylglutamate and schizophrenia (p-value=3.97E-07, Q-value=2.46E-04); this metabolite was observed to decrease

in schizophrenic patients and is one of five best discriminators between the schizophrenic patients and the controls[26]. We have identified palmitoyl-dihydrosphingomyelin as the most significantly associated metabolite with Alzheimer's disease (p-value=7.38E-05, Q-value=4.45E-02). Palmitoyl-dihydrosphingomyelin is a particular type of sphingomyelin and it has been reported that a significant increase of sphingomyelin levels have been observed in the cerebrospinal fluid from individuals with prodromal Alzheimer's Phenotype compared to cognitively normal controls [27]. For Autism, the most significantly associated metabolite is homocitrulline, and decreased homocitrulline concentration has been detected in Autistic patients[28]. In all the cases except methionine sulfone, the metabolite was implicated by multiple weak metabolite-QTL SNPs aligned to moderate GWAS SNPs, thus cannot be identified by genome-wide significant metabolite-QTL SNPs and GWAS SNPs alone (Figure 1.2).

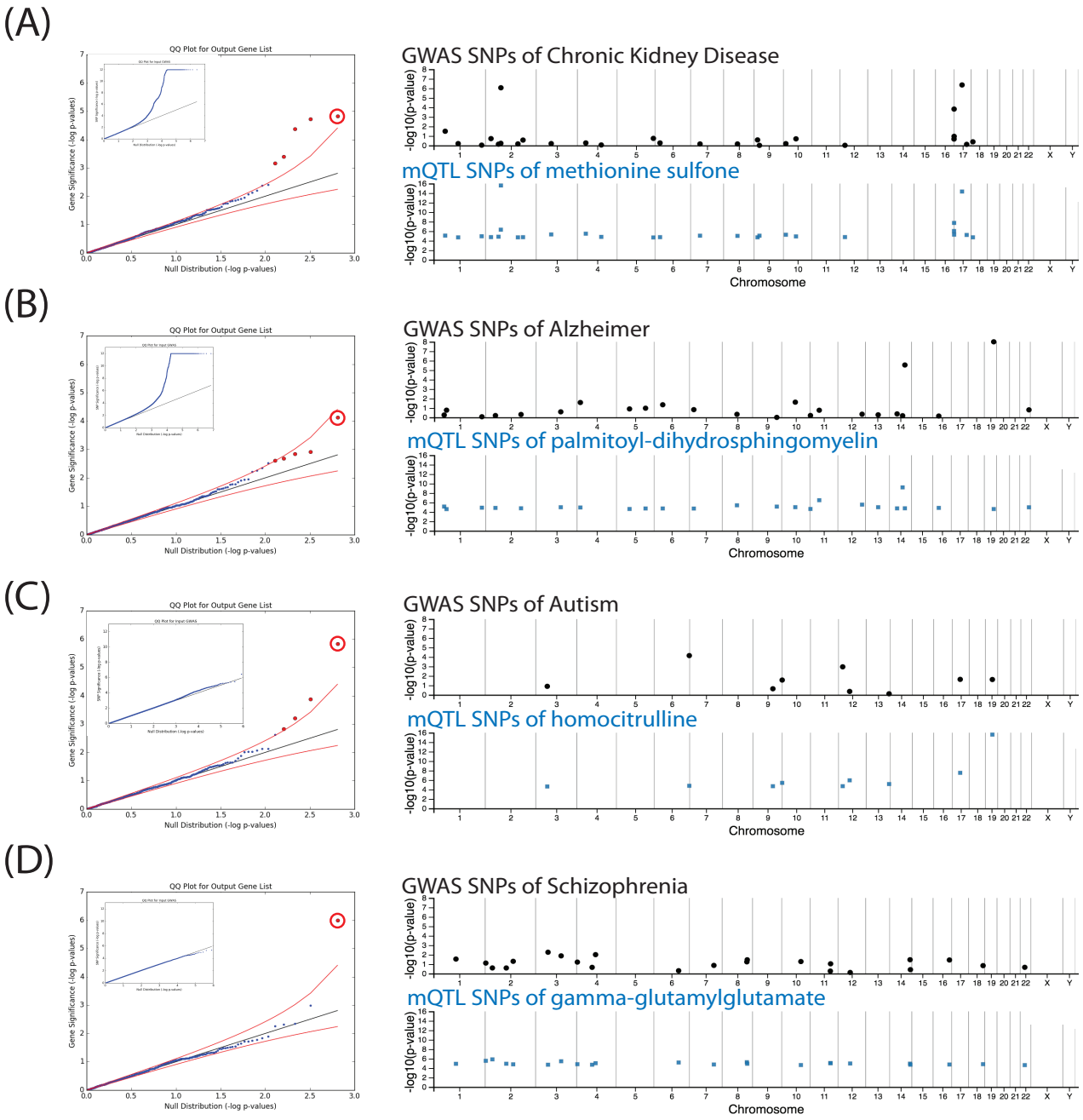


Figure 1. 2 Four examples of metabolite-phenotype associations

QQ plot for metabolite p-value distribution (left panel), QQ plot of original GWAS study (inserts in the left panel), and alignment between the metabolite-QTL and GWAS profiles (right panel) are presented. (A) Methionine sulfone is associated with chronic kidney disease. (B) Palmitoyl-dihydrosphingomyelin is associated with Alzheimer's disease. (C) Gamma-glutamylglutamate is associated with schizophrenia. (D) Homocitrulline is associated with Autism.

3.1.2. Gene phenotype associations

we have identified thousands of genes significantly associated with various phenotypes with Sherlock-II. Literature supporting these identified gene phenotype associations provides molecular detail that inspires mechanistic analysis. Take rheumatoid arthritis (RA) as an example. Three out of five most significant RA associated genes are from human leukocyte antigen family (HLA-DRB5, HLA-DPA1, HLA-DPB2), consistent with the established immune origin of RA [29]. The third most significant gene ITPR3(inositol 1,4,5-Trisphosphate Receptor Type 3) mediates the release of intracellular calcium, and has been identified as RA associated gene through a pathway based analysis [30]. The involvement of intracellular calcium signaling in the RA pathology is supported by single cell imaging showing abnormal intracellular calcium signaling in RA patients[29]. We also identified CACNB1(Calcium Voltage-Gated Channel Auxiliary Subunit Beta 1), further supporting the role of abnormal calcium signaling in RA. In addition, two other genes involved in regulating apoptosis (BAK1 and GSDMB) are significantly associated with RA, and cytokine induced cell apoptosis were observed in RA synovium[31].

3.1.3. Summary of gene/metabolite phenotype associations

Overall, we have identified 2114 gene-phenotype associations (with Q-value <1/3) covering 1421 genes and 70 phenotypes, and 469 metabolite-phenotype associations (Q<1/3) covering 308 metabolites and 51 phenotypes. To facilitate the analysis by the community we have provided a full list of the gene/metabolite-phenotype associations in appendix.

Using Sherlock-II, we have systematically analyzed the published GWAS data, using a set of eQTL and metabolite-QTL data we curated. The analyses include 445 GWAS datasets covering 88 phenotypes, 8 individual eQTL covering 6 unique tissues[32-39] , GTEx (version 7) data[40] , and 1 metabolite-QTL dataset [22]. To include the possibility that the expression of a gene can influence the phenotype through multiple tissues, we combined eQTL data across different tissues: when a SNP appears in multiple eQTL datasets, we assign the most significant p-value to it in order to capture the best eQTL signal across tissues. This merged version includes the eQTL profiles of 21,892 genes. For each GWAS dataset, Sherlock-II calculates a p-value for every gene measuring the statistical significance of the overlap between the merged eQTL profile of the gene and the GWAS profile. In many cases, although only weak individual associations are present in the GWAS data, Sherlock-II detects strong gene/metabolite-phenotype association signals as many GWAS SNPs are considering in aggregate during the transformation. By combining across multiple SNPs with weak associations, information present in GWAS associations at levels below traditional gnome-wide significance thresholds can be extracted. Our global analysis uncovers many gene/metabolite-phenotype associations that shed light on the genetic determinants of complex human traits, including disease risks.

The above results suggest that the significant gene/metabolite associations with other phenotypes identified by Sherlock-II are also likely to be biologically meaningful and detailed analysis can lead to new insight into the disease etiology.

3.2. Global analysis of the genetic similarity between phenotypes

3.2.1. Hierarchical clustering of phenotypes based on gene phenotype association profile

The global gene-phenotype association matrix not only allows us to identify genes associated with a specific phenotype but also enables the analysis of the genetic similarity between different

phenotypes at the gene level. Previous analyses of genetic similarity at the SNP level revealed certain overlapping SNPs between statistically associated phenotypes, lending genetic support to the observed co-occurence [41-43]. Genetic similarity at the gene level may yield stronger signal as multiple SNPs may converge at the same gene (e.g., influencing the expression of the same gene through different eQTL SNPs (eSNPs), thus one gene may be implicated in multiple phenotypes that do not necessarily share SNPs in the original GWAS data. Indeed, we have observed many instances where a gene's eQTL profile aligned to the GWAS profiles of two different phenotypes through different SNPs (see Figure 3.1 for an example). More importantly, genetic similarity analysis at the gene level also allows us to delineate the common cell/molecular mechanisms underlying different phenotypes, which we describe in more detail in the next section.

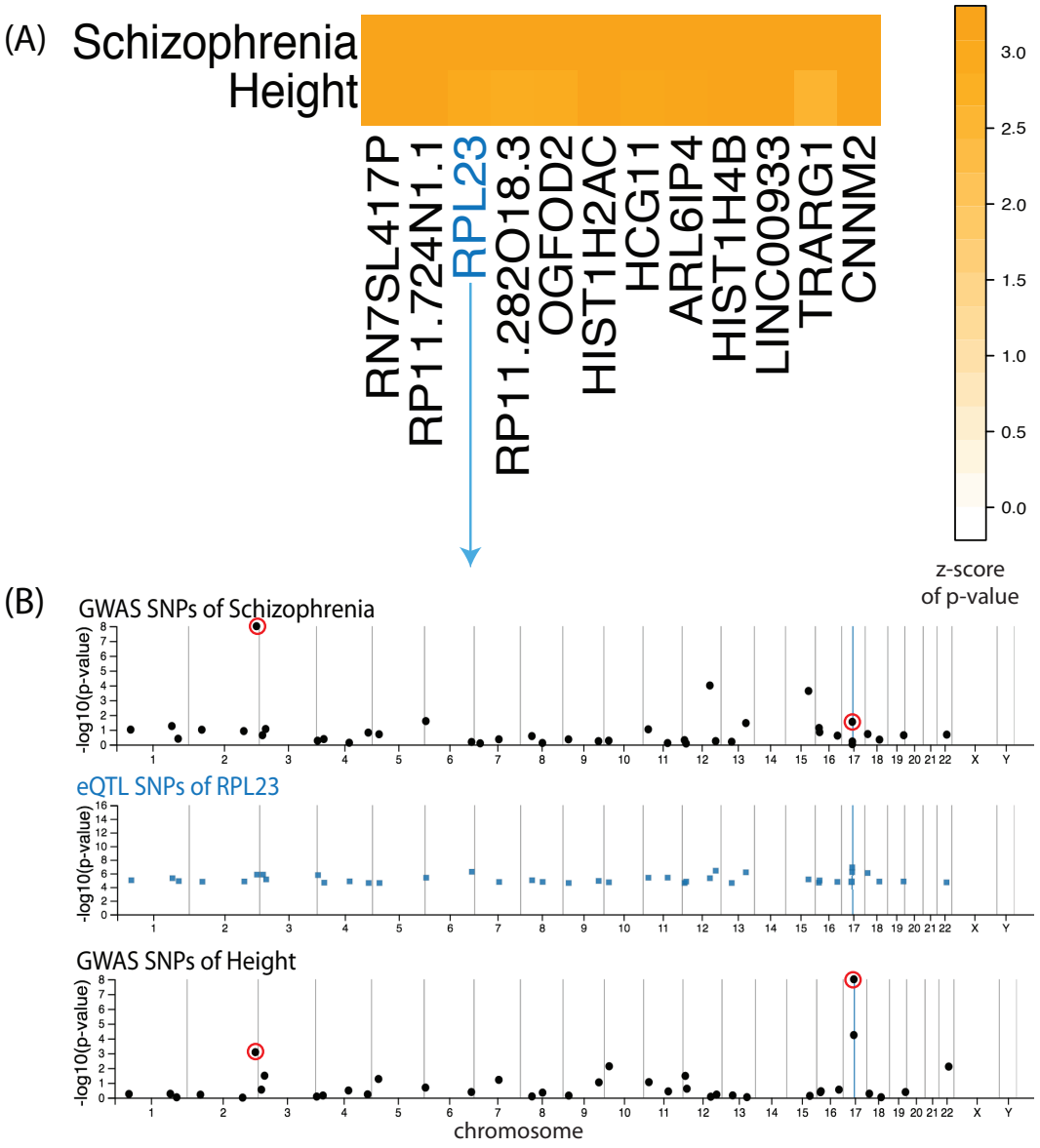


Figure 3. 1 example alignment

(A) Genetic similarity between schizophrenia and height supported by co-association with multiple genes. (B) the gene RPL23 is co-associated with both phenotypes, supported by eSNPs that aligned to non-overlapping GWAS peaks (highlighted with red circles) in two different phenotypes.

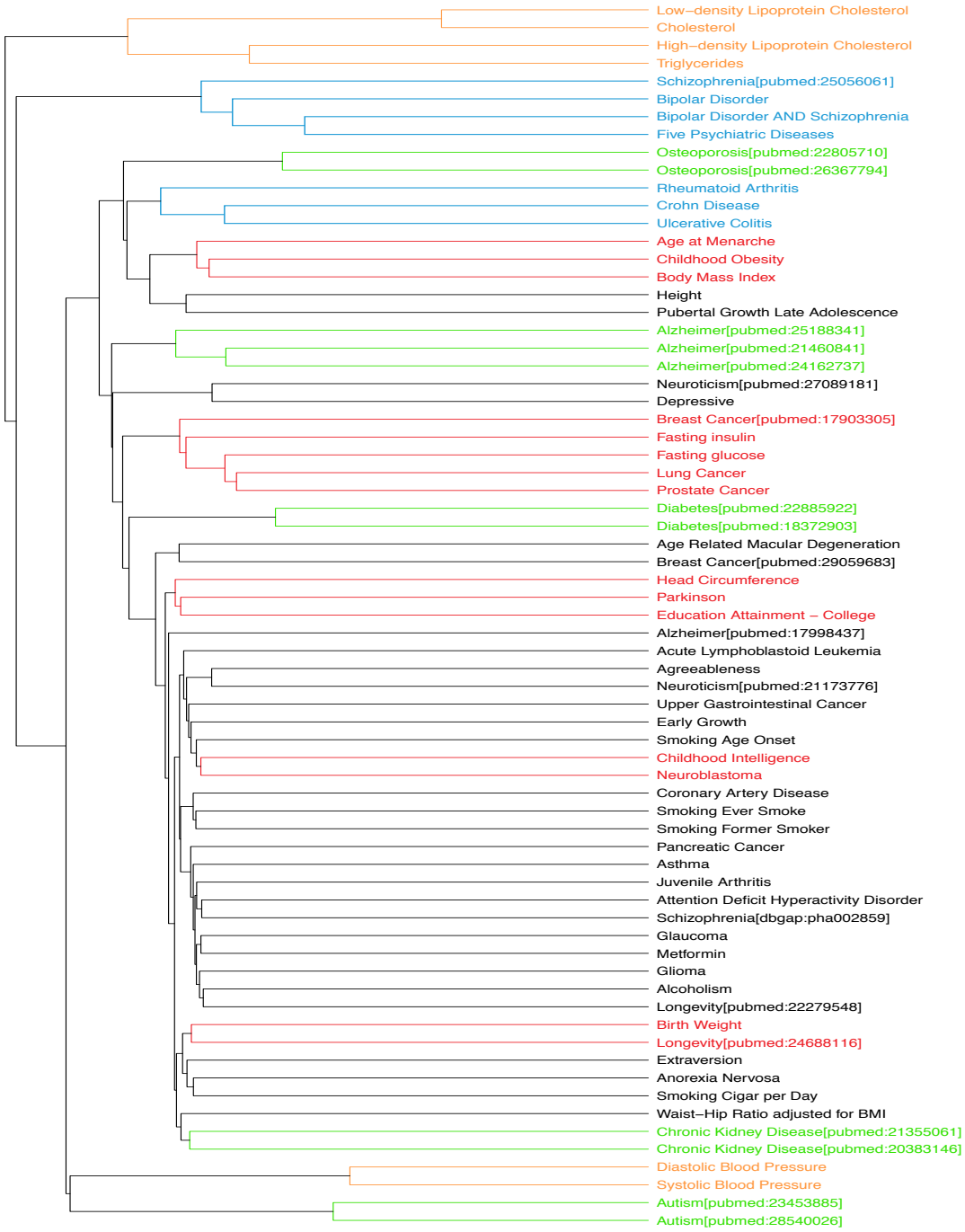


Figure 3. 2 global hierarchical clustering of phenotypes based on gene association profiles.

Highlighted branches discussed in the text: 1) GWASs of the same phenotypes (green); 2) phenotypes that are closely related and analyzed in the same GWAS (orange); 3) phenotypes that are known to share genetic etiology (blue); and 4) phenotypes with no obvious relations but co-occurences were observed based on previous epidemiological studies (red).

We first seek to identify the global similarity between two different phenotypes by comparing the gene-phenotype association vector (-log10(p-value) for each gene, with a dimension of the number of genes = 21892) for the two phenotypes. Intuitively, similar gene p-value vectors imply that the phenotypes respond similarly to the expression changes of individual genes. We applied hierarchical clustering to cluster all the phenotypes, using Euclidean distance between gene-phenotype vectors. The result of this global clustering analysis is a tree in which phenotypes with similar p-value vector are placed next to each other in a subtree (Figure 3.2). We find similar/related phenotypes were clustered together as expected. However, more interestingly we find phenotypes that are not obviously related are also clustered together, suggesting that they share similar genetic mechanisms.

The expected phenotype sub-clusters are formed by 1) GWASs of the same phenotype (e.g., osteoporosis, Alzheimer, diabetes, chronic kidney disease, and autism), indicating that at the global level the association was replicated; 2) phenotypes that are closely related and were studied in the same GWASs (e.g., LDL cholesterol, HDL cholesterol, total cholesterol, and Triglyceride; diastolic and systolic blood pressure); 3) phenotypes that are known to share genetic etiology in previous studies (e.g., bipolar disorder (BP), schizophrenia (SZ), the combination of the BP and SZ, and five psychiatric disorders including BP and SZ [44], and phenotypes with strong evidence supporting their shared mechanisms (e.g., rheumatoid arthritis, Crohn's disease, and ulcerative colitis, due to their common inflammatory/autoimmune nature). The fact that these different groups of related phenotypes were clustered together suggests that the gene-phenotype vector captured the major genetic information of the phenotype.

The global similarity analysis based on the gene-phenotype vector revealed a number of sub-trees in which phenotypes with no obvious relations are linked together (Figure 3.2, red

subtrees). Here we discuss a few examples: 1) a subtree that contains age at menarche, childhood obesity, body mass index; 2) a subtree with 5 phenotypes: breast cancer, fasting insulin, fasting glucose, lung cancer, and prostate cancer; 3) head circumference, Parkinson, and education attainment; 4) childhood intelligence with neuroblastoma; 5) longevity with birth weight. Upon literature search, we found that for each of the groups, there were previous epidemiologic studies that found the co-occurrence of the phenotypes: 1) a previous study found that girls experience early menarche are significantly often overweight/obese [45]; 2) hyperinsulinemia with insulin resistance is a significant risk factor for breast cancer independent of adiposity or body fat distribution [46]; 3) higher educational attainment significantly lower the risk for Parkinson's disease [47]; 4) children with neuroblastoma score a significantly higher IQ than the control [48] 5) mouse experiment showed that low birth weight correlates with increased longevity [49]. These were the observed correlations of the phenotypes, and the fact that they are clustered by the gene -phenotype association matrix indicates that they are linked by the associations with a set of genes that should shed light on the common underlying genetic mechanisms.

Delineating common genetic mechanisms underlying seemingly unrelated phenotypes

To identify potential common mechanisms that underlie two seemingly unrelated phenotypes, we search for genes that drive the similarity between the two gene-phenotype vectors. One straightforward approach is to analyze the most significantly associated genes with each of the phenotypes and test for significant overlap. This approach yielded insight for some of the phenotype subtrees. One example is the age-at-menarche/body-mass-index/childhood-obesity subtree. We found overlapping genes in each of the three pairwise comparisons (Figure 3.3). Multiple genes are associated with age-at-menarche and body-mass-index with a theme on cell growth and proliferation. For example, SMAD3 is significantly associated with both age-at-

menarche and body-mass-index and is a pivotal intracellular nuclear effector of the TGF-beta (transforming growth factor beta) signaling which play a critical role in regulating cell grow, differentiation and development[50]. MAP2K5 (Mitogen-Activated Protein Kinase Kinase 5) is also significantly associated with both the phenotypes and the signaling cascade mediated by this kinase is involved in growth factor stimulated cell growth and muscle differentiation [51]. Genes simultaneously associated with both childhood-obesity and body-mass-index include TFAP2B and CENPO. TFAP2B is a transcription factor of AP-2 family which stimulate cell proliferation and suppress terminal differentiation of specific cell types during embryonic development [52]. CENPO encodes a component of the interphase centromere complex which is required for bipolar spindle assembly, chromosome segregation and checkpoint signaling during mitosis thus is critical to cell division[53]. Thus examination of the shared significant gene associations suggests a common theme on cell growth and development underlying all three phenotypes.

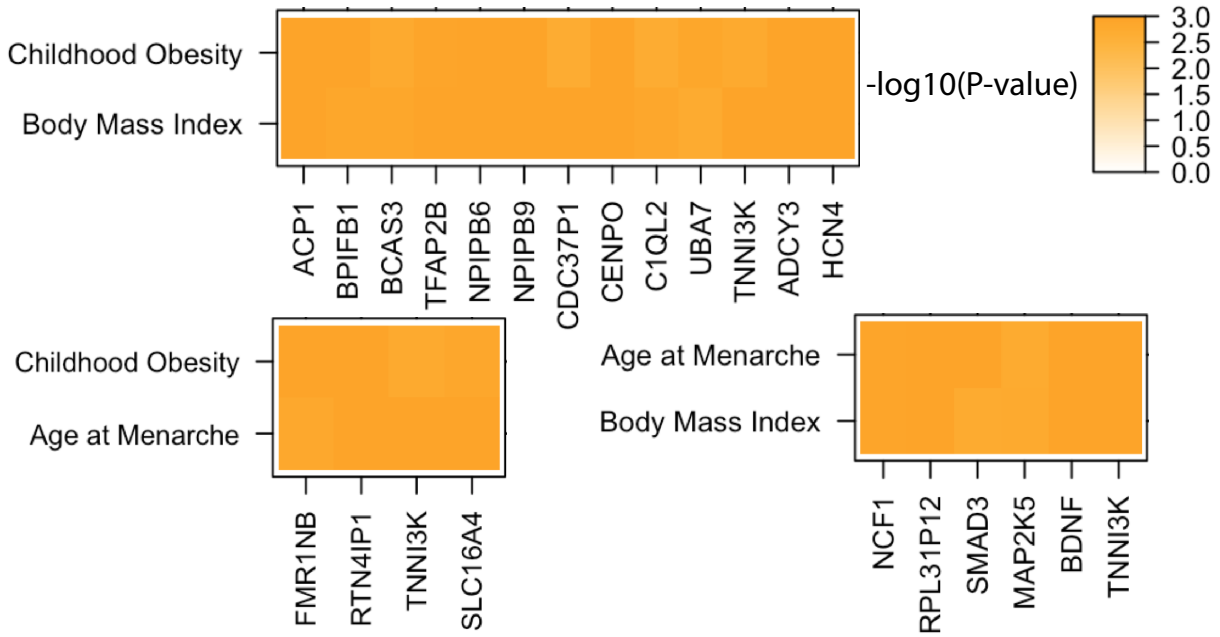


Figure 3. 3 overlapping genes between phenotype pairs

Overlapping genes in the three pair-wise comparisons of the phenotypes in the subtree (age-at-menarche/body-mass-index/childhood-obesity) from the hierarchical clustering of the phenotypes.

While comparing the top associated genes between phenotypes yields clues about the common mechanisms for some of the phenotype subtrees, its power of detection is limited. For many phenotype pairs that are similar based on their gene association p-value vector and thus clustered together, the top associated genes for the two phenotypes do not overlap (an example is the subtree with 5 phenotypes including fasting-insulin and breast-cancer). This suggests that the similarity is driven by more diffused signal distributed to a set of genes with weak associations. To identify such set of genes, we employ an approach previously developed to identify a subset of functionally related genes that drive the similarity between two gene expression profiles[24] . This approach assesses the contribution of a subset of genes defined by a GO category (or a pathway) to the Pearson correlation and assigns a z-score to the subset (see Method). This allows the ranking of the significance of different subsets of genes based on their z-score; we call it

3.2.2. Gene Ontology terms reveal common mechanism connecting phenotypes

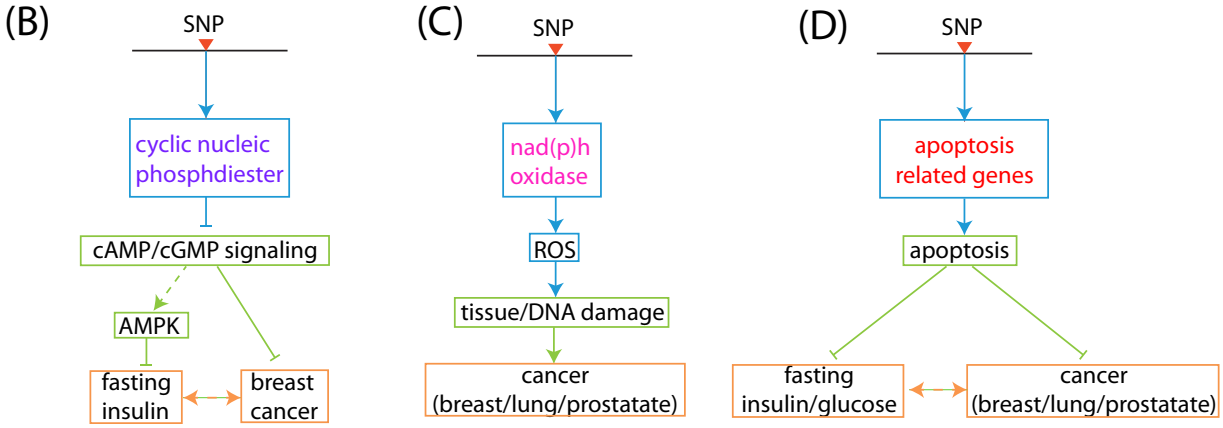
Using the Partial-Pearson-Zscore, we analyze the pairs of phenotypes in a subtree and found many significant GO categories that contribute to the similarity, suggestive of common underlying mechanisms for these phenotype pairs. As expected, for the subtree with rheumatoid-arthritis/Crohn's-disease/ulcerative-colitis, we found that the top GO categories that contribute to the similarity between RA and CD are "antigen processing and presentation of peptide or polysaccharide antigen via. MHC class II" (ranked 1st, z-score = 18.6) and similar GO categories ("MHC class II protein complex" ranked 2nd, and "peptide antigen binding" ranked 3rd), and pyroptosis (an inflammatory form of cell death) (ranked 6th, z-score= 14.8). Similarly, GO categories contributing to the similarity between RA and UC include pyroptosis (ranked ^{1st} z-score = 20.7). These results suggest that at least part of the common underlying mechanisms for these diseases are immune in nature and involve inflammation induced cell death.

Table 3. 2 Summary of the top GO categories connecting 5 phenotypes

(A): Top 5 non-redundant GO categories with highest Partial-Pearson-Zscore for each pair of the phenotypes. Recurrent GO terms across different pairwise comparisons were highlight with different colors. Also highlighted is the "cgmp catabolic process" that support the relation between breast cancer and fasting insulin. (B) to (D): proposed models with causal relations to explain the observed co-associations between the expression of the genes in the GO term and multiple phenotypes. Alternative models were considered in the discussion section.

(A)

Breast Cancer - Fasting insulin	Breast Cancer - Fasting glucose	Breast Cancer - Lung Cancer	Breast Cancer - Prostate Cancer	Fasting insulin - Fasting glucose	Fasting insulin - Lung Cancer	Fasting insulin - Prostate Cancer	Fasting glucose - Lung Cancer	Fasting glucose - Prostate Cancer	Lung Cancer - Prostate Cancer
cgmp catabolic process(12.1)	tail-anchored membrane protein insertion into er membrane(13.1)	nad(p)h oxidase activity(13.2)	apoptotic nuclear changes(32.8)	sorting endosome(32.1)	sterol homeostasis(12.0)	5'-3' exodeoxyribonuclease activity(10.7)	synaptic growth at neuromuscular junction(21.0)	tail-anchored membrane protein insertion into er membrane(12.5)	nad(p)h oxidase activity(14.7)
integral component of organelle membrane(8.9)	cytosolic large ribosomal subunit(11.2)	tail-anchored membrane protein insertion into er membrane(11.4)	ubiquitin-independent protein catabolic process via the multivesicular body sorting pathway(25.6)	oxidoreductase activity, acting on nad(p)h(17.6)	regulation of actin polymerization or depolymerization(8.5)	dendritic growth cone(8.2)	sterol homeostasis(9.8)	pore complex assembly(7.2)	tail-anchored membrane protein insertion into er membrane(9.2)
autophagy of mitochondrion(8.1)	regulation of lipid storage(9.2)	negative regulation of cysteine-type endopeptidase activity involved in apoptotic process(8.5)	polyubiquitin modification-dependent protein binding(25.4)	sterol homeostasis(17.2)	histone exchange(8.4)	negative regulation of synaptic plasticity(7.2)	regulation of signaling(9.1)	left/right axis specification(7.1)	sensory organ development(9.0)
establishment of protein localization to mitochondrion(8.1)	selenium binding(8.5)	positive regulation of inhibitory postsynaptic potential(7.9)	tail-anchored membrane protein insertion into er membrane(23.9)	zinc ion transmembrane transport(16.4)	gamma-tubulin complex(7.5)	peptidyl-diphthamide biosynthetic process from peptidyl-histidine(5.4)	regulation of intrinsic apoptotic signaling pathway(9.1)	acetylcholine receptor regulator activity(7.0)	apoptotic nuclear changes(8.7)
dendritic growth cone(7.9)	lsm1-7-pat1 complex(8.4)	outer dynein arm(7.9)	positive regulation of viral release from host cell(21.7)	negative regulation of intrinsic apoptotic signaling pathway(14.1)	apoptotic cell clearance(6.5)	ubiquitin-protein transferase activator activity(5.3)	tail-anchored membrane protein insertion into er membrane(7.7)	positive regulation of mrna 3'-end processing(6.7)	ubiquitin-protein transferase activator activity(7.2)



For the subtrees with phenotypes not obviously related, Partial-Pearson-Zscore analysis also revealed significant GO categories suggestive of the genetic mechanisms that connect these phenotypes together. An example is the subtree with 5 phenotypes including fast insulin and breast cancer (Figure 3.2). Significant GO categories were found for all 10 pair-wise comparisons (Figure 3.2). For example, we found that the "cgmap catabolic process" is the most significant GO category for the fast-insulin/breast-cancer pair. This category includes several

nucleotides phosphodiesterase, such as PDE4A and PDE2A. These are dual specificity cAMP and cGMP phosphodiesterase, which degrade cAMP and cGMP, and thus suppress cAMP/cGMP signaling. Inhibitors of these enzymes have been used as anti-cancer drugs[54], as the activation of cyclic nucleotide signaling is sufficient to inhibit proliferation and activate apoptosis in many type of cancer cells[54, 55]. cAMP signaling pathway is also strongly connected to insulin secretion. Among the various intracellular signals involved, cAMP is particularly important for amplifying insulin secretion[56]. Thus altered cAMP/cGMP signaling seems to connect with both breast cancer and fast-insulin.

Among the most significant GO terms from the pair-wise comparisons, "NAD(P)H oxidase activity" occurred multiple times and connects all three cancer types (breast cancer, prostate cancer, lung cancer). NAD(P)H oxidases are known to express at high level and produce ROS in cancer cells. Over-expression of these enzymes (possibly induced by inflammatory signal) has been linked to tissue injury and DNA damage from ROS that accompany pre-malignant conditions, contributing to the initiation and progression of a wide range of solid and hematopoietic malignances[57].

One of the most frequently appeared GO terms is related to apoptosis (Figure 3.2, colored red). This theme occurred in 6 out of 10 phenotype pairs and connects all 5 phenotypes. Defect in the normal programmed cell death mechanisms play a major role in the pathogenesis of tumor [58]. Apoptosis is also known to play an important role in the pathophysiology of both type I and type II diabetes and excess apoptosis is the underlying cause for beta-cell loss [59]. Thus altered apoptosis might be the common mechanism connecting the three cancer phenotypes and changed fasting-insulin/fasting-glucose level in normal people.

Another GO term that connect all the phenotypes is “tail-anchored membrane protein insertion into ER”. This GO term includes several genes (such as SGTA and BAG6) involved in the post-translational pathway through which tail-anchored membrane proteins are recognized and targeted to ER membrane[60] . This pathway is important in many cellular processes including the regulation of cell cycle. SGTA over-expression was involved in the pathogenesis of breast cancer [61] and BAG6 is involved in the control of apoptosis and is required for the acetylation of P53 upon DNA damage, linking them to cancer risk [53]. ER membrane targeting is also important for insulin secretion, potentially connecting these genes to the fasting insulin/glucose phenotype.

We propose models with causal relations to explain the observed co-associations between the expression of a set of functionally related genes and multiple phenotypes (Figure 3.2). These models are not unique, as there exists alternative and possibly less parsimonious models that can also explain the observed co-associations (see Discussion).

3.2.3. Two way bi-clustering detects genes associated with multiple phenotypes

The Partial-Pearson-Zscore approach can detect the co-association of a subset of gene defined by a GO term or pathway with a pair of phenotypes, and examination of all the pairwise comparisons can reveal common GO terms/pathways associated with multiple phenotypes clustered in a sub-tree. With the complex relations between physiological processes and phenotypes (Figure 1.1), it is possible that multiple genes not necessarily belonging to a predefined functional category can co-associate with multiple phenotypes not necessarily in the same sub-tree.

To detect the co-association of a subset of genes with multiple phenotypes, we applied a two-way bi-clustering algorithm – the Iterative Signature Algorithm (ISA) [62] to the global gene-

phenotype association matrix. This algorithm was developed previously to analyze gene expression data in order to detect a subset of genes with similar expression patterns across a subset of experimental conditions. The ISA algorithm starts with a seed of a sub-matrix and iterates through phenotypes and genes to recruit/drop members till the set of phenotypes and genes converge to a stable submatrix. Statistical significance of detected phenotype-gene sub-clusters are calculated by random permutation of the input p-value matrix within each phenotype (see Method).

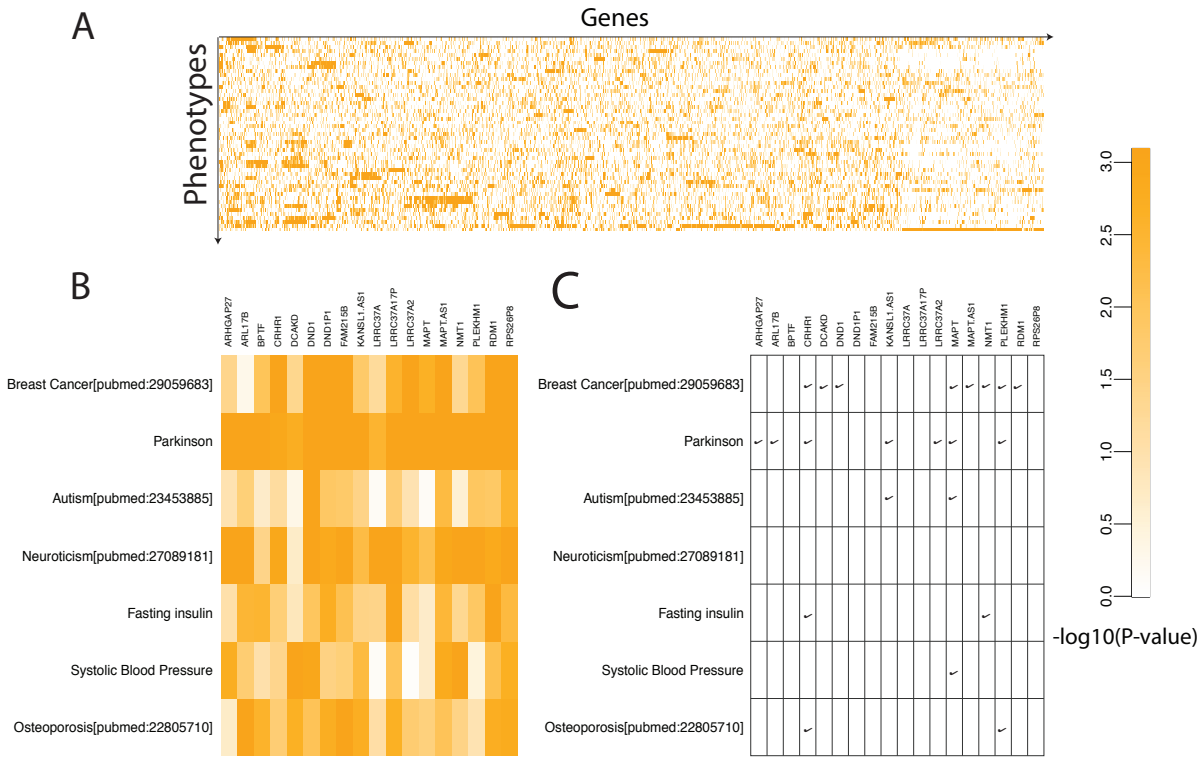


Figure 3. 4 Subclustering of phenotype-gene association matrix

Subclustering analysis revealed a set of genes co-associated with multiple phenotypes. (A) the global gene-phenotype association matrix, filtered to enrich the signal and reduce the dimension. The rows and the columns of the filter matrix were reordered by the hierarchical clustering algorithm. Certain subclusters are already visible. (B) One of the sub-clusters identified by the ISA algorithm. This subcluster consists of 18 genes and 7 phenotypes including a few neuro-psychiatric diseases, osteoporosis, breast cancer and physiological traits such as insulin level and blood pressure. (C) gene-phenotype associations in the subcluster supported by literature (check marks).

Starting from a filtered global gene-phenotype association matrix to improve the signal and reduce the search space for the ISA algorithm (see Figure 3.4 and Method), we have identified a number of significant sub-clusters that connect multiple genes with multiple phenotypes. In general, these sub-clusters involve phenotypes that are expected to be together as well as phenotypes that are seemingly unrelated, and genes with a range of different functional categories, suggesting a complex web of connections between physiological processes and phenotypes (Figure 1.1). Here we discuss the most significant sub-cluster and provide a list of sub-clusters discovered by the bi-clustering algorithm in Figure 3.5.

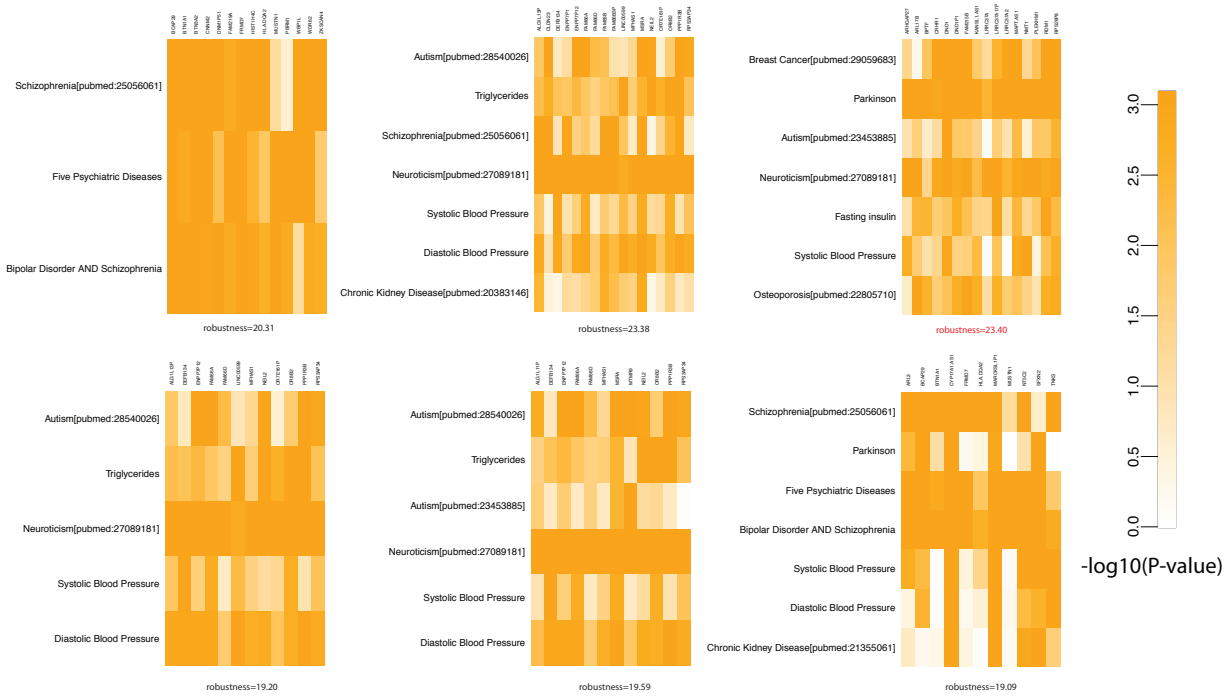


Figure 3. 5 all sub-clustering from phenotype-gene association matrix

Other examples of the sub-clusters discovered by applying the ISA bi-clustering algorithm to the gene-phenotype association matrix.

The most significant sub-cluster contains 7 diverse phenotypes and 18 genes with a broad range of biological functions (p-value = 1E-6 Figure 3.4). According to gene annotations [53], this set

of genes seems to be enriched for regulatory function: ARHGAP27 is a Rho GTPase activating protein, CRHR1 encodes a corticotropin-releasing-hormone-receptor, BPTF is a transcription factor highly expressed in the brain of Alzheimer's patients, DND1 is an inhibitor of miRNA mediated repression, FAM215B is a non-coding RNA, KANSL1 is a component of histone acetylation complexes, and PLEKHM1 is an adaptor protein that regulates Rab7 dependent fusion events in the endolysosomal system. Some of the gene-phenotype associations predicted by this sub-cluster are supported by literature (Figure 3.4 and table 3.3), and several genes associate with multiple phenotypes across the sub-cluster. For examples, MAPT (microtubule associated protein tau, known to be associated with multiple neurodegenerative and neuropsychiatric diseases [63]) is associated with Parkinson[64], Autism[65], and Systolic blood pressure [66]. Consistently we found MAPT antisense transcript with similar association patterns, as the expression of MAPT antisense is negatively correlated with MAPT expression [65, 67-70]. PLEKHM1 is known to be associated with breast cancer, Parkinson, and osteoporosis, and mutations in this gene are associated with autosomal recessive osteopetrosis type 6 [71-74]. These examples suggest that other gene-phenotype associations in this submatrix may also be true and biological meaningful associations. A comprehensive summary of literature support for gene-phenotype association is shown in table 3.3.

Among the seven phenotypes in this sub-cluster, several were known to be correlated based on previous epidemiological studies. For example, in addition to the connection between fasting insulin and breast cancer discussed above, insulin was linked to several phenotypes in this cluster: 1) there is a high prevalence of insulin resistance in non-diabetic Parkinson's disease patients [75] ; 2) it was suggested that abnormal insulin signaling underlies increased autism risk based on the evidence that mutations that hyperactivate PI3K pathway cause autism [76]; 3)

insulin resistance co-occurs with hypertension[77]; 4) insulin may work to stimulate osteoblast differentiation, and diabetes have a higher risk of bone fracture[78]. These evidences suggest that insulin signaling may be one of the physiological processes that connect these phenotypes together.

Functional analysis of genes in the sub-cluster added further support that regulation of insulin signaling is a common mechanism, with three of the genes in the cluster having strong connection with insulin. The first gene CRHR1 encodes a corticotropin-releasing-hormone-receptor. Administration of corticotropin, the downstream functional hormone of CRHR1, results in a 5- to 10-fold rise in plasma insulin levels in mice [79]. In ER positive breast cancer which constitutes about 80% of all breast cancers, estrogen alters the splicing of CRHR1 and disrupts CRH-mediated signaling, which contributes to the cancer growth driven by estrogen[80]. Thus CRHR1 may connect insulin and breast cancer through the downstream insulin signaling. The second gene is NMT1(N-myristoyltransferase). Its association with insulin is supported by mice experiment that liver particulate N-myristoyltransferase activity appears to be inversely proportional to the level of plasma insulin [81]. The third gene is PLEKHM1 which is an adaptor protein that regulates Rab7 dependent fusion events in the endolysosomal system[53]. PLEKHM1 participates in the degradation of insulin granule by the fusion of secretory granule with lysosomal, which is highly up-regulated upon starvation [82]. Thus loss of function mutations in this gene may lead to high insulin level that drives osteopetrosis.

Table 3. 3 literature search for the most significant sub-cluster. Number indicates pubmed ID

/	Breast Cancer	Parkinson	Autism	Neuroticism	Fasting Insulin	Systolic Blood Pressure	osteoporosis
ARHGAP27	/	25370933	/	/	/	/	/
ARL17B	/	25370933	/	/	/	/	/
BPTF	/	/	/	/	/	/	/
CRHR1	17097220/24526122	25370933	/	/	14301515	/	19801982
DCAKD	23227362	/	/	/	/	/	/
DND1	28191469	/	/	/	/	/	/
DND1P1	/	/	/	/	/	/	/
FAM215B	/	/	/	/	/	/	/
KANSL1.AS1	/	25370933	26777411	/	/	/	/
LRRC37A	/	/	/	/	/	/	/
LRRC37A17P	/	/	/	/	/	/	/
LRRC37A2	/	25370933	/	/	/	/	/
MAPT	20579400	20070850/21292315	26479762	/	/	23969178	/
MAPT.AS1	29441192	/	/	/	/	/	/
NMT1	30446635	/	/	/	8240341	/	/
PLEKHM1	28229117	25370933/21292315	/	/	/	/	17404618
RDM1	22016618	/	/	/	/	/	/
RPS26P8	/	/	/	/	/	/	/

4. Discussion

4.1. Advantage of similarity study on gene/metabolite level than on SNP level

We have developed a computational framework to study the genetic similarity between different phenotypes and to delineate the potential mechanisms underlying the similarity by integrating GWASs of multiple phenotypes with association studies of IMT. We demonstrated the approach by a global analysis of 445 GWASs covering 88 complex human phenotypes. Our analysis revealed expected as well as unexpected relationships between phenotypes, and suggested specific mechanisms that connect these phenotypes together. Our analysis also revealed many specific gene-phenotype and metabolite-phenotype associations that should motivate more detailed mechanistic studies in the future.

Comparing with genetic similarity analysis using GWAS data at the SNP level, our analysis framework has several advantages. First, comparing the genetic profiles at the gene level can yield stronger signal, as multiple SNPs can converge on the same gene. Even if the GWAS peaks of two different phenotypes do not align by proximity, they can be “aligned” through a gene, i.e., they may influence the expression of the same gene. We provided one example in

which the gene RPL23 is co-associated with both schizophrenia and height (Figure 3.1); the correlation of the two phenotypes has been observed in epidemiological studies [21] . The co-association is supported by two GWAS SNPs associated with the two phenotypes respectively. These two SNPs do not align by proximity; they align to two different eSNPs for the same gene.

Secondly, genetic similarity analysis at the gene level readily suggests potential mechanisms connecting different phenotypes. We have presented two different approaches to delineate genes/pathways that contribute to the genetic similarity: Partial-Pearson-Zscore analysis to identify a group of functionally related genes that contribute to the similarity between a pair of phenotypes, and a sub-clustering analysis of the global gene-phenotype association matrix that identifies a set of genes co-associated with multiple phenotypes. Analyses using these approaches led to specific hypotheses regarding the mechanisms that connect expected as well as unexpected relationships between phenotypes, and we expect that the power of detection will grow as more GWAS and IMT data accumulate.

4.2. Alternative mechanisms supported by the same genetic evidence

It is important to note that the data we obtained using this analysis framework are co-associations between genes and phenotypes supported by multiple GWAS and eQTL SNPs, not causal relations. In suggesting mechanistic models, we proposed specific causal relationships that are consistent with the data (Figure 3.2). However, there are often alternative (and perhaps less parsimonious) models that are consistent with the observed data. For example, we proposed that SNPs influence genes involved in apoptosis which can independently influence fasting insulin/glucose levels and cancer risks (Figure 3.2). However, there exists alternative models that are also consistent with the data. For example, SNPs can influence fasting insulin level which in turn influences apoptosis associated genes' expression which in turn influences cancer risk. A

variation of this model is that insulin can influence apoptosis associated genes' expression and cancer risk independently (Figure 4.1). These alternative models are also consistent with the co-associations between SNPs and all the phenotypes (including the gene expression phenotype), but less parsimonious in that there is a missing molecular trait in between SNPs and fasting insulin/glucose. However, the true mechanistic model is not necessarily the most parsimonious. Clearly more experiments/analyses are needed in order to discriminate between different models.

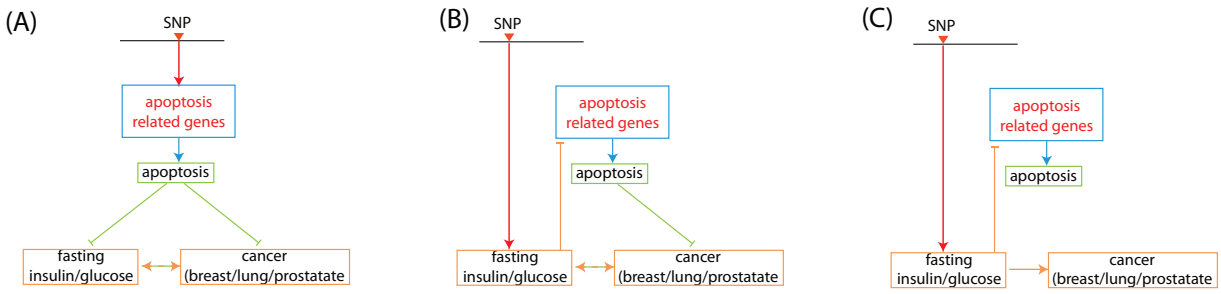


Figure 4. 1 Alternative mechanisms

Alternative models explaining the co-associations among the expression of apoptosis related genes, fasting insulin/glucose, and cancers. (A) the model we proposed in Figure 3.2. (B)-(C): alternative models that also explain the observed co-associations.

5. Method

5.1. Sherlock-II algorithm

Sherlock-II detects IMT-phenotype association by comparing the GWAS SNPs with the QTL SNPs of the IMT (eQTL or metabolite-QTL), with the assumption that if an IMT is causal to the phenotype, SNPs that influence the IMT should also influence the phenotype, thus the QTL profile of the IMT should have significant overlap with the GWAS profile of the phenotype. To test the significance of the overlap, Sherlock-II first selects the QTL SNPs of an IMT that pass a

threshold, aligns these SNPs to the corresponding SNPs in the GWAS, and test whether the aligned GWAS SNPs have p-values significantly better than those picked at random.

When testing the significance of a selected set of loci in a given GWAS, Sherlock-II uses a simple scheme to compute a score for the n p-values of the GWAS SNPs that best tag these loci. It then uses a convolution-based approach to construct an empirical null distribution for scores involving n p-values drawn randomly from the full set of GWAS p-values. In certain instances, particularly with eQTL for many genes, this test will be performed repeatedly for all intermediate molecular traits in a study. If dependence between the sets of IMTs is detected — for example, due to pleiotropic loci that regulate many genes — a provision of adjusting the null distribution is included. Through extensive testing and simulation, we show that our empirical strategy produces p-values that are highly resistant to inflation. In this section, we describe the individual components of the Sherlock-II method in detail in the order in which they occur.

Identifying Tag SNPs for Independent Blocks

Sherlock-II statistics is based on the assumption that the collection of SNPs tested are tagging independently segregating haplotype blocks in both the GWAS and eQTL cohorts. Understanding the true block structure of the cohorts is critical to accurately estimating gene significance. The inadvertent inclusion of dependent blocks can inflate the significance of genes by over-counting true associations; conversely, a conservative approach that enforces wide separation between tagging SNPs may miss many independent loci. The preferred approach to identifying independent blocks requires genotypes for both the GWAS and functional data to define block boundaries based on regions of rapid decay in linkage-disequilibrium (LD). However, since the difficulty of obtaining genotypes across a large number of GWAS studies makes this impractical, we constructed a database of LD using Caucasian cohorts from the 1000

Genomes Project. We use PLINK (v1.07) to compute r^2 linkage between common SNPs for 379 individuals in the CEPH, CEU, TSI, GBR, FIN, and IBS cohorts of 1000 Genomes release v3.20101123. This permits the alignment of associated loci in the functional data to corresponding SNPs in GWAS data, and it enables the identification of independently-segregating tag SNPs in the combined data for use in the statistical test.

We first match the GWAS SNPs to all the loci present in a given functional data set that pass a specific threshold. For the eQTL data used here, the nominal threshold is typically eSNP p-values $< 10^{-5}$. If matching SNPs are not present in both data sets, we use the closest GWAS SNPs in r^2 LD > 0.85, if available. We then use an agglomerative clustering approach to identify non-overlapping blocks of SNPs in r^2 LD < 0.2. From each block, the functional SNP with the strongest association is selected as a tag, and its corresponding GWAS SNP is included in the set subjected to statistical test. Since discrepancies between the various cohorts — GWAS, functional, and linkage — are inevitable, a minimum 100 kb distance between tag SNPs is enforced. In addition, we exclude SNPs from the human leukocyte antigen (HLA) region between 6p22.1 to 6p21.3 due to its complex linkage structure.

Core Statistical Method

Once a set of SNPs related to a specific molecular trait is identified, we compute a score s by simply combining their GWAS log p-values:

$$s = -\sum_{i=1}^{n} log_{10}(p_i)$$

This converts our collection of SNPs into a scalar quantity that, when referenced against the appropriate null distribution, indicates the statistical significance of selecting this subset from the pool of all independent GWAS loci. Our approach is analogous to Fisher's combined probability test using the scoring function $-2\sum_{i=1}^{n} log_e(p_i)$which follows a χ^2 distribution with 2n degrees of freedom when p-values are drawn from a uniform distribution on (0,1). With our method, instead of assuming a particular form for the distribution of GWAS p-values, we use discrete convolution to compute an empirically-derived distribution of scores when combining n p-values. Since s represents a specific score, we let S represent a random variable of all possible scores. We form the discrete probability distribution function (PDF) of S using bins of width b log units, where the probability of scores in the range $0 \leq s < b$ is the first element, $b \leq s < 2b$ is the second element, and so forth. Thus, $f_n[s] = P(S = s)$ is the discrete PDF for a score comprising n independent GWAS p-values. For the simplest case of scores involving only a single SNP, the PDF $f_1[s]$ is essentially a normalized histogram of p-values for all independent (unlinked) SNPs in the GWAS that aligned to a eQTL SNP across all genes:

$$f_1[k] = \frac{1}{N}\sum_{i=1}^{N} I[kb \leq -log_{10}(p_i) < (k+1)b]$$

where I is the indicator function, N is the total number of SNPs, b is the bin width, and k = s/b is the array index. For computational efficiency, the minimum GWAS p-value is typically truncated at 10^{-9}, well below genome-wide significance, yielding 900 bins when spacing b = 0.01 is used. This single-locus PDF forms the basis from which null distributions for any arbitrary number of functionally-related loci are formed. Since the sum of two independent random variables has a PDF equal to the convolution of their individual probability distributions,

scores involving two GWAS p-values have PDF $f_2 = f_1 * f_1$. In the general case for any value of $n > 1$, the PDF $f_n[s]$ is formed from n-1 recursive convolutions of $f_1[s]$:

$$f_n[k] = \sum_{l=0}^{k} f_1[l] \cdot f_{n-1}[k-l]$$

where again k = s/b is the array index. For the special case of a truly uniform p-value distribution, we have an exact continuous solution to the recursive convolution for n p-values (which can be obtained from the chi-square distribution)

$$f_n(s) = \ln(10)^n \prod_{i=0}^{n-1} \frac{1}{i} s^{n-1} 10^{-s}$$

To motivate the use of our empirical approach, we compare this theoretical result to our empirical method for GWAS of differing power. For low-powered association studies with p-value distributions close to uniform, the tail of the empirical distribution is similar to the tail of the theoretical distribution, leading to roughly identical estimates of significance for a given score. For well-powered meta analyses that appear to contain inflation from various sources, significant differences in the distribution tails can lead to an appreciable overestimation of significance for high scores when using the theoretical distribution. With the empirical distribution derived from the real GWAS data, Sherlock-II is insensitive to the input GWAS inflation.

Correcting for Pleiotropic and Sampling Effect

Another source of inflation stems from a lack of true independence between molecular traits across a given functional dataset. For example, pleiotropic loci may appear to regulate hundreds of genes in the eQTL data for a given tissue. In our simulations, these loci may inflate the output

test statistic due to chance alignment with significant GWAS SNPs; in our tests, this occurs at an appreciable level in perhaps five percent of the cases. Since our method permits the use of a unique distribution f_1 for each p-value added to the score, it enables a simple scheme for reducing inflation by adjusting each distribution based on the number of molecular traits affected by the locus. For the eQTL example, this involves conditioning the GWAS distribution based on the number of genes influenced by a locus across the entire eQTL data set. When chance alignment of pleiotropic loci and significant GWAS SNPs occurs, the overrepresentation of small p-values is reflected in the null distribution for affected genes. Thus, in practice, the actual distribution incorporated into a Sherlock-II score with eQTL data is conditioned on the number of genes that are regulated by the same locus:

$$f_{1|c\in C_{family}}[k] = \frac{\sum_{i=1}^{N} I[kb \le -log_{10}(p_i) < (k+1)b] \cap I[C_i \in C_{family}]}{\sum_{i=1}^{n} I[C_i \in C_{family}]}$$

where I is the indicator function, N is the total number of SNPs, k is the index, b is the bin width, and the ith SNP is assigned gene count C_i while the locus in question has count c. C_{family} is several non-overlapping sets, each including a range of gene count c. For example $C_{family6}$ includes all SNPs that regulate 6 to 10 genes. Now the distribution contains only the p-values of SNPs that belong to the same count family. With this change, instead of computing the significance of a set of SNPs drawn from GWAS, our method computes the significance of a set of SNPs drawn given their individual pleiotropies count family. In our simulations, this results in a minor change in the significance and rank order of output genes in most cases but prevents inflation whenever chance alignment of significant GWAS SNPs and pleiotropic loci occurs

(Figure 5.1). Specifically, we penalize SNPs regulating the expression of a large number of genes.

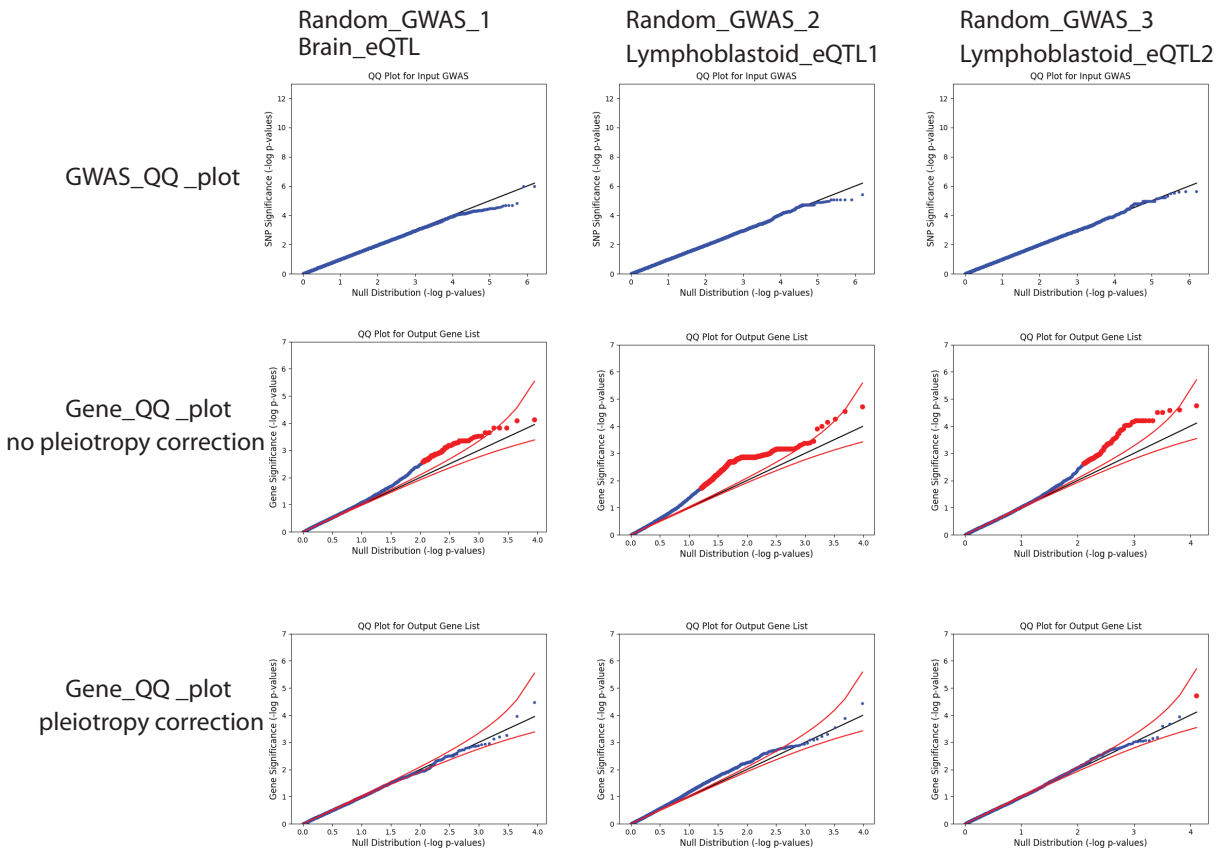


Figure5. 1 Sherlock-II is robust to inflation

Using p-value distribution conditioned on the pleiotropic counts of the SNPs prevents the inflation of the gene-phenotype association p-values. Top Row: QQ plots from case-control randomization GWASs show no signal. Middle Row: The chance overlap of pleiotropic loci in eQTL data and low p-value SNPs from the randomized GWAS can result in inflation of gene-phenotype association p-values. Bottom Row: the use of PDFs conditioned on SNP pleiotropy corrected the inflation.

5.2. Global analysis to identify molecular similarity between phenotypes

We first filtered the gene-phenotype association matrix to eliminate redundancies. For each phenotype with multiple GWAS dataset from overlapping cohorts, we chose a representative one with the strongest signal, defined as the one with the largest number of genes with p-value of

association smaller than 0.001. Note that the representative GWAS doesn't have to be the one with the biggest sample size because some signal is only statistically significant in a sub-population.

For genetic similarity analysis, we transformed the p-values in the gene-phenotype association matrix to -log10(p-value), so that genes with small p-values can have a stronger weight in the distance calculation.

5.2.1. Hierarchical clustering

Hierarchical clustering is applied to the gene-phenotype association matrix to group phenotypes together based on their genetic similarity. We used R and its default distance measure, the Euclidean distance, to calculate distance between phenotype p-value vectors. We then use Ward's method to cluster the phenotypes [83]

5.2.2. Partial-Pearson-Zscore analysis

We employed a simple statistical approach to detect gene ontology terms that contribute significantly to the genetic similarity between two different phenotypes. Gene Ontology gene annotations were downloaded from Ensembl. We removed the GO terms with less than 5 genes (not enough statistics) and more than 100 genes (too non-specific) from the analysis, after which 6796 GO terms were used. For a specific pair of phenotypes, we calculate a Partial-Pearson-Zscore for each of the 6796 GO terms using the following formula (previously developed for analyzing similarity between two gene expression profiles [24])

$$z_J = \sum_{i \in J} (x_i - u_x)(y_i - u_y)/\sigma_x \sigma_y \sqrt{n_J}$$

where xi and yi are -log10(p-value) for gene i associated with the two phenotypes, $u \quad and\ \sigma$ are the mean and standard deviation for x and y respectively, J is the gene set defined by the GO term, and n_J is the number of genes in the GO term. This formula is similar to the one for calculating Pearson correlation (with a normalization $\frac{n}{\sqrt{N_j}}$), except that the summation is partial – only for the genes in the subset. We argued that the null distribution for Z_j should be standard normal distribution, and verified that this is indeed the case by randomly sample all phenotypes pairs and generate Z_j for all the GO terms (Figure 5.2).

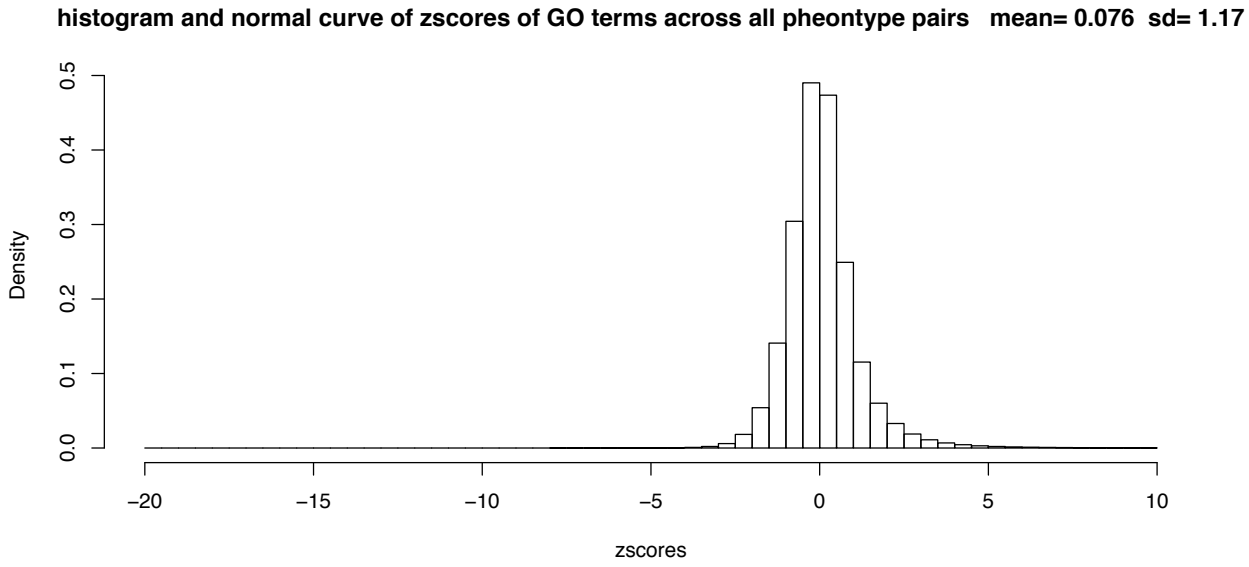


Figure 5. 2 partial-pearson-zscore distribution

Distribution of the Partial-Pearson-Zscore generated by analyzing GO terms for all the phenotype pairs. The distribution has a mean of 0.066 and a standard deviation of 1.06, which are very close to that from the standard normal distribution as expected from a z-score.

5.2.3. Two-way bi-clustering algorithm

In order to enhance signal and reduce the search space, we further filtered gene-phenotype p-value matrix before sub-clustering algorithm is applied. We removed the columns and rows in

which none of the gene-phenotype associations passed a Q-value cutoff of 1/3. Thus we only keep the phenotypes and genes with at least one significant association.

We used the ISA algorithm (Iterative Signature Algorithm) [62] to detect sub-clusters. This algorithm was developed to detect a subset of genes with similar expression patterns in a subset of conditions. The ISA algorithm starts with a seed of a sub-matrix and iterates through phenotypes and genes to recruit/drop members till the set of phenotypes and genes converge to a stable submatrix. By default, ISA randomly choose columns/rows as seed to begin iteration with the assumption that a large number of random seeds will cover most starting conditions that converge to stable submatrices. We went further by eliminating the randomness from seeding with an exhaustive set of seeds. More specifically we generated n(n-1)/2 number of seeds for n phenotypes to exhaust all possible combination of starting conditions. Standard post-processing steps with R ISA package [84] are conducted to merge similar sub-clusters.

Statistical significance of the discovered sub-cluster is calculated by permutation test. One permutation is defined by permuting the p-value vector of each phenotype in the gene-phenotype p-value matrix, keeping the p-value distribution within each phenotype invariant. Then ISA algorithm is applied with the same parameters used in the discovery phase to detect sub-clusters from the permutated matrix. A robustness score defined by ISA[62, 85, 86] is calculated for each sub-cluster discovered from the permutated matrix. The highest robustness of these sub-clusters is recorded for each permutation. Sub-cluster p-value is calculated by comparing its robustness score with the distribution of the robustness score generated from all the permutations.

6. Visualization

With high throughput statistical framework Sherlock II, we have generated millions of IMT-phenotype associations, each supported by dozens of SNPs. With this rich information comes the difficulty for data retrieval and inspection. By default, Sherlock-II outputs one txt file and one static html file for each GWAS-QTL pair. Although the html file provides a basic degree of visualization for individual Sherlock-II result file, it provides little help for the exploration of phenotype connections where multiple Sherlock-II result files are required to be analyzed together.

we need an interactive tool to allow user to make selection of phenotype combination and generate figures accordingly because the total number of phenotype combination is too large. Theoretically, the number of all combination of phenotypes scales exponentially (2^n) with respect to the number of individual Sherlock-II results. Given 445 GWAS datasets paired with several QTL datasets, the sheer number of Sherlock-II results is estimated to exceed 10^{40} which makes it time-wise too expensive and storage-wise inefficient to prepare visualization for all 2^n phenotype pairs.

6.1. Interactive web-interface

The interactive tool is called Sherlock-Vis and is implemented as a browser based visualization tool consist of three layers as shown in figure 6.1 flowing the MVC design pattern[87]. The top layer is a dynamic drawing layer where plots are drawn/updated right after user makes a selection. I implemented this layer using standard HTML and Javascript library called Data Driven Document(D3)[88] and rely on a third party server Flask to make the website public . The bottom layer can be thought of a warehouse storing all the Sherlock-II generated result

information. I base this layer on Oracle MySQL database. The middle layer that connects dynamic drawing layer and warehouse layer is called business logic layer. This layer is responsible first for communication that retrieve/store data from/to database and forward/receive data/command from browser. It is also responsible for all the data structure manipulation such as sorting and filtering. I implemented this layer from scratch using the general purpose programming language Python.

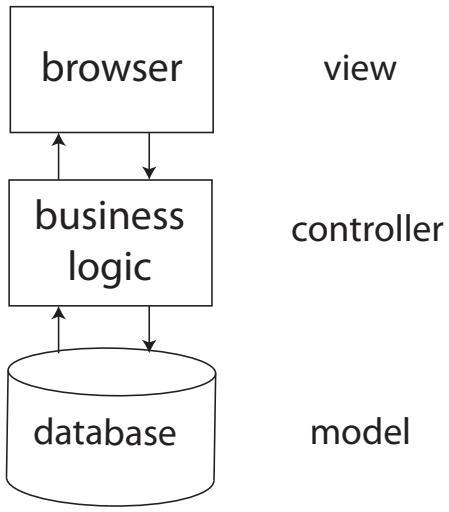


Figure 6. 1 design pattern of web interface

The MVC design pattern is implemented in three connected layers. Instruction is initiated from the browser, forwarded to database after data structure manipulation in the business logic layer with python. After data is retrieved from the database, they are feed back to the browser for display.

6.1.1. Browse gene-phenotype associations

In order to allow user selection, the web interface need basic input boxes including a way to select source GWAS, QTL. GWASs are not listed individually in a drop box because it can be more intuitive to allow users to select from 70 unique diseases rather than 445 GWAS datasets. This is partially feasible because there are rarely too many GWASs for the same disease. On the contrary, QTL datasets are listed individually due to the limited number of choices.

Another necessary input field is number of genes to display for each GWAS-QTL pair because firstly webpage cannot accommodate 20,000 genes at the same time and secondly genes with insignificant association statistics are excluded in most statistical analysis thus only genes with significant association statistics are frequently selected for display. To accommodate this use pattern, user can specify a number N (default 30) of genes to display which will be the top N genes with most significant association statistics. In this case multiple GWAS-QTL pairs are selected, top N genes for each pair will be joint together and the joint set are all displayed. In the case the joint set contains too many genes to display simultaneously, excess genes can be viewed by scroll the gene list to the right direction by mouse.

Finally, there is a unique use pattern when user needs only transcription factor displayed for each GWAS-QTL pair. A separate check box is created to accommodate this use case.

The browser interface is created as Figure 6.2.

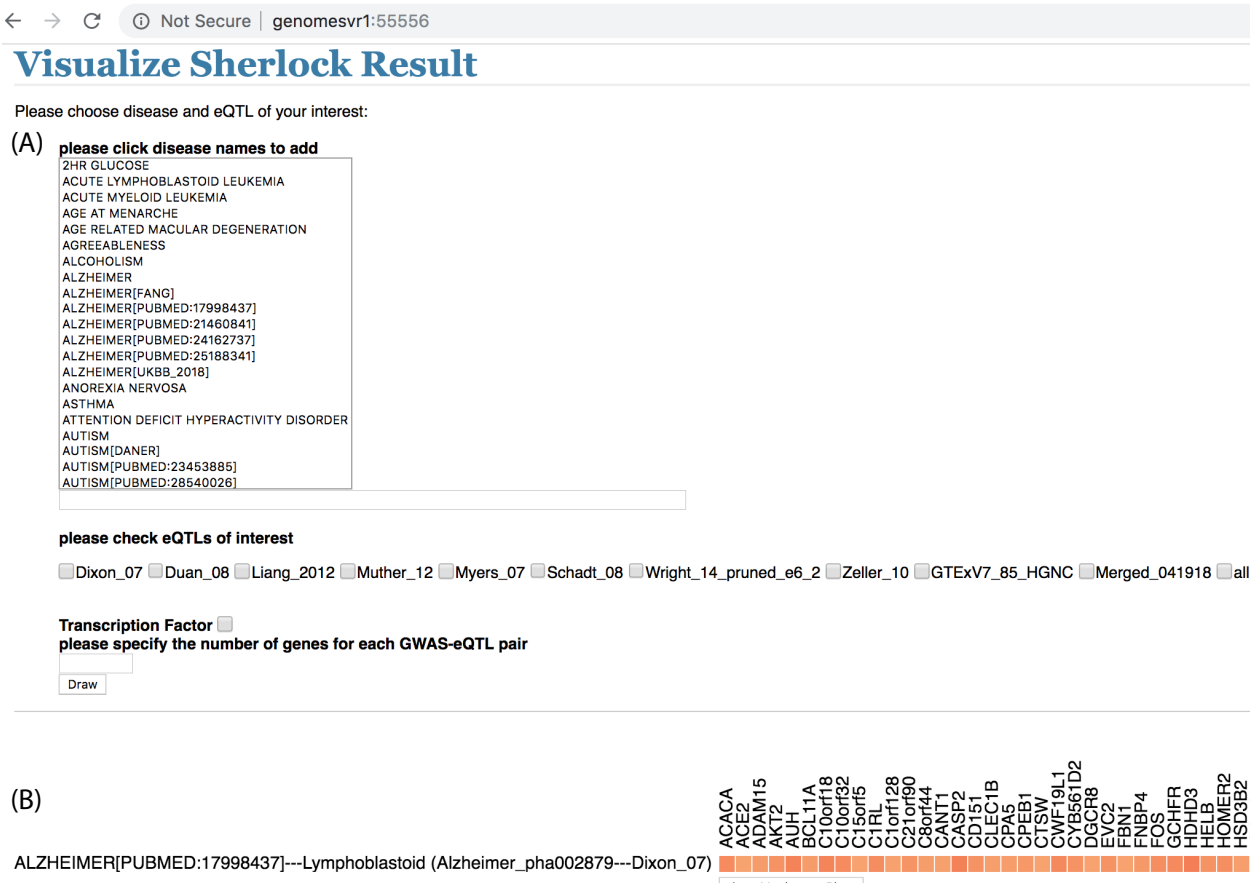


Figure 6. 2 web-interface overview

The web-interface. Part A: selection interface allows user to choose source disease and eQTL, specify the number of genes with most significant association to display, and choose whether only transcription factors be displayed. Part B: gene list generated for the selection.

After user makes a selection, a list of genes will be displayed for each GWAS-QTL pair. The GWAS_QTL pair is named in the format disease---tissue(GWAS---QTL) where disease corresponds to GWAS and tissue corresponds to QTL origin. By default, 30 genes with the most significant gene-phenotype p-value are displayed. The gene list is sorted in alphabetic order and color coded for gene-phenotype association p-value, more intensive color for more significant p-value. The exact association statistics can be inspected once cursor is hovered over the gene-disease association box.

Recall we need an Sherlock-Vis to accommodate phenotype combination rather than individual GWAS-QTL pair. To meet this need, Sherlock-Vis allows multiple GWAS-QTL pairs be displayed in the same plot as a response to multiple selection of source diseases as shown in Figure 6.3. The set of genes to display are the union of top genes from all GWAS-QTL pairs. A black box indicates that the gene-disease association is not available. This way it is easy to compare a gene's association with multiple phenotypes. For example, search for genes significantly associated with multiple phenotypes.

One recurrent search pattern for multiple disease-QTL pairs is enrichment study. Specifically, test whether genes significantly associated with one disease also have significant association with another disease. To address this need, Sherlock-Vis provides sorting function by clicking the target GWAS-QTL pair as shown in Figure 6.3. The gene list for all GWAS-QTL pairs will be reordered in the increasing order of gene-disease association p-value for target GWAS-QTL pair.

Similarly, it is fair to look for enrichment of diseases given a target gene. Sherlock-Vis provides sorting function by clicking the target gene as shown in Figure 6.3. The GWAS-QTL pairs will be reordered in the increasing order of gene-disease association p-value for target gene.

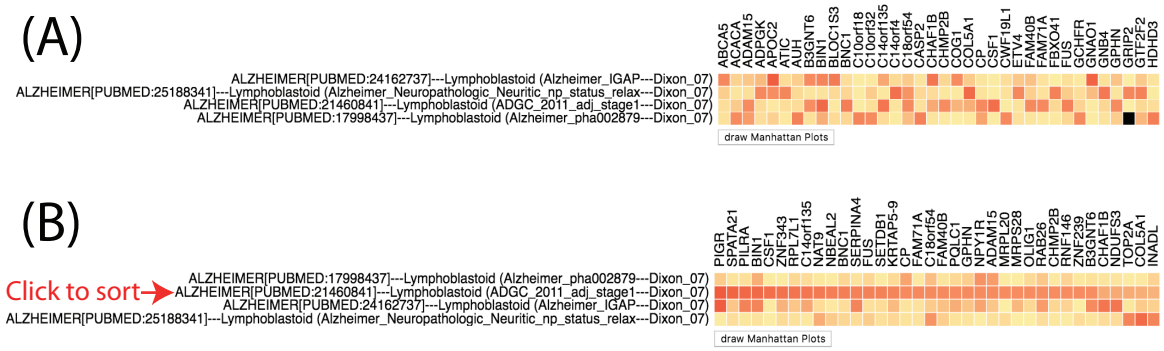


Figure 6. 3 multiple phenotypes on web-interface and sorting

(A) Multiple phenotype-eQTL pairs are displayed, gene set are the union of top 30 for each phenotype-eQTL pair, genes not displayed in Figure6.3 can be viewed by scrolling to the right by mouse. (B) gene list can be sorted with respect to a chosen phenotype-eQTL pair by clicking on the pair name. In this example, gene list is sorted by the second phenotype-eQTL pair, ALZHEIMER[PUBMED:21460841]---Lymphoblastoid (ADGC_2011_adj_stage1---Dixon_07) .

6.1.2. Browse SNPs supporting gene-phenotype associations

SNPs provides quantitative evidence supporting gene-disease association. Given Sherlock-II's capability of integrating influence from trans-SNP and SNP with mediocre association strength, it will be helpful to visualize all supporting SNPs for a target gene-disease association.

To meet this need, Sherlock-Vis provides separate plots to visualize GWAS SNPs, eQTL SNPs and their alignment relation. User can select a target gene to plot by right click the gene name in gene list of a GWAS-QTL pair (chosen gene will turn green to indicate selection). Clicking the "draw Manhattan plots" button will initiate a search for all SNPs supporting the gene-disease association.

The Manhattan plots are drawn for GWAS and QTL separately but using the same x-axis scale to facilitate alignment inspection. For both case, x-axis are chromosomes. Y-axis indicates association strength on the log scale because otherwise the difference between significant association statistics will be hard to tell. For SNP-GWAS associations, the y-axis is capped at global significance level 1E-8 so that any SNP with a more significant p-value can be truncated

to 1E-8 without losing information for the purpose of inspecting SNP's support for gene-disease association. Similarly, Y-axis is capped in a similar way. When Manhattan plot is drawn for only one gene, the GWAS SNPs are drawn as black dots while QTL SNPs are drawn as blue squares. Gene location on the genome is indicated by a blue vertical line.

While inspecting supporting SNPs on a genome level gives a global idea of all SNPs, it is necessary to inspect SNPs on a local scale for more details. For example, Sherlock-II allows SNP alignment by LD so a GWAS SNP and eQTL SNP can be aligned and possess different locations on the same chromosome. Such detail can only be viewed within a small range of genome. To accommodate for this need, Sherlock-Vis support zooming function along the x-axis. User can use the mouse scroll wheel as shown in Figure 6.4. On a small scale, the vertical blue line indicating gene location transformed into a blue square with both ends indicating location of first nucleotide of first exome and last nucleotide of last exome of the target gene, as shown in Figure 6.4.

While inspecting supporting SNPs on global and local scale provides qualitative information, quantitative details are also embedded in Sherlock-Vis. Recall Sherlock-II identified genes does not have to be the closest gene to GWAS SNP. For the purpose of comparison, Sherlock-Vis calculates the closest gene for each SNP and display this information once cursor is hovered over the target SNP. Together displayed also include the exact SNP-GWAS or SNP-QTL p-value on log scale and aligned GWAS SNP when cursor is hovered over an eQTL SNP.

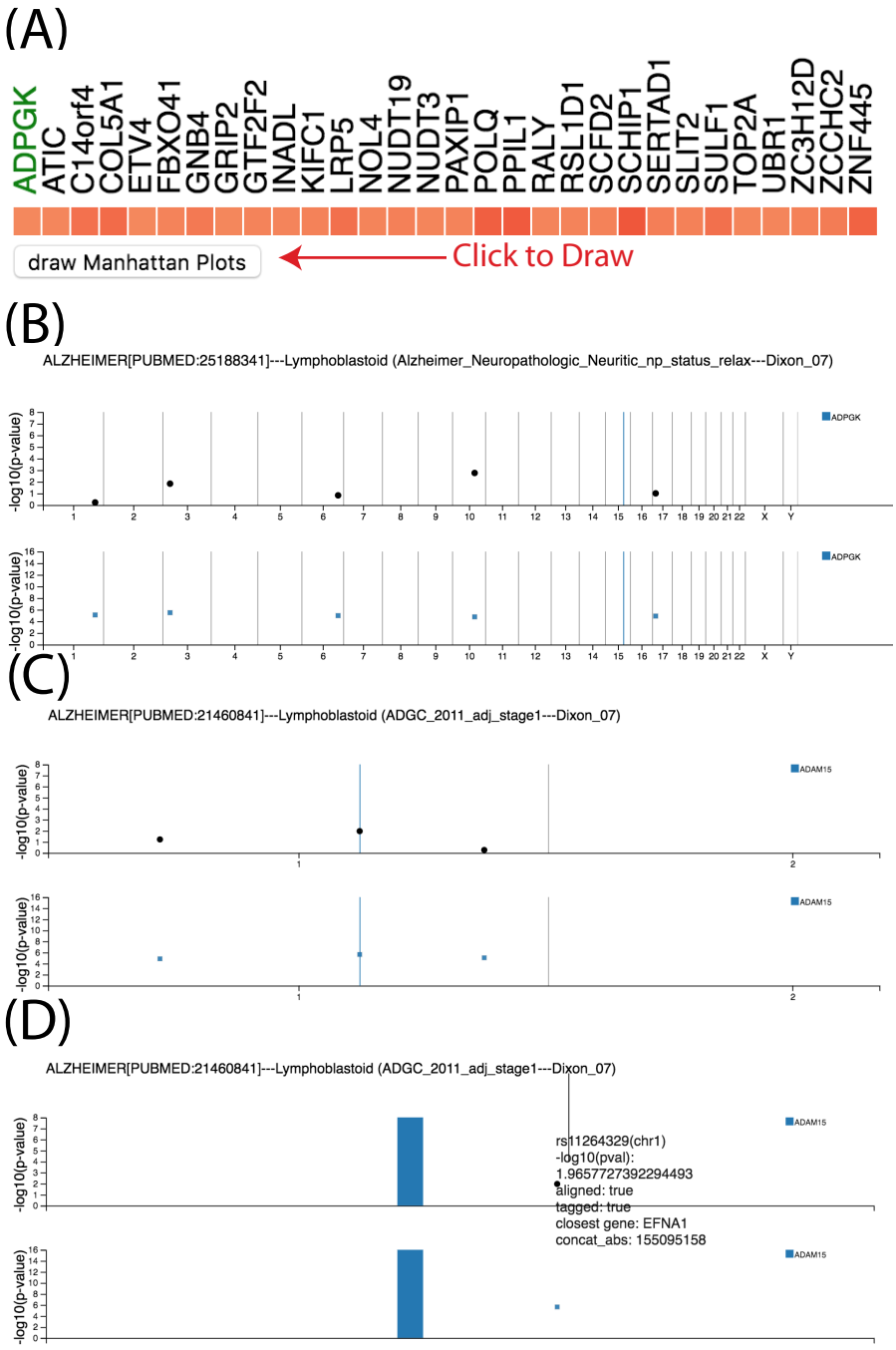


Figure 6. 4 Manhattan plot on web interface

Manhattan plot showing aligned SNPs supporting gene ADAM15's association with Alzheimer's disease. (A) Target gene is colored green to indicate selection. (B) Target gene's Manhattan plot consists of two halves, each having x-axis as chromosome. The top half has y-axis as GWAS p-value on log scale for each GWAS SNPs, indicated by black dots. The bottom half has y-axis as eQTL p-value on log scale for each eQTL SNP, indicated by blue squares. A color bar indicates gene location on the genome scale (in this example it is drawn on chromsome 1 due to target gene) (C-D) zoom-in view for the aligned SNPs can be achieved by the mouse wheel, SNP detail can be inspected by hovering over the SNP indicator(black dots for GWAS and blue square for eQTL).

Multiple genes, if simultaneously associated with the same phenotype, provide better evidence for the underlying mechanism. For example, existing knowledge such as gene ontology can be employed to explain the target phenotype. To provide genetic support for such multi-gene evidence, Sherlock-Vis allows the user to plot Manhattan plots for multiple genes at the same time as shown in Figure6.5. In the GWAS Manhattan plot panel, SNP sets supporting individual genes are united into one set. Multiple lines indicating gene location are plotted and color coded. In the eQTL Manhattan plot panel, SNPs supporting each gene are plotted separately with different color.

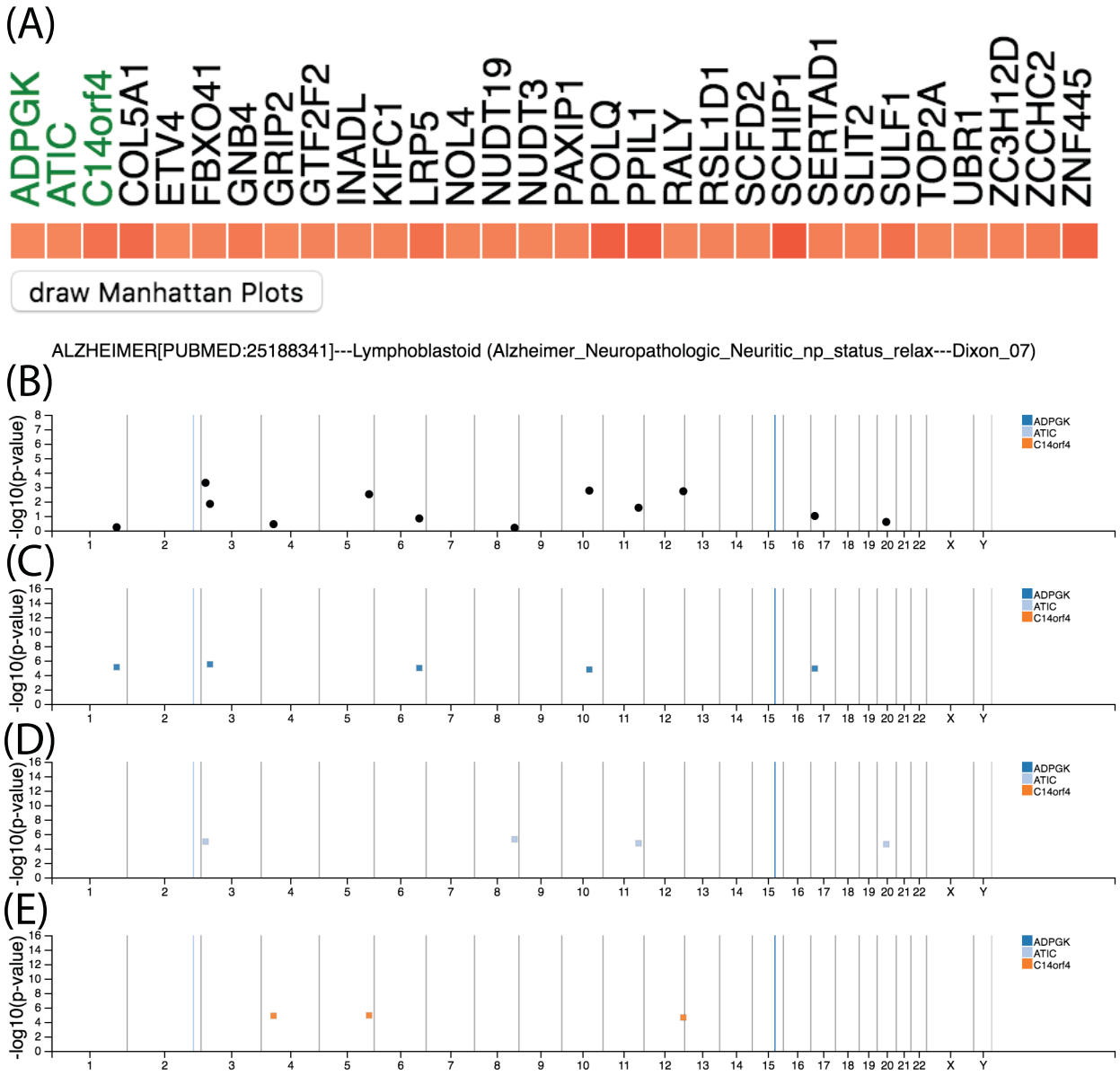


Figure 6. 5 Manhattan plot for multiple genes on web interface

Manhattan plot showing aligned SNPs supporting gene ADAM15, BIN1 and BNC1. (A) Target genes are colored green to indicate selection. (B) GWAS SNPs supporting all genes are plotted in the same panel. (C) eQTL SNPs of gene ADAM15 is plotted in blue color. (D) eQTL SNPs of gene BIN1 is plotted in cyan color. (E) eQTL SNPs of gene BNC1 is plotted in orange color.

Sherlock-II calculate gene-disease association statistics based on a subset of eQTL SNPs that align with GWAS SNP and are independent of each other. The independence here is defined by LD score of 0.2, specifically any two eQTL SNPs with LD score larger than 0.2 are considered

dependent and only one of them will be chose as tag SNP in subsequent calculation, the other SNPs are treated as untagged SNPs. The chosen one has the most significant influence on target gene expression.

It can be interesting to look at not only tag eQTL SNps, but also untagged eQTL SNPs as an intuitive inspection of Sherlock-II workflow. Sherlock-Vis is capable to reproduce Sherlock-II's entire alignment process as shown in Figure 6.6.

By selecting the untagged checkbox, untagged eQTL SNPs will be drawn besides tag eQTL SNPs by unfilled squares just like tag eQTL SNPs. All the details such as association statistics and nearest gene are included as well. The only difference between untagged and tag eQTL SNPs are their shape, indicating untagged eQTL SNPs do not contribute to gene-disease association calculation. Their corresponding aligned GWAS SNPs will be drawn as unfilled circles as shown in Figure 6.6.

Sherlock-II calculates gene-disease association first selecting a subset of GWAS SNPs that also influence target gene's expression. It then compares this subset of GWAS SNPs' p-values with all GWAS SNPs' p-value. A gene will have a significant association statistics if its subset of GWAS SNPs are highly enriched for significant GWAS p-values. Given this background, it could be straightforward if the comparison can be displayed. Sherlock-Vis satisfy this need by allowing all GWAS SNPs and eQTL SNPs for the target gene to be displayed at the same time. Similar to displaying untagged eQTL SNPs, target gene's subset of GWAS SNPs are solid circles while the other are unfilled circles as shown in Figure 6.6. A GWAS p-value threshold is enforced to filter out too weak GWAS signal which can be noise. The default threshold is set as p-value 1E-3.

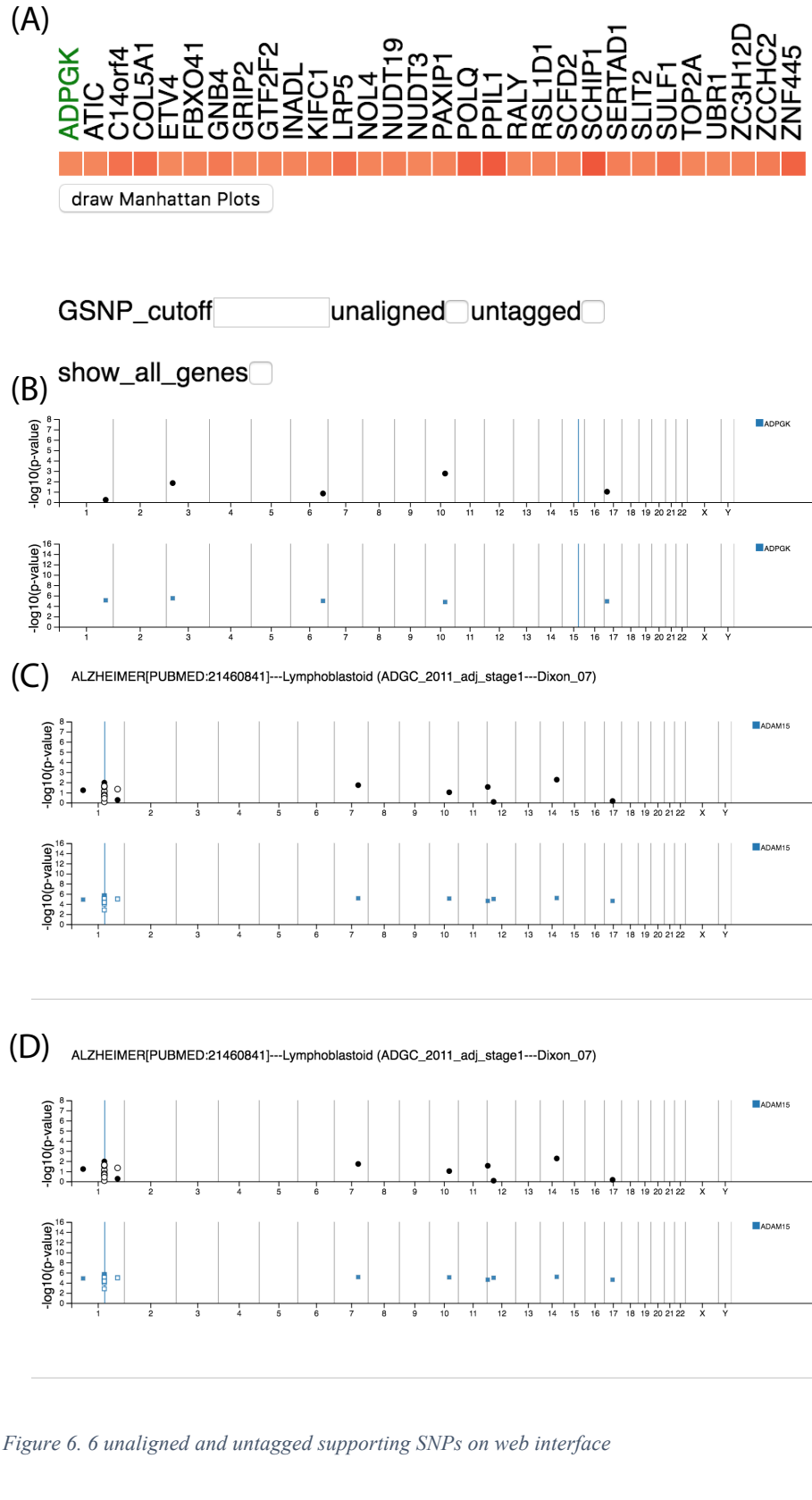


Figure 6. 6 unaligned and untagged supporting SNPs on web interface

However, Sherlock-II do this in a step-wise manner and include all eQTL SNPs that in LD with other eQTL SNPs

The Model part is supported by MySQL database storing SherlockII computed IMT-phenotype association and the supporting SNPs. Multiple schemas divide the whole logic into tables linked to each other through

6.1.3. Database models

The Model part is supported by MySQL database storing SherlockII computed IMT-phenotype association and the supporting SNPs. Multiple schemas divide the whole logic into tables linked to each other through foreign keys. The most important two tables are named gene_fields and metabolite_fields which stores IMT-phenotype association including the source GWAS, eQTL/metabolite-QTL and the association p-value and Q-value. The gene_fields table and its supporting tables are shown in figure 6.7. Metabolite_fields table has the same schema.

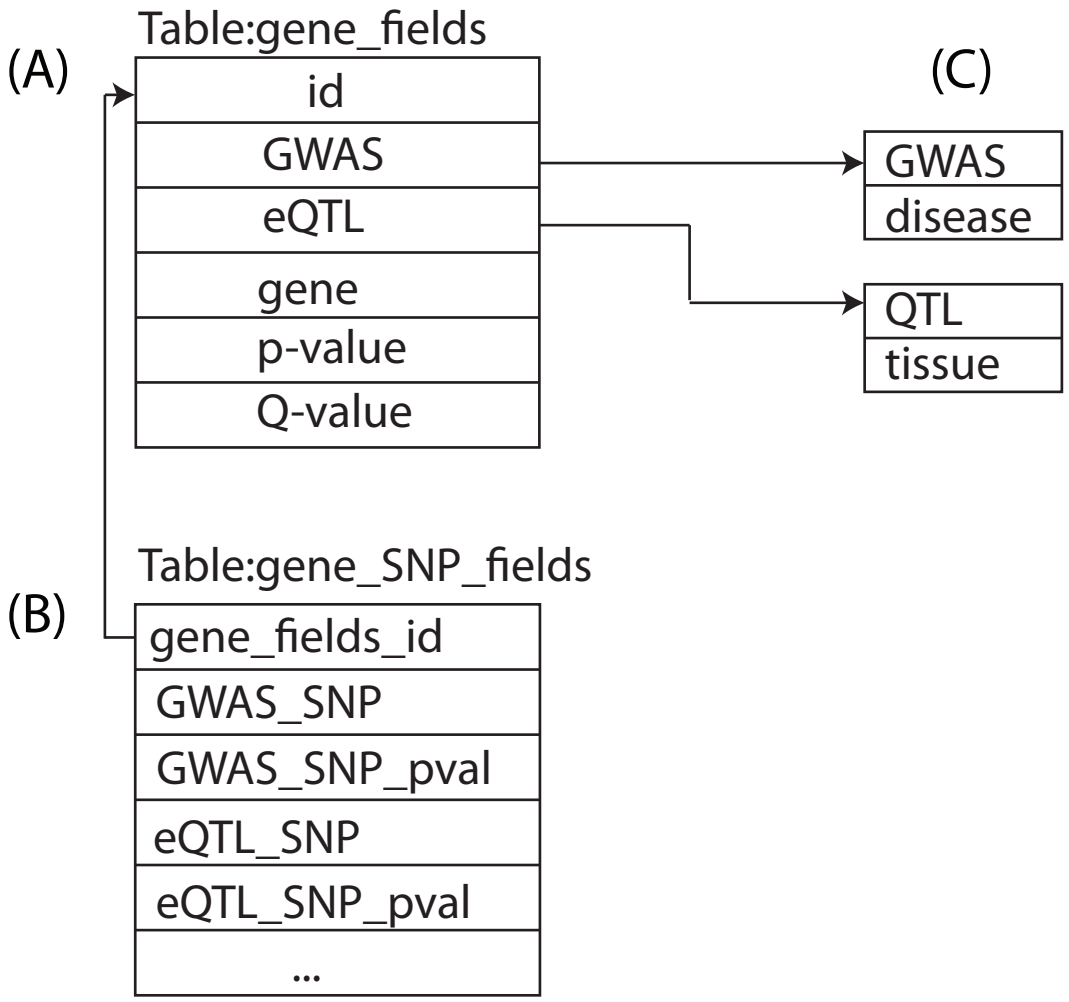


Figure6. 7 database schema

Data base schema for gene-phenotype association. (A) The main table gene_fields stores source GWAS, eQTL information and the associated gene and association statistics. (B) Supporting SNPs are stored in table gene_SNP_fields, linked to table gene_fields through foreign key gene_fields_id (C) GWAS's phenotype annotation and eQTL's tissue origin is stored in separate tables.

7. Partially completed Attempts and future directions

7.1. Allele direction

Motivation

Sherlock-II identifies IMT-disease association by summarizing the supporting SNPs' p-values. It will also be interesting to inspect individual SNP's direction. An example question to ask is whether minor allele of all SNPs supporting the same IMT are influencing its expression/concentration in the same direction? For example, if the minor allele of three SNPs supporting the same metabolite-disease association all reduces the metabolite's concentration and increases disease risk, more confidence will we have to conclude that decreased metabolite concentration will increase disease risk. Inference like this will be useful in the design of clinical intervention and could also help refine our hypothesis of disease etiology.

Current discovery

In a pilot study, I have detected SNP consistency supporting guanidinoacetate's association with chronic kidney disease. This association is supported by three SNPs rs7969761, rs10519022 and rs10519022. Specifically, A/T allele on rs7969761 reduces CKD risk and reduces guanidinoacetate concentration, A/T allele on GWAS SNP rs10519022 increases CKD risk and A/T allele on its aligned mQTL rs1153854 increases guanidinoacetate concentration, A/T allele

on GWAS SNP rs8101881 reduces CKD risk and A/T allele on its aligned mQTL rs10418164 reduces guanidinoacetate concentration.

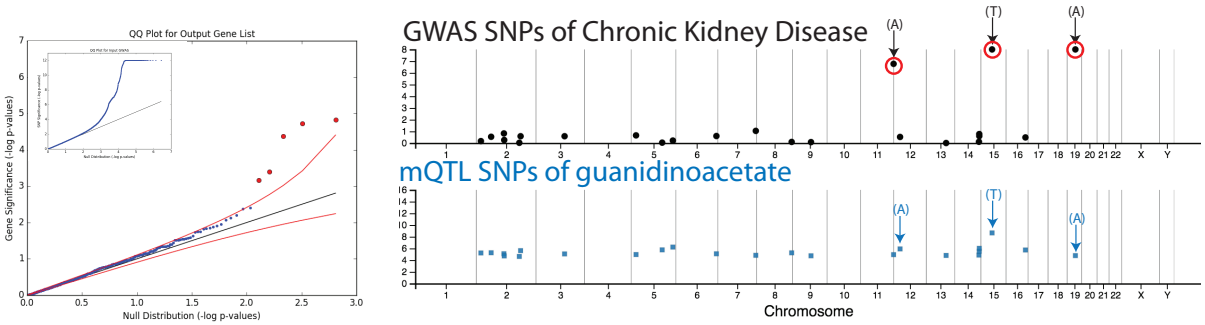


Figure 7. 1 SNP direction

three SNPs on three different chromosomes are modulating the gene expression level consistently and influencing disease risk at the same direction.

Method in the pilot study

The SNP consistency supporting guanidinoacetate's association with chronic kidney disease is discovered by first applying statistical filtering on supporting SNPs. Sherlock-II integrates multiple aligned GWAS SNPs together to generate gene/metabolite score, the majority of which have weak p-values. Although meaningful in aggregate, GWAS SNPs with weak p-value are vulnerable to statistical noise when considered individually. Thus it is only safe to consider GWAS SNPs with mediocre to significant p-values. Following this logic, I first set a mediocre statistical threshold for GWAS SNP's p-values (1E-3). In the example shown in Figure 7.1, such a GWAS SNP threshold excludes most supporting SNPs but does not affect the few comparatively significant ones. Applying this filtering step on all significant gene/metabolite-phenotype associations reduces the number of potentially interesting SNPs greatly which makes manual inspection a manageable task.

Points to improve

A complete consistency like Figure 7.1 is rare. In most cases, a majority but not one hundred percent of supporting SNPs have the same allelic direction. A statistical interpretation is needed to explain and benefit from such incomplete consistency. Once designed, such statistical interpretation can also help inspecting individual SNP after the SNP-level filtering and replacing the still-time-consuming manual inspection on individual SNPs.

7.2. eQTL Mechanism

Motivation

eQTL dataset provides the experimental observation of SNP-gene expression association, but does not provide the association mechanism, especially the eQTL SNPs that are far away from gene coding region. How can these variant influence a remote gene's expression? Such mechanism, once known, will provide more detail about disease etiology and further our understand. It will also reduce the risk of eQTL SNP expression associations being statistical noises and make Sherlock-II prediction more reliable.

Current discovery

To elucidate the eQTL mechanism, I collaborate with Shen lab in UCSF where they conduct promoter capture HiC experiment. The basic idea is that a eQTL SNP influences remote gene expression by falling into an enhancer region and affect target gene expression through promoter-enhancer interactions, or from random polymer looping, where undirected physical motion of chromatin causes loci to collide.

Using promoter capture data from Shen lab, I have detected triple alignment among GWAS, eQTL and promoter capture datasets. One such triple alignment contains a SNP that falls in a

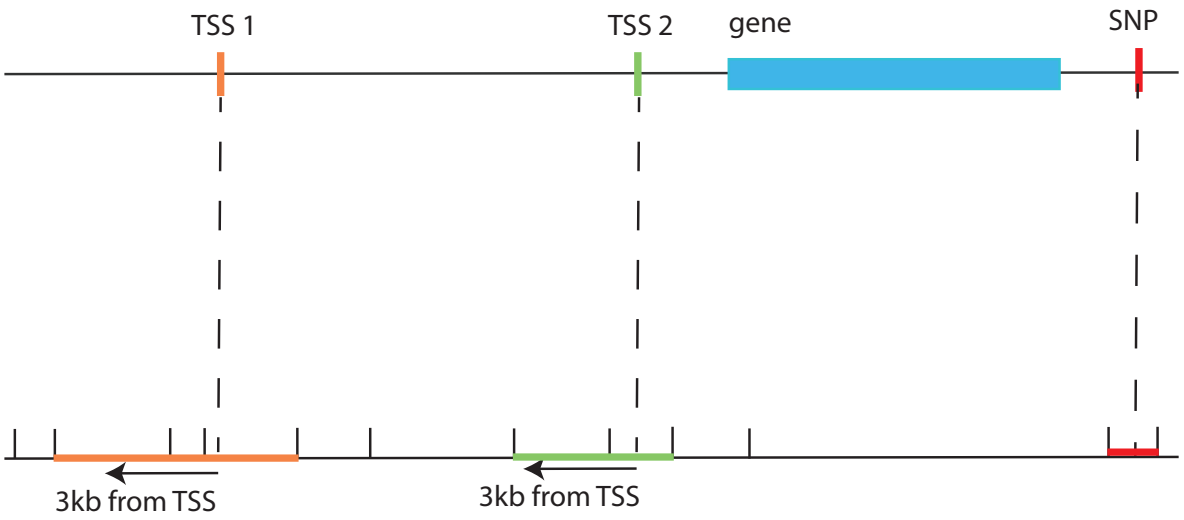
promoter capture enhancer/promoter region, being itself a eQTL SNP and also influence disease risk. The triple alignment was done using 445 GWAS datasets, 8 eQTL datasets and 7 promoter capture datasets including astrocyte, excitatory, hippocampal region and fetal interneuron, fetal inferior parietal lobule, fetal neuron, fetal Retinal ganglion cells. The triple alignment results cover all 88 diseases.

We can try to explain disease etiology by combining eQTL-promoter capture mapping with Sherlock-II results. In the case when an eQTL SNP align influences disease risk and also gene expression through promoter capture, we can hypothesize the SNP influences disease risk by first influencing gene expression in the enhancer-promoter interaction fashion, then the altered gene expression increases/decreases disease risk.

Method in the pilot study

The promoter capture dataset from Shen lab consist of pairs of chromosome segments that interact with each other. Each end of the pair consists of chromosome, chromosome start and chromosome end. At least one half of a pair is a gene promoter region while the other end can be a remote enhancer region or another promoter region.

One SNP-gene association provided by eQTL dataset is mapped to multiple promoter-capture associations because one gene can have multiple transcription start sites. For a given gene, I include all promoter capture pairs whose promoter region falls within a few kilo base pairs away from each of transcription start site as shown in Figure 7.2.



gene	SNP	SNP_bait	TSS1	TSS1_promoter_1
gene	SNP	SNP_bait	TSS1	TSS1_promoter_2
gene	SNP	SNP_bait	TSS1	TSS1_promoter_3
gene	SNP	SNP_bait	TSS2	TSS2_promoter_1
gene	SNP	SNP_bait	TSS2	TSS2_promoter_1

Figure 7. 2 promoter enhancer interaction

Points to improve

It will be interesting to investigate the proportion of eQTL SNPs explained by promoter capture data. We can also compare Sherlock-II results supported by eQTL SNPs explained by promoter capture data and the rest. For example, do Sherlock-II results supported by eQTL SNPs explained by promoter capture data have a less significant association statistic than the other? Answer to this question can be positive because the majority of eQTL SNPs explained by

promoter capture data are expected to be trans-eQTL and harbor a less significant influence on gene expression level. It will be equally interesting to look for disease enrichment in the result subset explained by promoter capture data. For example, enrichment of neural disease in the portion explained by promoter capture dataset, if identified, will provide detailed genetic evidence of both the underlying etiology and genetic similarity across multiple neural diseases.

It will also be interesting to look for eQTL mechanism other than promoter capture. One promising type of functional dataset is methylation where methylation of the base altered the proximity of binding proteins and influence gene expression. One particular advantage of the methylation dataset is its classification of association by tissues which match eQTL well. For example, the roadmap epigenomics consortium have mapped more than 100 tissues and cells types[89]. Such a rich collection of source tissue will allow fine mapping between eQTL and methylation datasets. At the same time, it is worth mentioning epigenomics dataset's caveat that methylation datasets do not provide high resolution on the genome scale. Biologically, multiple proximal nucleotides can be affected by the same methylation modification, making it hard to differentiate individual SNP when explaining eQTL datasets.

7.3. Dimension reduction

Motivation

Comparing phenotype similarity using more than 20,000 genes is possible but less intuitive comparing with the alternative if we can extract the most important few features for each phenotype. In the extreme case where we can use two aggregate pseudo-gene to represent 20,000 genes, the 88 phenotypes can be drawn on a two dimensional surface. This way phenotype similarity can be inspected visually.

Method in the pilot study

Principle Component Analysis(PCA) is a widely used dimension reduction technique. The intuition is to project each phenotype vector in the input gene-phenotype association matrix onto its eigenvectors. By definition, such projection will decompose input onto multiple orthogonal dimensions with the first dimension capturing the most variance, the second dimension capturing the second most variance … One particular convenient feature of PCA is that the variance captured by each orthogonal dimension is the eigenvalue in the process of eigen-decomposition. In this way we can visualize the reducing trend of variance explained by each dimension and dimension reduction can be achieved when the first few dimensions explains the majority of toal variance. For example, if the first 2 dimensions explain 95% of the total variance in a four dimensional dataset, we can safely reduce each data point in the original data set from four dimension to 2 dimension.

Our gene-phenotype association matrix is composed of 69 phenotypes and 21892 genes. Applying PCA on this association matrix gives 68 principle components because the number of principle components is bounded by the minimum of phenotypes and genes.

Unfortunately, the first few dimensions do not represent the majority of input variances as shown in Figure 7.3. The first two principle components explain less than 7 percent of the total variance which makes it inappropriate to discard the other dimensions because doing so will lose too

much information.

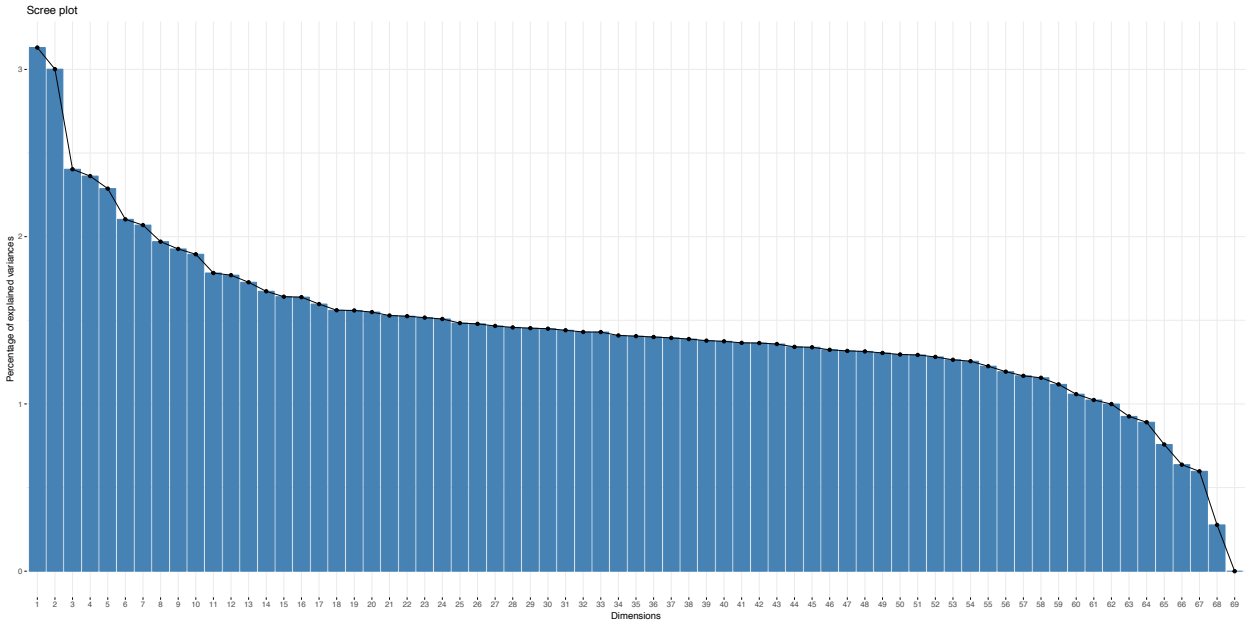


Figure 7. 3 PCA percentage variance

percent of variance explained by PCA principle components. The x-axis is principle component 1 to principle component 68, y-axis is percentage of variance explained.

Points to improve

Because the first few principle components do not explain the majority of variance, dimension reduction in this case will lose much information. Further exploration of data preprocessing is needed in order to allow PCA to function well in dimension reduction. For example, Shelrock-II outputs gene/metabolite-phenotype association p-values and by far they are transformed into log scale for visual inspection of extreme small p-values. There may exist other p-value transformations that provides clear separation of extreme small p-values but at the same time offer a better dimension reduction potential.

One transformation other than log scale transformation is called inverse normal transformation where p-value are assumed to be the complement of cumulative probability function of a hidden normal variable. In this way, a p-value 0.5 is mapped to z-score of 0, a p-value 0.001 is mapped

to 3.090232 because the probability for a standard normal variable to take value at least 3.090232 is 0.001.

Comparing with log scale transformation, the inverse normal transformation provides the same resolution for the region between 0.00051 and 0.30106 but provides better resolution for the less significant region 0.02742 to 0.30106 as shown in Figure 7.4. This will help differentiating significant versus non-significant p-values. Another difference lies for the region between 0.5 and 1 which are all mapped to 0 in inverse normal transformation numerically. This mapping does not result in a loss of information for the purpose of clustering because nothing within this region is considered significant associations.

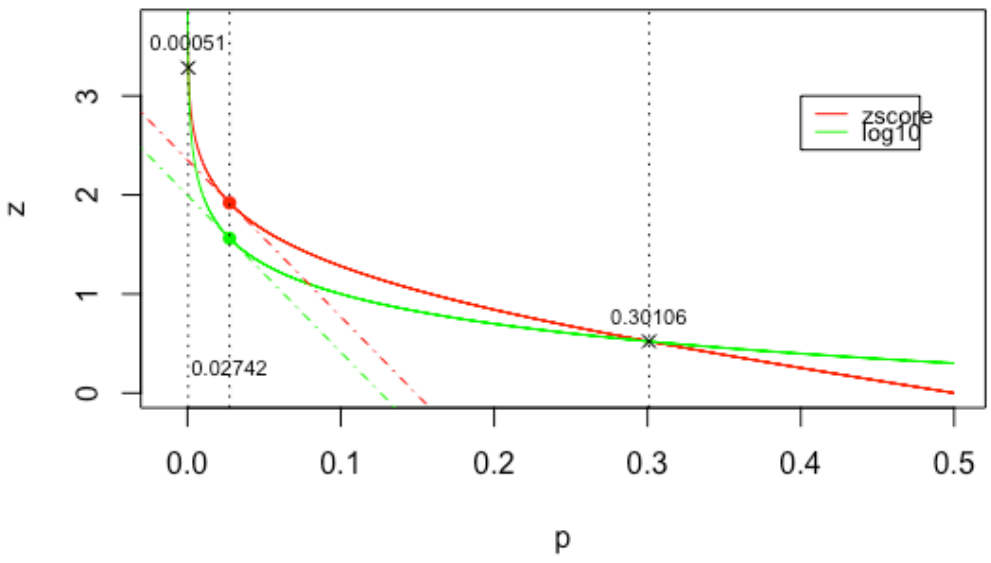


Figure 7. 4 logarithm and inverse normal transformation

Comparison between log scale transformation(green) and inverse normal transformation(red).

7.4. More phenotype connection supported by gene ontology terms

Motivation

Figure 3.2 gives the phenotype hierarchical clustering results and we investigate the gene ontology terms supporting subtrees. However, not all phenotype connections are captured by the hierarchical clustering algorithm, especially local similarity defined by a subset of genes (for example GO terms). For example, connection between schizophrenia and height has been reported in epidemiological study but they are not placed adjacent to each other in hierarchical clustering. A further investigation of this phenotype pair reveals GO terms supporting this connection. This proves that phenotype pairs not adjacent to each other in the hierarchical clustering result can have local similarities supported by GO terms.

Current discovery

The first three Go ontology terms supporting schizophrenia and height are positive regulation of myoblast proliferation (zscore 19.7) magnesium ion homeostasis (zscore 13.3) and negative regulation of megakaryocyte differentiation (zscore 13.2). Reduced level of magnesium level is observed in schizophrenia patients[90], thus directly link to schizophrenia. Meanwhile, magnesium ion influences height through muscle tissue that magnesium bind competitively to calcium ion binding sites and thus participates in the regulation of muscle contraction[91]. Similarly, literatures support megakaryocyte's influence on both schizophrenia and height that megakaryocyte controls the biogenesis of platelet[92] and altered platelet physiology has been observed in schizophrenia[93]. Megakaryocyte is associated with height by serving as a regulator of bone marrow homeostasis and physiology[94].

In order to detect more phenotype connections, I find the three most similar phenotypes for each phenotype by comparing their gene association profile. The three most similar phenotypes to height are as pubertal growth late adolescence, osteoporosis and fasting glucose. The top GO terms supporting each phenotype pair are consistent with phenotypes that the first GO term connecting height with pubertal growth late adolescence is chondroblast differentiation which is, by definition, related to bone physiology. The first GO term connecting height with osteoporosis is glucose-6-phosphate transport, linking osteoporosis to the third most similar phenotype fasting glucose. The top first GO term linking height and fasting glucose, insulin receptor substrate binding, gives more mechanistic evidence on a finer granularity.

In seek of more interesting GO terms linking seemingly unrelated phenotypes, I find one GO term, positive regulation of telomere maintenance via telomere lengthening, linking Alzheimer's disease, bipolar disorder, glioma and longevity as shown in Figure 7.1.

Table 7. 1 telomere maintenance connecting 5 phenotypes

(A)

	('Alzheimer[pubmed :21460841]', 'Alzheimer[pubmed: 25188341]')	('Alzheimer[pubmed :21460841]', 'Longevity[pubmed: 24688116]')	('Alzheimer[pubmed :21460841]', 'Glioma')	('Alzheimer[pubmed :21460841]', 'Bipolar Disorder')	('Alzheimer[pubmed :25188341]', 'Longevity[pubmed: 24688116]')	('Alzheimer[pubmed :25188341]', 'Glioma')	('Alzheimer[pubmed :25188341]', 'Bipolar Disorder')	('Longevity[pubmed: 24688116]', 'Glioma')	('Longevity[pubmed: 24688116]', 'Bipolar Disorder')	('Glioma', 'Bipolar Disorder')
1th GO term	very-low-density lipoprotein particle clearance(10.7)	receptor-mediated virion attachment to host cell(14.9)	hormone receptor binding(8.5)	rna transport(10.8)	receptor-mediated virion attachment to host cell(13.5)	neurotransmitter transport(10.0)	regulation of brown fat cell differentiation(9.2)	integral component of postsynaptic specialization membrane(9.3)	dna protection(6.8)	positive regulation of telomere maintenance via telomere lengthening(12.7)
2th GO term	cdc73/paf1 complex(10.1)	positive regulation of histone h4 acetylation(14.2)	positive regulation of telomere maintenance via telomere lengthening(5.6)	nerve growth factor binding(9.3)	positive regulation of sensory perception of pain(10.3)	creatine kinase activity(7.6)	choline transmembrane transporter activity(5.5)	negative regulation of epidermal growth factor-activated receptor activity(8.0)	cardiac muscle cell development(5.2)	regulation of cellular response to stress(8.4)
3th GO term	positive regulation of telomere maintenance via telomere lengthening(9.5)	epsilon dna polymerase complex(10.6)	transmembrane receptor protein tyrosine phosphatase activity(5.6)	positive regulation of telomere maintenance via telomere lengthening(7.4)	dna repair complex(9.0)	replication fork protection complex(6.8)	t cell chemotaxis(4.9)	anterior commissure morphogenesis(6.5)	glycerolipid metabolic process(5.2)	spectrin(6.4)
4th GO term	dsrna transport(7.1)	torc1 complex(6.5)	bitter taste receptor activity(5.4)	diacylglycerol kinase activity(6.7)	positive regulation of telomere maintenance via telomere lengthening(8.9)	sphingomyelin metabolic process(6.6)	trna splicing, via endonucleolytic cleavage and ligation(4.9)	macrophage activation involved in immune response(6.4)	xenobiotic transmembrane transporting atpase activity(4.8)	positive regulation of transcription initiation from rna polymerase ii promoter(6.2)
5th GO term	cerebellar purkinje cell differentiation(7.1)	metanephric collecting duct development(6.5)	maintenance of cell polarity(5.2)	fucose binding(6.5)	carbohydrate:proton symporter activity(8.8)	susceptibility to natural killer cell mediated cytotoxicity(6.3)	rna transport(4.7)	negative regulation of coagulation(6.3)	positive regulation of granulocyte differentiation(4.6)	regulation of protein export from nucleus(6.0)

(B) Telomere-phenotype literatures:

Alzheimer	[PMID: 27312549] Telomere Shortening in Alzheimer's Disease Patients
Longevity	[PMID: 24350925] Telomeres and their role in aging and longevity
Glioma	[PMID: 26014050] Telomere maintenance and the etiology of adult glioma
Bipolar Disorder	[PMID: 28621334] Telomere Length and Bipolar Disorder

Telomeres are DNA-protein structures that form protective caps at the end of eukaryotic chromosomes. Accelerated rate of telomere shortening has been observed in Alzheimer's disease patients[95]. Multiple evidences has linked telomere biology to aging and mechanism leading to longevity[96]. Telomere maintenance has been identified as an important role in glioma susceptibility, initiation, and prognosis[97] Shortened telomere length is associated with familial risk for bipolar disorder[98].

Method in the pilot study

Computing the three most similar phenotypes for a chosen target phenotype is straightforward but there exist caveats. It is tempting to choose the three phenotypes from hierarchical clustering as three most similar ones. However, hierarchical clustering is history dependent, meaning the most similar phenotype to target can be joined earlier due to a smaller distance to a third phenotype. For this reason, the three most similar phenotypes to target phenotype need to be calculated separately.

On the contrary, finding GO term linking multiple phenotypes involves multiple steps. This cannot be easily completed due to the large number (more than 7000) of GO terms for each phenotype pair and the exponential number of phenotype combinations (it is the power set of all phenotypes which scale to 2^{69})

In order to find GO term linking multiple phenotypes in a feasible time scale, I divided the searching process in two steps. First search for any candidate GO terms that link target phenotype with the most other phenotypes. Second validate the candidate GO terms across phenotype combinations.

In the first step, I start seeking GO term candidate by choosing a target from the 69 phenotypes and a zscore threshold (set to zscore 4 by default). Then all GO terms associating the target phenotype with all other 68 phenotypes are sorted by number of phenotypes pairs passing the threshold. For example, the GO term connecting the most phenotypes with glioma passing zscore threshold is positive regulation of telomere maintenance via telomere lengthening. This GO terms connects glioma with 6 phentoypes (bipolar disorder, Crohn's disease, ever smoked, depressive, ulcerative colitis and Alzheimer's disease). The GO term connecting the second most phenotypes with glioma passing zscore threshold is flavonoid glucuronidation, associating glioma with 4 phenotypes (acute lymphoblastoid leukemia, cholesterol, systolic blood pressure and high density lipoprotein). This process is repeated for all phenotypes by choosing each as target. At the end of the first step, I have a few GO terms for each phenotype connecting it with at least two other phenotypes where the GO terms pass zscore threshold.

In the second step, I apply several criteria to the top few GO terms of each phenotype. The first criterion is the number of phenotypes GO term connecting with the target phenotype. Inspecting a GO term that links three phenotypes with target one makes it qualify to pass the first step but may not be very promising in linking multiple phenotype together. This criterion makes the search process a heuristic rather than algorithm but significantly accelerates the search process. The second criterion requires the GO terms to be specific than broad concept. This is necessary to ensure GO terms conveys specific physiological process connecting phenotypes. Otherwise, we may find a GO term linking multiple phenotypes but not convey information only because that GO term is protein binding and thus too broad to further our understanding. GO terms satisfying these two criteria is run against all phenotype combinations only the phenotype pairs passing the same zscore cutoff set in the first step is kept.

Points to improve

When looking for the three most similar phenotypes for a chosen target phenotype, Euclidean distance was used to measure phenotype similarity. Although intuitive, Euclidean distance is sensitive to noise accumulated through the dimensions. In our case, weak dissimilarity across the twenty thousand genes can contribute to the similarity metric and bias the results. To resolve this problem, other distance metric can be attempted which are more robust to dimensionality difference.

One particular solution for the above mentioned defect can be pearson's correlation coefficient which is confined between -1 and 1. Shifting this domain to 0 to 2 gives a well-defined distance range and I expect this metric to be robust against small noised from high dimension.

When looking for GO terms connecting multiple phenotypes together, I applied two criteria to reduce the number of candidate GO terms. This process can be biased to my knowledge base and susceptible to falsely exclude interesting connection. A parameter-free approach is expected to remove this bias and help the search to capture more phenotype connections.

One particular solution for the above mentioned defect can be a background distribution of GO term zscores. Given first few GO terms from each target phenotype that connecting it with at least two other phenotypes, the GO terms' zscore can be collected and used to build a global background distribution across all phenotypes. The significant few out of this background zscore distribution can be used to choose candidate GO terms for further inspections in a parameter free fashion.

References

1. Council, N.R., *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. 2011, Washington, DC: The National Academies Press. 142.
2. Loscalzo, J., I. Kohane, and A.L. Barabasi, *Human disease classification in the postgenomic era: a complex systems approach to human pathobiology.* Mol Syst Biol, 2007. **3**: p. 124.
3. Seidman, J.G. and C. Seidman, *The genetic basis for cardiomyopathy: from mutation identification to mechanistic paradigms.* Cell, 2001. **104**(4): p. 557-67.
4. Collins, F.S. and H. Varmus, *A new initiative on precision medicine.* N Engl J Med, 2015. **372**(9): p. 793-5.
5. Paez, J.G., et al., *EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy.* Science, 2004. **304**(5676): p. 1497-500.
6. Pao, W., et al., *Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain.* PLoS Med, 2005. **2**(3): p. e73.
7. Hall, P., et al., *Hormone-replacement therapy influences gene expression profiles and is associated with breast-cancer prognosis: a cohort study.* BMC Med, 2006. **4**: p. 16.
8. Hedenfalk, I., et al., *Gene-expression profiles in hereditary breast cancer.* N Engl J Med, 2001. **344**(8): p. 539-48.
9. Visscher, P.M., et al., *10 Years of GWAS Discovery: Biology, Function, and Translation.* Am J Hum Genet, 2017. **101**(1): p. 5-22.

10. Sud, A., B. Kinnersley, and R.S. Houlston, *Genome-wide association studies of cancer: current insights and future perspectives.* Nat Rev Cancer, 2017. **17**(11): p. 692-704.

11. Gamazon, E.R., et al., *A gene-based association method for mapping traits using reference transcriptome data.* Nat Genet, 2015. **47**(9): p. 1091-8.

12. Liu, J.Z., et al., *A versatile gene-based test for genome-wide association studies.* Am J Hum Genet, 2010. **87**(1): p. 139-45.

13. He, X., et al., *Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS.* Am J Hum Genet, 2013. **92**(5): p. 667-80.

14. Gusev, A., et al., *Integrative approaches for large-scale transcriptome-wide association studies.* Nat Genet, 2016. **48**(3): p. 245-52.

15. Zhu, Z., et al., *Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets.* Nat Genet, 2016. **48**(5): p. 481-7.

16. Sivakumaran, S., et al., *Abundant pleiotropy in human complex diseases and traits.* Am J Hum Genet, 2011. **89**(5): p. 607-18.

17. Stearns, F.W., *One hundred years of pleiotropy: a retrospective.* Genetics, 2010. **186**(3): p. 767-73.

18. Paaby, A.B. and M.V. Rockman, *The many faces of pleiotropy.* Trends Genet, 2013. **29**(2): p. 66-73.

19. Solovieff, N., et al., *Pleiotropy in complex traits: challenges and strategies.* Nat Rev Genet, 2013. **14**(7): p. 483-95.

20. Zammit, S., et al., *Height and body mass index in young adulthood and risk of schizophrenia: a longitudinal study of 1 347 520 Swedish men.* Acta Psychiatr Scand, 2007. **116**(5): p. 378-85.

21. Bacanu, S.A., X. Chen, and K.S. Kendler, *The genetic overlap between schizophrenia and height.* Schizophr Res, 2013. **151**(1-3): p. 226-8.

22. Pickrell, J.K., et al., *Detection and interpretation of shared genetic influences on 42 human traits.* Nat Genet, 2016. **48**(7): p. 709-17.

23. Turley, P., et al., *Multi-trait analysis of genome-wide association summary statistics using MTAG.* Nat Genet, 2018. **50**(2): p. 229-237.

24. McCarroll, S.A., et al., *Comparing genomic expression patterns across species identifies shared transcriptional profile in aging.* Nat Genet, 2004. **36**(2): p. 197-204.

25. Long, T., et al., *Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites.* Nat Genet, 2017. **49**(4): p. 568-578.

26. Do, K.Q., et al., *gamma-Glutamylglutamine and taurine concentrations are decreased in the cerebrospinal fluid of drug-naive patients with schizophrenic disorders.* J Neurochem, 1995. **65**(6): p. 2652-62.

27. Kosicek, M., et al., *Elevated cerebrospinal fluid sphingomyelin levels in prodromal Alzheimer's disease.* Neurosci Lett, 2012. **516**(2): p. 302-5.

28. West, P.R., et al., *Metabolomics as a tool for discovery of biomarkers of autism spectrum disorder in the blood plasma of children.* PLoS One, 2014. **9**(11): p. e112445.

29. Ji, H., et al., *Arthritis critically dependent on innate immune system players.* Immunity, 2002. **16**(2): p. 157-68.

30. Eleftherohorinou, H., et al., *Pathway-driven gene stability selection of two rheumatoid arthritis GWAS identifies and validates new susceptibility genes in receptor mediated signalling pathways.* Hum Mol Genet, 2011. **20**(17): p. 3494-506.

31. Firestein, G.S., M. Yeo, and N.J. Zvaifler, *Apoptosis in rheumatoid arthritis synovium.* J Clin Invest, 1995. **96**(3): p. 1631-8.

32. Dixon, A.L., et al., *A genome-wide association study of global gene expression.* Nat Genet, 2007. **39**(10): p. 1202-7.

33. Duan, S., et al., *Genetic architecture of transcript-level variation in humans.* Am J Hum Genet, 2008. **82**(5): p. 1101-13.

34. Liang, L., et al., *A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines.* Genome Res, 2013. **23**(4): p. 716-26.

35. Grundberg, E., et al., *Mapping cis- and trans-regulatory effects across multiple tissues in twins.* Nat Genet, 2012. **44**(10): p. 1084-9.

36. Myers, A.J., et al., *A survey of genetic human cortical gene expression.* Nat Genet, 2007. **39**(12): p. 1494-9.

37. Schadt, E.E., et al., *Mapping the genetic architecture of gene expression in human liver.* PLoS Biol, 2008. **6**(5): p. e107.

38. Wright, F.A., et al., *Heritability and genomics of gene expression in peripheral blood.* Nat Genet, 2014. **46**(5): p. 430-7.

39. Zeller, T., et al., *Genetics and beyond--the transcriptome of human monocytes and disease susceptibility.* PLoS One, 2010. **5**(5): p. e10693.

40. *The Genotype-Tissue Expression (GTEx) project.* Nat Genet, 2013. **45**(6): p. 580-5.

41. Bulik-Sullivan, B., et al., *An atlas of genetic correlations across human diseases and traits.* Nat Genet, 2015. **47**(11): p. 1236-41.

42. Zheng, J., et al., *LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis.* Bioinformatics, 2017. **33**(2): p. 272-279.

43. O'Connor, L.J. and A.L. Price, *Distinguishing genetic correlation from causation across 52 diseases and complex traits.* Nat Genet, 2018. **50**(12): p. 1728-1734.

44. Lee, S.H., et al., *Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs.* Nat Genet, 2013. **45**(9): p. 984-94.

45. Bralic, I., et al., *Association of early menarche age and overweight/obesity.* J Pediatr Endocrinol Metab, 2012. **25**(1-2): p. 57-62.

46. Bruning, P.F., et al., *Insulin resistance and breast-cancer risk.* Int J Cancer, 1992. **52**(4): p. 511-6.

47. Glatt, S.L., et al., *Risk factors for dementia in Parkinson's disease: effect of education.* Neuroepidemiology, 1996. **15**(1): p. 20-5.

48. Butler, R.W., N.-K.V. Cheung, and J.H. Eddy, *Increased intellectual functioning in children with Neuroblastoma.* Child Neuropsychology, 1996. **2**(2): p. 77-82.

49. Langley-Evans, S.C. and D.V. Sculley, *The association between birthweight and longevity in the rat is complex and modulated by maternal protein intake during fetal life.* FEBS Lett, 2006. **580**(17): p. 4150-3.

50. Kingsley, D.M., *The TGF-beta superfamily: new members, new receptors, and new genetic tests of function in different organisms.* Genes Dev, 1994. **8**(2): p. 133-46.

51. Katz, M., I. Amit, and Y. Yarden, *Regulation of MAPKs by growth factors and receptor tyrosine kinases.* Biochim Biophys Acta, 2007. **1773**(8): p. 1161-76.

52. Eckert, D., et al., *The AP-2 family of transcription factors.* Genome Biol, 2005. **6**(13): p. 246.

53. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.* Nucleic Acids Res, 2016. **44**(D1): p. D733-45.

54. Fajardo, A.M., G.A. Piazza, and H.N. Tinsley, *The role of cyclic nucleotide signaling pathways in cancer: targets for prevention and treatment.* Cancers (Basel), 2014. **6**(1): p. 436-58.

55. Azevedo, M.F., et al., *Clinical and molecular genetics of the phosphodiesterases (PDEs).* Endocr Rev, 2014. **35**(2): p. 195-233.

56. Seino, S., et al., *Roles of cAMP signalling in insulin granule exocytosis.* Diabetes Obes Metab, 2009. **11 Suppl 4**: p. 180-8.

57. Roy, K., et al., *NADPH oxidases and cancer.* Clin Sci (Lond), 2015. **128**(12): p. 863-75.

58. Reed, J.C., *Dysregulation of apoptosis in cancer.* J Clin Oncol, 1999. **17**(9): p. 2941-53.

59. Tomita, T., *Apoptosis in pancreatic beta-islet cells in Type 2 diabetes.* Bosn J Basic Med Sci, 2016. **16**(3): p. 162-79.

60. Leznicki, P., et al., *The association of BAG6 with SGTA and tail-anchored proteins.* PLoS One, 2013. **8**(3): p. e59590.

61. Zhu, T., et al., *Expression and prognostic role of SGTA in human breast carcinoma correlates with tumor cell proliferation.* J Mol Histol, 2014. **45**(6): p. 665-77.

62. Bergmann, S., J. Ihmels, and N. Barkai, *Iterative signature algorithm for the analysis of large-scale gene expression data.* Phys Rev E Stat Nonlin Soft Matter Phys, 2003. **67**(3 Pt 1): p. 031902.

63. Pirscoveanu, D.F.V., et al., *Tau protein in neurodegenerative diseases - a review.* Rom J Morphol Embryol, 2017. **58**(4): p. 1141-1150.

64. Seto-Salvia, N., et al., *Dementia risk in Parkinson disease: disentangling the role of MAPT haplotypes.* Arch Neurol, 2011. **68**(3): p. 359-64.

65. Kadak, M.T., et al., *Low Serum Level alpha-Synuclein and Tau Protein in Autism Spectrum Disorder Compared to Controls.* Neuropediatrics, 2015. **46**(6): p. 410-5.

66. Nation, D.A., et al., *Pulse pressure in relation to tau-mediated neurodegeneration, cerebral amyloidosis, and progression to dementia in very old adults.* JAMA Neurol, 2015. **72**(5): p. 546-53.

67. Pan, Y., et al., *Knockdown of LncRNA MAPT-AS1 inhibites proliferation and migration and sensitizes cancer cells to paclitaxel by regulating MAPT expression in ER-negative breast cancers.* Cell Biosci, 2018. **8**: p. 7.

68. Ikeda, H., et al., *The estrogen receptor influences microtubule-associated protein tau (MAPT) expression and the selective estrogen receptor inhibitor fulvestrant downregulates MAPT and increases the sensitivity to taxane in breast cancer cells.* Breast Cancer Res, 2010. **12**(3): p. R43.

69. Edwards, T.L., et al., *Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease.* Ann Hum Genet, 2010. **74**(2): p. 97-109.

70. Glodzik, L., et al., *Blood pressure decrease correlates with tau pathology and memory decline in hypertensive elderly.* Neurobiol Aging, 2014. **35**(1): p. 64-71.

71. Medina-Aguilar, R., et al., *DNA methylation data for identification of epigenetic targets of resveratrol in triple negative breast cancer cells.* Data Brief, 2017. **11**: p. 169-182.

72. Liu, G., et al., *Identifying the Association Between Alzheimer's Disease and Parkinson's Disease Using Genome-Wide Association Studies and Protein-Protein Interaction Network.* Mol Neurobiol, 2015. **52**(3): p. 1629-1636.

73. Nalls, M.A., et al., *Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies.* Lancet, 2011. **377**(9766): p. 641-9.

74. Van Wesenbeeck, L., et al., *Involvement of PLEKHM1 in osteoclastic vesicular transport and osteopetrosis in incisors absent rats and humans.* J Clin Invest, 2007. **117**(4): p. 919-30.

75. Hogg, E., et al., *High Prevalence of Undiagnosed Insulin Resistance in Non-Diabetic Subjects with Parkinson's Disease.* J Parkinsons Dis, 2018. **8**(2): p. 259-265.

76. Stern, M., *Insulin signaling and autism.* Front Endocrinol (Lausanne), 2011. **2**: p. 54.

77. Salvetti, A., et al., *The inter-relationship between insulin resistance and hypertension.* Drugs, 1993. **46 Suppl 2**: p. 149-59.

78. Klein, G.L., *Insulin and bone: Recent developments.* World J Diabetes, 2014. **5**(1): p. 14-6.

79. Genuth, S. and H.E. Lebovitz, *STIMULATION OF INSULIN RELEASE BY CORICOTROPIN.* Endocrinology, 1965. **76**: p. 1093-9.

80. Gunter, M.J., et al., *Insulin, insulin-like growth factor-I, and risk of breast cancer in postmenopausal women.* J Natl Cancer Inst, 2009. **101**(1): p. 48-60.

81. King, M.J., et al., *Membrane-associated N-myristoyltransferase activity is reduced in obese (fa/fa) Zucker rat liver.* Biochem Biophys Res Commun, 1993. **196**(2): p. 665-70.

82. Goginashvili, A., et al., *Insulin granules. Insulin secretory granules control autophagy in pancreatic beta cells.* Science, 2015. **347**(6224): p. 878-82.

83. Ward, J.H., *Hierarchical Grouping to Optimize an Objective Function.* Journal of the American Statistical Association, 1963. **58**(301): p. 236-244.

84. Csardi, G., Z. Kutalik, and S. Bergmann, *Modular analysis of gene expression data with R.* Bioinformatics, 2010. **26**(10): p. 1376-7.

85. Ihmels, J., et al., *Revealing modular organization in the yeast transcriptional network.* Nat Genet, 2002. **31**(4): p. 370-7.

86. Ihmels, J., S. Bergmann, and N. Barkai, *Defining transcription modules using large-scale gene expression data.* Bioinformatics, 2004. **20**(13): p. 1993-2003.

87. Leff, A. and J.T. Rayfield. *Web-application development using the Model/View/Controller design pattern*. in *Proceedings Fifth IEEE International Enterprise Distributed Object Computing Conference*. 2001.

88. Bostock, M., V. Ogievetsky, and J. Heer, *D³ Data-Driven Documents.* IEEE Transactions on Visualization and Computer Graphics, 2011. **17**(12): p. 2301-2309.

89. Kundaje, A., et al., *Integrative analysis of 111 reference human epigenomes.* Nature, 2015. **518**(7539): p. 317-30.

90. Kirov, G.K. and K.N. Tsachev, *Magnesium, schizophrenia and manic-depressive disease.* Neuropsychobiology, 1990. **23**(2): p. 79-81.

91. Potter, J.D., S.P. Robertson, and J.D. Johnson, *Magnesium and the regulation of muscle contraction.* Fed Proc, 1981. **40**(12): p. 2653-6.

92. Patel, S.R., J.H. Hartwig, and J.E. Italiano, Jr., *The biogenesis of platelets from megakaryocyte proplatelets.* J Clin Invest, 2005. **115**(12): p. 3348-54.

93. Asor, E. and D. Ben-Shachar, *Platelets: A possible glance into brain biological processes in schizophrenia.* World J Psychiatry, 2012. **2**(6): p. 124-33.

94. Malara, A., et al., *The secret life of a megakaryocyte: emerging roles in bone marrow homeostasis control.* Cell Mol Life Sci, 2015. **72**(8): p. 1517-36.

95. Liu, M., et al., *Telomere Shortening in Alzheimer's Disease Patients.* Ann Clin Lab Sci, 2016. **46**(3): p. 260-5.

96. Tzanetakou, I.P., et al., *Telomeres and their role in aging and longevity.* Curr Vasc Pharmacol, 2014. **12**(5): p. 726-34.

97. Walsh, K.M., et al., *Telomere maintenance and the etiology of adult glioma.* Neuro Oncol, 2015. **17**(11): p. 1445-52.

98. Powell, T.R., et al., *Telomere Length and Bipolar Disorder.* Neuropsychopharmacology, 2018. **43**(2): p. 445-453.

Appendix

Metabolite-phenotype associations with Q-value threshold 0.333

phenotype	metabolite	p-value	Q-value
ACUTE MYELOID LEUKEMIA	nicotinamide	0.00145	0.278
ACUTE MYELOID LEUKEMIA	1-arachidoyl-GPC(20:0)	0.000417	0.2
ACUTE MYELOID LEUKEMIA	mannose	0.000692	0.2
AGE RELATED MACULAR	1-(1-enyl-palmitoyl)-2-palmitoyl-GPC(P-16:0/16:0)	0.000148	0.091
AGE RELATED MACULAR	2-arachidonoyl-GPE(20:4)	0.000658	0.141
AGE RELATED MACULAR	1-stearoyl-2-linoleoyl-GPC(18:0/18:2)	0.000689	0.141
AGREEABLENESS	N-acetylmethionine	0.000717	0.214
AGREEABLENESS	3-methylglutarylcarnitine(1)	0.000696	0.214
ALZHEIMER	X-12688	0.000355	0.137
ALZHEIMER	O-sulfo-L-tyrosine	0.000542	0.297
ALZHEIMER	sphingomyelin(d18:2/14:0_d18:1/14:1)	0.00219	0.281
ALZHEIMER	X-17167	0.00299	0.3
ALZHEIMER	N-acetyl-1-methylhistidine	0.00109	0.253
ALZHEIMER	X-24065	0.000661	0.137
ALZHEIMER	N6-carbamoylthreonyladenosine	0.00249	0.329
ALZHEIMER	trans-urocanate	0.00146	0.3
ALZHEIMER	alpha-glutamyltyrosine	0.00198	0.329
ALZHEIMER	2-hydroxyglutarate	0.0015	0.281
ALZHEIMER	palmitoyldihydrosphingomyelin(d18:0/16:0)	7.38E-05	0.0445
ALZHEIMER	1-palmitoyl-2-linoleoyl-GPC(16:0/18:2)	0.000963	0.297
ALZHEIMER	X-21628	0.000953	0.137
ALZHEIMER	X-11261	0.00113	0.253
ALZHEIMER	sphingomyelin(d18:1/15:0_d16:1/17:0)	0.00216	0.322
ALZHEIMER	behenoylsphingomyelin(d18:1/22:0)	0.00142	0.285
ALZHEIMER	N-acetylmethionine	0.00277	0.281
ALZHEIMER	3-methylglutarylcarnitine(2)	0.002	0.308
ALZHEIMER	glycocholenatesulfate	0.00259	0.278

phenotype	metabolite	p-value	Q-value
ALZHEIMER	X-24278	0.00268	0.281
ALZHEIMER	4-androsten-3beta_17beta-diolmonosulfate(2)	0.00327	0.281
ALZHEIMER	sphingomyelin(d18:1/20:0_d16:1/22:0)	0.000812	0.273
ALZHEIMER	1-arachidonoyl-GPC(20:4n6)	0.00205	0.294
ALZHEIMER	oleoyl-linoleoyl-glycerol(18:1/18:2)[2]	0.00307	0.329
ALZHEIMER	palmitoyl-linoleoyl-glycerol(16:0/18:2)[2]	0.00152	0.281
ALZHEIMER	tryptophanbetaine	0.00134	0.329
ASTHMA	sphingomyelin(d18:1/14:0_d16:1/16:0)	0.00206	0.319
ASTHMA	gulonicacid	0.00125	0.319
ASTHMA	2-docosahexaenoyl-GPE(22:6)	0.00102	0.319
ASTHMA	1-docosapentaenoyl-GPC(22:5n3)	0.00213	0.319
ATTENTION DEFICIT	2-hydroxybutyrate/2-hydroxyisobutyrate	0.00084	0.293
ATTENTION DEFICIT	pyroglutamylglycine	0.00141	0.293
ATTENTION DEFICIT	gamma-glutamylglutamine	0.00114	0.293
ATTENTION DEFICIT	dodecanedioate	0.00207	0.323
AUTISM	1-adrenoyl-GPC(22:4)	0.00712	0.309
AUTISM	X-24061	0.00628	0.309
AUTISM	1_2_3-benzenetriolsulfate(2)	0.000532	0.158
AUTISM	X-21892	0.00195	0.271
AUTISM	X-16071	0.00171	0.292
AUTISM	X-17189	0.000353	0.227
AUTISM	stearoylcarnitine	0.00151	0.249
AUTISM	1-stearyl-GPC(O-18:0)	0.001	0.297
AUTISM	cholate	0.0012	0.318
AUTISM	1-palmitoyl-2-meadoyl-GPC(16:0/20:3n9)	0.00355	0.286
AUTISM	X-22147	0.00237	0.305
AUTISM	X-12206	0.0029	0.292
AUTISM	theophylline	0.000235	0.117
AUTISM	2-palmitoyl-GPC(16:0)	0.00097	0.312
AUTISM	trigonelline(N'-methylnicotinate)	0.00133	0.291
AUTISM	5alpha-androstan-3beta_17beta-dioldisulfate	0.000694	0.0894
AUTISM	1-stearoyl-GPE(18:0)	0.00286	0.307
AUTISM	alpha-hydroxyisocaproate	0.000211	0.136
AUTISM	1-methylhistidine	0.000363	0.158
AUTISM	gamma-glutamylhistidine	0.00292	0.235
AUTISM	X-19141	0.00106	0.3
AUTISM	dehydroisoandrosteronesulfate(DHEA-S)	5.25E-05	0.0169
AUTISM	N-acetylputrescine	0.00159	0.302
AUTISM	pregnsteroidmonosulfate	0.00262	0.315
AUTISM	tiglylcarnitine	0.00277	0.261
AUTISM	margarate(17:0)	0.00271	0.305
AUTISM	17-methylstearate	0.00154	0.188
AUTISM	X-11850	7.55E-05	0.0427
AUTISM	paraxanthine	0.000631	0.264
AUTISM	oleoylcarnitine	0.000502	0.151
AUTISM	caproate(6:0)	0.00196	0.307
AUTISM	dihydroorotate	0.00248	0.228
AUTISM	methylglucopyranoside(alpha+beta)	0.0019	0.332
AUTISM	X-11308	1.80E-05	0.0104
AUTISM	proline	0.00104	0.318
AUTISM	5alpha-pregnan-3beta_20alpha-diolmonosulfate(2)	0.00134	0.258
AUTISM	2-palmitoleoyl-GPC(16:1)	0.00119	0.127
AUTISM	citrate	0.0015	0.286
AUTISM	1-stearoyl-2-arachidonoyl-GPI(18:0/20:4)	0.0011	0.249
AUTISM	arachidate(20:0)	0.00176	0.209
AUTISM	N1-methyladenosine	0.00296	0.272
AUTISM	X-24241	0.00671	0.309
AUTISM	androsteroidmonosulfate(1)	0.000454	0.073

phenotype	metabolite	p-value	Q-value
AUTISM	X-11540	0.00226	0.272
AUTISM	creatinine	0.000377	0.227
AUTISM	4-acetylphenolsulfate	0.000661	0.195
AUTISM	X-11905	0.00201	0.292
AUTISM	pseudouridine	0.000416	0.261
AUTISM	taurodeoxycholate	0.000187	0.0602
AUTISM	X-15728	0.00181	0.286
AUTISM	X-24435	0.00159	0.209
AUTISM	kynurenine	0.00111	0.303
AUTISM	phosphatidylcholine(18:0/20:5_16:0/22:5n6)	0.00171	0.271
AUTISM	sphingomyelin(d18:1/20:0_d16:1/22:0)	0.000419	0.27
AUTISM	bilirubin(Z_Z)	0.00108	0.33
AUTISM	4-androsten-3beta_17beta-diolmonosulfate(1)	0.00111	0.251
AUTISM	X-23782	0.00176	0.249
AUTISM	androsteronesulfate	0.000226	0.145
AUTISM	X-11787	0.00153	0.305
AUTISM	dimethylglycine	0.000335	0.216
AUTISM	2-methylcitrate/homocitrate	0.00159	0.249
AUTISM	N-acetylglutamine	0.000352	0.0833
AUTISM	palmitoyl-linoleoyl-glycerol(16:0/18:2)[1]	0.000111	0.0672
AUTISM	succinylcarnitine	0.00112	0.307
AUTISM	1-palmitoyl-2-arachidonoyl-GPE(16:0/20:4)	0.00062	0.302
AUTISM	propionylglycine	0.00313	0.286
AUTISM	X-12822	0.0016	0.297
AUTISM	homocitrulline	1.47E-06	0.000932
AUTISM	X-14056	0.00164	0.242
AUTISM	DSGEGDFXAEGGGVR	0.00149	0.3
AUTISM	1-palmitoyl-2-palmitoleoyl-GPC(16:0/16:1)	0.00273	0.292
AUTISM	tryptophanbetaine	0.00225	0.307
AUTISM	1-linoleoyl-GPC(18:2)	0.00132	0.251
AUTISM	X-12007	0.0027	0.272
AUTISM	1-palmitoyl-2-linoleoyl-GPC(16:0/18:2)	5.11E-05	0.0169
AUTISM	epiandrosteronesulfate	0.000418	0.0833
AUTISM	5-hydroxyindoleacetate	0.00254	0.307
AUTISM	biliverdin	0.00242	0.312
AUTISM	gamma-glutamyl-epsilon-lysine	0.000681	0.271
AUTISM	beta-alanine	0.00107	0.31
AUTISM	X-15469	0.000691	0.33
AUTISM	alpha-glutamylglycine	9.38E-05	0.0604
AUTISM	1-methylurate	0.000897	0.332
AUTISM	2-isopropylmalate	0.00162	0.331
AUTISM	1-palmitoyl-2-docosahexaenoyl-GPE(16:0/22:6)	0.00159	0.233
AUTISM	C-glycosyltryptophan	0.000396	0.233
AUTISM	16a-hydroxyDHEA3-sulfate	0.000131	0.028
AUTISM	palmitoylcarnitine	0.001	0.331
AUTISM	eicosapentaenoate(EPA;20:5n3)	0.00339	0.223
AUTISM	3-aminoisobutyrate	0.00181	0.216
AUTISM	glycerol3-phosphate	0.00149	0.271
AUTISM	pregnenolonesulfate	7.04E-05	0.0454
AUTISM	imidazolelactate	0.00589	0.309
AUTISM	dihomo-linolenate(20:3n3orn6)	0.00153	0.318
AUTISM	1_2-dimyristoyl-GPC(14:0/14:0)	0.000456	0.294
AUTISM	X-23026	0.000254	0.148
AUTISM	N-acetyl-1-methylhistidine	0.00181	0.307
AUTISM	fumarate	0.0049	0.316
AUTISM	nonanoylcarnitine	0.0012	0.0984
AUTISM	21-hydroxypregnenolonedisulfate	0.000616	0.13
AUTISM	tryptophan	0.00183	0.242

phenotype	metabolite	p-value	Q-value
AUTISM	X-15503	5.51E-05	0.0355
AUTISM	behenoylsphingomyelin(d18:1/22:0)	0.000133	0.0695
AUTISM	X-21365	0.00207	0.261
AUTISM	X-01911	0.00435	0.248
AUTISM	N-acetylhistidine	0.00286	0.307
AUTISM	4-androsten-3alpha_17alpha-diolmonosulfate(3)	0.00274	0.176
AUTISM	stearoylsphingomyelin(d18:1/18:0)	0.00279	0.235
AUTISM	4-androsten-3beta_17beta-diolmonosulfate(2)	0.00115	0.204
AUTISM	xanthine	0.00211	0.296
AUTISM	10-undecenoate(11:1n1)	0.000843	0.255
AUTISM	X-23639	0.00107	0.303
BIPOLAR DISORDER	1-palmitoyl-2-meadoyl-GPC(16:0/20:3n9)	0.000406	0.23
BIPOLAR DISORDER	ergothioneine	8.01E-05	0.0484
BIPOLAR DISORDER AND	imidazolepropionate	0.00159	0.273
BIPOLAR DISORDER AND	X-11880	0.00221	0.284
BIPOLAR DISORDER AND	methioninesulfone	0.000244	0.149
BIPOLAR DISORDER AND	X-11441	4.34E-05	0.0223
BODY MASS INDEX	trigonelline(N'-methylnicotinate)	0.000235	0.11
BODY MASS INDEX	2-isopropylmalate	0.00115	0.171
BODY MASS INDEX	1-myristoylglycerol(14:0)	0.00135	0.171
BODY MASS INDEX	3-hydroxypyridinesulfate	0.00146	0.171
BREAST CANCER	sulfate	9.21E-05	0.0584
CHOLESTEROL	oleoyl-linoleoyl-glycerol(18:1/18:2)[1]	0.00139	0.14
CHOLESTEROL	2-linoleoyl-GPE(18:2)	0.000711	0.0799
CHOLESTEROL	X-24061	0.00242	0.18
CHOLESTEROL	X-24278	0.0002	0.0643
CHOLESTEROL	1-stearoyl-2-linoleoyl-GPI(18:0/18:2)	0.000507	0.0799
CHOLESTEROL	palmitoleoyl-oleoyl-glycerol(16:1/18:1)[2]	0.000745	0.0799
CHOLESTEROL	thyroxine	0.00294	0.21
CHOLESTEROL	1-palmitoleoyl-2-linoleoyl-GPC(16:1/18:2)	0.00503	0.27
CHOLESTEROL	2-arachidonoyl-GPE(20:4)	9.36E-05	0.0603
CHOLESTEROL	1-linolenoylglycerol(18:3)	0.000965	0.148
CHOLESTEROL	1-methylurate	0.00359	0.21
CHOLESTEROL	2-linoleoyl-GPC(18:2)	0.00466	0.245
CHOLESTEROL	1-dihomo-linolenoyl-GPE(20:3n3or6)	0.00107	0.14
CHOLESTEROL	1-palmitoyl-2-stearoyl-GPC(16:0/18:0)	0.00155	0.111
CHOLESTEROL	1-stearoyl-2-arachidonoyl-GPI(18:0/20:4)	0.000333	0.138
CHOLESTEROL	1-linoleoyl-GPE(18:2)	0.000175	0.0563
CHOLESTEROL	1-oleoylglycerol(18:1)	0.00127	0.148
CHOLESTEROL	X-24241	0.00439	0.257
CHOLESTEROL	1-palmitoyl-2-oleoyl-GPE(16:0/18:1)	0.00252	0.18
CHOLESTEROL	1-palmitoyl-2-arachidonoyl-GPI(16:0/20:4)	0.00305	0.196
CHOLESTEROL	1-palmitoyl-2-dihomo-linolenoyl-GPC(16:0/20:3n3or6)	0.00123	0.111
CHOLESTEROL	dihomo-linolenate(20:3n3orn6)	0.000417	0.0799
CHRONIC KIDNEY DISEASE	pyroglutamine	0.000398	0.0641
CHRONIC KIDNEY DISEASE	quinolinate	0.0022	0.165
CHRONIC KIDNEY DISEASE	5-hydroxytryptophan	0.00197	0.165
CHRONIC KIDNEY DISEASE	X-21821	0.00476	0.313
CHRONIC KIDNEY DISEASE	hypoxanthine	0.00157	0.165
CHRONIC KIDNEY DISEASE	cyclo(leu-pro)	0.00131	0.165
CHRONIC KIDNEY DISEASE	homocitrulline	4.26E-05	0.00915
CHRONIC KIDNEY DISEASE	gamma-glutamyl-epsilon-lysine	0.00447	0.313
CHRONIC KIDNEY DISEASE	X-11305	0.00114	0.175
CHRONIC KIDNEY DISEASE	2-aminooctanoate	0.000533	0.165
CHRONIC KIDNEY DISEASE	guanidinoacetate	1.89E-05	0.00609
CHRONIC KIDNEY DISEASE	lactate	0.00243	0.165
CHRONIC KIDNEY DISEASE	1-(1-enyl-stearoyl)-2-oleoyl-GPC(P-18:0/18:1)	0.0029	0.267
CHRONIC KIDNEY DISEASE	homoarginine	0.000135	0.041

phenotype	metabolite	p-value	Q-value
CHRONIC KIDNEY DISEASE	methioninesulfone	1.50E-05	0.00609
CHRONIC KIDNEY DISEASE	N-acetylhistidine	0.000741	0.165
CHRONIC KIDNEY DISEASE	xanthine	0.000178	0.041
CONSCIENTIOUSNESS	1-stearoyl-2-oleoyl-GPE(18:0/18:1)	0.000257	0.151
CONSCIENTIOUSNESS	margarate(17:0)	0.00135	0.266
CONSCIENTIOUSNESS	pyridoxate	0.00125	0.266
CORONARY ARTERY DISEASE	2-hydroxy-3-methylvalerate	0.00253	0.288
CORONARY ARTERY DISEASE	hypoxanthine	0.000273	0.161
CORONARY ARTERY DISEASE	X-12411	0.00251	0.288
CORONARY ARTERY DISEASE	5-hydroxytryptophan	0.00283	0.288
CORONARY ARTERY DISEASE	X-24125	0.000648	0.191
CORONARY ARTERY DISEASE	xanthine	0.00293	0.288
CROHN DISEASE	carnitine	0.00136	0.284
CROHN DISEASE	1-stearoyl-2-oleoyl-GPC(18:0/18:1)	0.00143	0.284
CROHN DISEASE	1-stearoyl-2-docosapentaenoyl-GPC(18:0/22:5n6)	0.000565	0.284
CROHN DISEASE	1-(1-enyl-palmitoyl)-2-arachidonoyl-GPC(P-16:0/20:4)	0.00253	0.302
CROHN DISEASE	isovalerylcarnitine	0.00289	0.229
CROHN DISEASE	myristoylcarnitine	0.00166	0.229
CROHN DISEASE	alpha-glutamyltyrosine	0.00245	0.229
CROHN DISEASE	1-stearoyl-2-linoleoyl-GPI(18:0/18:2)	0.00244	0.229
CROHN DISEASE	1-stearoyl-2-docosapentaenoyl-GPC(18:0/22:5n3)	0.000545	0.151
CROHN DISEASE	stearoylcarnitine	0.00431	0.298
CROHN DISEASE	tryptophanbetaine	0.000528	0.151
CROHN DISEASE	1-dihomo-linoleoyl-GPC(20:2)	0.0027	0.229
DIABETES	gamma-glutamylalanine	0.00064	0.218
DIABETES	1-myristoyl-GPC(14:0)	0.00124	0.218
DIABETES	hippurate	0.00155	0.218
DIABETES	pyroglutamylglutamine	0.00117	0.218
EARLY GROWTH	X-18901	0.00145	0.227
EARLY GROWTH	1-stearoyl-GPI(18:0)	0.00158	0.227
EARLY GROWTH	X-12063	0.00188	0.227
EARLY GROWTH	pseudouridine	0.000235	0.142
EARLY GROWTH	X-23765	0.00058	0.175
EARLY GROWTH	1-dihomo-linolenoyl-GPC(20:3n3or6)	0.00118	0.289
EARLY GROWTH	1-(1-enyl-palmitoyl)-2-linoleoyl-GPC(P-16:0/18:2)	0.00269	0.232
EARLY GROWTH	N-acetyltaurine	0.00172	0.289
EARLY GROWTH	1_2-dimyristoyl-GPC(14:0/14:0)	0.00102	0.289
EARLY GROWTH	X-22147	0.00259	0.232
EARLY GROWTH	oleoyl-linoleoyl-glycerol(18:1/18:2)[2]	0.0018	0.289
EXTREME_HEIGHT	X-11540	0.00332	0.306
EXTREME_HEIGHT	X-11381	0.00198	0.306
EXTREME_HEIGHT	carnitine	0.000708	0.306
EXTREME_HEIGHT	X-12450	0.00262	0.306
EXTREME_HEIGHT	hexanoylcarnitine	0.00297	0.306
EXTREME_HEIGHT	stearoylcarnitine	0.00146	0.306
EXTREME_HEIGHT	octadecanedioate	0.00293	0.306
FASTING GLUCOSE	1-palmitoyl-2-stearoyl-GPC(16:0/18:0)	0.000809	0.26
FASTING GLUCOSE	2-arachidonoyl-GPE(20:4)	0.000306	0.197
FASTING GLUCOSE	carnitine	0.000139	0.0757
GLIOMA	X-11787	0.000286	0.14
GLIOMA	X-23026	0.00135	0.282
GLIOMA	imidazolelactate	0.00278	0.282
GLIOMA	1-methylnicotinamide	0.0018	0.282
GLIOMA	pyroglutamine	0.0028	0.282
GLIOMA	DSGEGDFXAEGGGVR	0.00267	0.282
GLIOMA	X-12729	0.001	0.282
GLYCATED	lysine	0.000701	0.19
GLYCATED	1-arachidonoyl-GPI(20:4)	0.00179	0.324

phenotype	metabolite	p-value	Q-value
GLYCATED	X-17189	0.000222	0.12
HEIGHT	carnitine	0.000366	0.234
HIGH-DENSITY LIPOPROTEIN	1-linolenoylglycerol(18:3)	0.00113	0.161
HIGH-DENSITY LIPOPROTEIN	1-palmitoyl-GPE(16:0)	0.00288	0.161
HIGH-DENSITY LIPOPROTEIN	2-linoleoyl-GPC(18:2)	0.00479	0.315
HIGH-DENSITY LIPOPROTEIN	1-palmitoyl-2-oleoyl-GPE(16:0/18:1)	0.00273	0.161
HIGH-DENSITY LIPOPROTEIN	2-linoleoyl-GPE(18:2)	0.00144	0.178
HIGH-DENSITY LIPOPROTEIN	X-12688	0.0047	0.218
HIGH-DENSITY LIPOPROTEIN	1-stearoyl-2-arachidonoyl-GPI(18:0/20:4)	0.0021	0.161
HIGH-DENSITY LIPOPROTEIN	1-dihomo-linolenoyl-GPE(20:3n3or6)	0.00118	0.161
HIGH-DENSITY LIPOPROTEIN	1-stearoyl-2-linoleoyl-GPC(18:0/18:2)	0.00372	0.189
HIGH-DENSITY LIPOPROTEIN	X-24061	0.000629	0.178
HIGH-DENSITY LIPOPROTEIN	X-24278	0.0014	0.161
HIGH-DENSITY LIPOPROTEIN	2-arachidonoyl-GPE(20:4)	8.41E-05	0.0498
HIGH-DENSITY LIPOPROTEIN	1-linoleoyl-GPI(18:2)	0.006	0.257
HIGH-DENSITY LIPOPROTEIN	oleoyl-linoleoyl-glycerol(18:1/18:2)[1]	0.00219	0.161
HIGH-DENSITY LIPOPROTEIN	1-stearoyl-2-arachidonoyl-GPE(18:0/20:4)	0.00675	0.269
HIGH-DENSITY LIPOPROTEIN	xanthine	0.00815	0.303
HIGH-DENSITY LIPOPROTEIN	1-palmitoleoyl-2-linoleoyl-GPC(16:1/18:2)	0.00243	0.24
HIGH-DENSITY LIPOPROTEIN	1-palmitoyl-2-docosahexaenoyl-GPE(16:0/22:6)	0.00992	0.325
HIP	2-isopropylmalate	0.000614	0.296
HIP	carnitine	0.00151	0.316
HIP	N6-acetyllysine	0.000603	0.316
HIP	N-acetylputrescine	0.0011	0.296
HIP	N-palmitoylglycine	0.00265	0.332
HIP	1-methylxanthine	0.00132	0.316
INSULIN SECRETION	X-21341	0.00182	0.321
INSULIN SECRETION	2-hydroxyglutarate	0.000308	0.185
INSULIN SECRETION	pyroglutamylglutamine	0.00152	0.243
INSULIN SECRETION	ornithine	0.00122	0.321
INSULIN SECRETION	X-21666	0.000799	0.24
INSULIN SECRETION	methioninesulfone	0.00235	0.321
INSULIN SECRETION	X-23644	0.00224	0.321
INSULIN SECRETION	X-12026	0.00417	0.323
INSULIN SECRETION	X-23765	0.00403	0.323
INSULIN SECRETION	mannitol/sorbitol	0.00162	0.243
JUVENILE ARTHRITIS	myristoleoylcarnitine	0.00172	0.3
JUVENILE ARTHRITIS	X-22147	0.00251	0.3
JUVENILE ARTHRITIS	3-hydroxy-3-methylglutarate	0.00056	0.171
JUVENILE ARTHRITIS	leucylalanine	0.000344	0.171
JUVENILE ARTHRITIS	3-hydroxyoctanoate	0.00265	0.3
JUVENILE ARTHRITIS	creatinine	0.00295	0.3
LONGEVITY	urea	0.000987	0.212
LONGEVITY	gamma-glutamyltyrosine	0.00164	0.298
LONGEVITY	X-23314	0.000302	0.0973
LONGEVITY	palmitoleoyl-oleoyl-glycerol(16:1/18:1)[2]	0.00111	0.251
LONGEVITY	X-11478	0.000421	0.252
LONGEVITY	cortisol	0.00202	0.261
LONGEVITY	sphingomyelin(d18:1/20:1_d18:2/20:0)	0.0048	0.325
LONGEVITY	glycocholenatesulfate	0.00294	0.251
LONGEVITY	X-24309	8.03E-05	0.0517
LONGEVITY	xanthine	0.00144	0.298
LONGEVITY	X-24242	0.00177	0.251
LONGEVITY	X-11491	0.00311	0.251
LONGEVITY	carnitine	0.000185	0.102
LONGEVITY	5-hydroxylysine	0.000326	0.17
LOW-DENSITY LIPOPROTEIN	X-24278	0.000172	0.0535
LOW-DENSITY LIPOPROTEIN	1-stearoyl-2-arachidonoyl-GPI(18:0/20:4)	0.000523	0.112

phenotype	metabolite	p-value	Q-value
LOW-DENSITY LIPOPROTEIN	1-stearoyl-2-linoleoyl-GPI(18:0/18:2)	0.00137	0.198
LOW-DENSITY LIPOPROTEIN	oleoyl-linoleoyl-glycerol(18:1/18:2)[1]	0.00159	0.198
LOW-DENSITY LIPOPROTEIN	dihomo-linolenate(20:3n3orn6)	0.00328	0.245
LOW-DENSITY LIPOPROTEIN	1-(1-enyl-palmitoyl)-2-arachidonoyl-GPC(P-16:0/20:4)	0.00354	0.245
LOW-DENSITY LIPOPROTEIN	thyroxine	0.00324	0.245
LOW-DENSITY LIPOPROTEIN	1-oleoylglycerol(18:1)	0.000121	0.0535
LOW-DENSITY LIPOPROTEIN	1-palmitoyl-2-stearoyl-GPC(16:0/18:0)	0.000944	0.196
LUNG CANCER	X-21666	0.0003	0.125
LUNG CANCER	gluconate	0.000455	0.149
METFORMIN	1-stearoyl-2-docosapentaenoyl-GPC(18:0/22:5n6)	0.000237	0.153
NEUROBLASTOMA	p-cresolsulfate	0.000136	0.0849
NEUROTICISM	5-hydroxylysine	0.00102	0.291
NEUROTICISM	X-11372	0.00117	0.273
NEUROTICISM	2-aminooctanoate	0.000795	0.291
NEUROTICISM	glycine	0.00119	0.273
NEUROTICISM	1-(1-enyl-stearoyl)-2-docosahexaenoyl-GPC(P-18:0/22:6)	0.00238	0.302
NEUROTICISM	oleate/vaccenate(18:1)	0.00141	0.273
NEUROTICISM	epiandrosteronesulfate	0.0026	0.302
OSTEOPOROSIS	X-23782	0.00221	0.222
OSTEOPOROSIS	caproate(6:0)	0.0019	0.272
OSTEOPOROSIS	X-21666	0.0013	0.265
OSTEOPOROSIS	N-acetylaspartate(NAA)	0.00165	0.262
OSTEOPOROSIS	4-androsten-3beta_17beta-dioldisulfate(1)	0.00281	0.272
OSTEOPOROSIS	X-11470	0.000766	0.265
OSTEOPOROSIS	1-oleoyl-GPI(18:1)	0.00149	0.262
OSTEOPOROSIS	taurolithocholate3-sulfate	0.00391	0.266
OSTEOPOROSIS	1-palmitoleoyl-2-linoleoyl-GPC(16:1/18:2)	0.00229	0.272
OSTEOPOROSIS	ethylmalonate	0.00143	0.215
OSTEOPOROSIS	1-palmitoyl-2-dihomo-linolenoyl-GPC(16:0/20:3n3or6)	0.000772	0.262
OSTEOPOROSIS	uridine	0.000345	0.213
OSTEOPOROSIS	X-17327	0.00298	0.266
OSTEOPOROSIS	X-16570	0.00348	0.266
OSTEOPOROSIS	1-dihomo-linolenoyl-GPC(20:3n3or6)	0.00144	0.232
OSTEOPOROSIS	malonate	0.00112	0.265
OSTEOPOROSIS	tartarate	0.00108	0.232
OSTEOPOROSIS	androsteroidmonosulfate(1)	0.00191	0.298
OSTEOPOROSIS	N-acetyl-beta-alanine	0.00474	0.3
OSTEOPOROSIS	1-palmitoyl-2-linoleoyl-GPI(16:0/18:2)	0.000471	0.22
OSTEOPOROSIS	N-acetylalanine	0.0011	0.262
OSTEOPOROSIS	1-palmitoyl-2-oleoyl-GPE(16:0/18:1)	0.00317	0.272
OSTEOPOROSIS	cholate	0.00147	0.298
OSTEOPOROSIS	1-(1-enyl-palmitoyl)-2-arachidonoyl-GPC(P-16:0/20:4)	0.000876	0.248
OSTEOPOROSIS	X-16935	0.00205	0.266
OSTEOPOROSIS	guanidinoacetate	0.00251	0.272
OSTEOPOROSIS	4-guanidinobutanoate	0.00345	0.272
OSTEOPOROSIS	1-oleoyl-2-dihomo-linolenoyl-GPC(18:1/20:3)	0.000792	0.248
OSTEOPOROSIS	X-18899	0.0037	0.266
OSTEOPOROSIS	X-12822	0.000519	0.272
OSTEOPOROSIS	cholinephosphate	0.00198	0.272
OVERWEIGHT	1-stearoyl-GPE(18:0)	0.00026	0.159
PANCREATIC CANCER	X-17299	8.33E-05	0.0468
PANCREATIC CANCER	glycerate	0.000289	0.163
PROSTATE CANCER	malate	0.000324	0.206
PROSTATE CANCER	X-13529	0.00103	0.314
PROSTATE CANCER	X-23639	0.000621	0.314
PUBERTAL GROWTH LATE	pyroglutamine	0.00205	0.231
PUBERTAL GROWTH LATE	1-palmitoyl-2-dihomo-linolenoyl-GPC(16:0/20:3n3or6)	0.000558	0.298
PUBERTAL GROWTH LATE	X-24435	0.00284	0.262

phenotype	metabolite	p-value	Q-value
PUBERTAL GROWTH LATE	1-oleoyl-2-dihomo-linolenoyl-GPC(18:1/20:3)	0.00163	0.298
PUBERTAL GROWTH LATE	N-acetylglutamine	0.00281	0.308
PUBERTAL GROWTH LATE	quinolinate	0.000192	0.123
PUBERTAL GROWTH LATE	2-hydroxyphenylacetate	0.0024	0.308
PUBERTAL GROWTH LATE	5alpha-pregnan-3beta_20alpha-dioldisulfate	0.00126	0.231
PUBERTAL GROWTH LATE	cis-aconitate	0.00147	0.231
PUBERTAL GROWTH LATE	5-dodecenoate(12:1n7)	0.00215	0.231
PUBERTAL GROWTH LATE	gamma-glutamylleucine	0.0018	0.231
PUBERTAL GROWTH LATE	lactate	0.000352	0.208
PUBERTAL GROWTH LATE	1-dihomo-linolenoyl-GPC(20:3n3or6)	0.00163	0.298
RHEUMATOID ARTHRITIS	X-23997	0.000504	0.324
RHEUMATOID ARTHRITIS	isobutyrylcarnitine	0.000123	0.0681
SCHIZOPHRENIA	X-11372	0.000293	0.177
SCHIZOPHRENIA	methioninesulfone	2.15E-06	0.00139
SCHIZOPHRENIA	X-14838	0.000218	0.14
SCHIZOPHRENIA	3-methoxytyrosine	0.000972	0.313
SCHIZOPHRENIA	1_2_3-benzenetriolsulfate(2)	0.000579	0.142
SCHIZOPHRENIA	gamma-glutamylglutamate	3.97E-07	0.000246
SCHIZOPHRENIA	N-methylproline	0.00101	0.314
SMOKING AGE ONSET	epiandrosteronesulfate	0.000158	0.101
SMOKING AGE ONSET	3-ureidopropionate	0.000981	0.316
SMOKING CIGAR PER DAY	5-hydroxylysine	0.000385	0.166
SMOKING CIGAR PER DAY	X-22162	0.000581	0.166
SMOKING EVER SMOKE	myristoleate(14:1n5)	0.00032	0.182
SMOKING FORMER SMOKER	X-17179	0.000475	0.271
TRIGLYCERIDES	2-linoleoyl-GPC(18:2)	0.0111	0.326
TRIGLYCERIDES	dihomo-linolenate(20:3n3orn6)	0.0042	0.18
TRIGLYCERIDES	1-oleoylglycerol(18:1)	0.00166	0.119
TRIGLYCERIDES	1-palmitoyl-2-arachidonoyl-GPI(16:0/20:4)	0.000508	0.0979
TRIGLYCERIDES	1-stearoyl-2-arachidonoyl-GPI(18:0/20:4)	0.000193	0.0979
TRIGLYCERIDES	1-stearoyl-2-linoleoyl-GPI(18:0/18:2)	0.00113	0.116
TRIGLYCERIDES	2-linoleoyl-GPE(18:2)	0.00111	0.116
TRIGLYCERIDES	palmitoleoyl-oleoyl-glycerol(16:1/18:1)[2]	0.00299	0.138
TRIGLYCERIDES	1-palmitoyl-2-oleoyl-GPE(16:0/18:1)	0.00291	0.138
TRIGLYCERIDES	1-palmitoyl-2-stearoyl-GPC(16:0/18:0)	0.00921	0.298
TRIGLYCERIDES	1-palmitoleoyl-2-linoleoyl-GPC(16:1/18:2)	0.00479	0.193
TRIGLYCERIDES	X-24278	0.000608	0.0979
TRIGLYCERIDES	1-dihomo-linolenoyl-GPE(20:3n3or6)	0.00199	0.125
TRIGLYCERIDES	1-linoleoyl-GPI(18:2)	0.00973	0.298
TRIGLYCERIDES	oleoyl-linoleoyl-glycerol(18:1/18:2)[1]	0.00233	0.125
TRIGLYCERIDES	2-arachidonoyl-GPE(20:4)	0.000569	0.0979
TRIGLYCERIDES	1-linoleoyl-GPE(18:2)	0.00126	0.116
TRIGLYCERIDES	theobromine	0.00973	0.298
TRIGLYCERIDES	1-palmitoyl-2-dihomo-linolenoyl-GPC(16:0/20:3n3or6)	0.00215	0.125
TRIGLYCERIDES	1-palmitoyl-2-alpha-linolenoyl-GPC(16:0/18:3n3)	0.00564	0.214
TRIGLYCERIDES	1-linolenoylglycerol(18:3)	0.00144	0.116
TRIGLYCERIDES	X-24061	0.00876	0.298
ULCERATIVE COLITIS	1-arachidonoyl-GPE(20:4n6)	0.00109	0.33
ULCERATIVE COLITIS	cis-4-decenoylcarnitine	0.00273	0.33
ULCERATIVE COLITIS	X-12688	0.00145	0.33
ULCERATIVE COLITIS	1-palmitoyl-2-dihomo-linolenoyl-GPC(16:0/20:3n3or6)	0.0026	0.33
WAIST CIRCUMFERENCE	1-methylxanthine	0.000408	0.262
WAIST CIRCUMFERENCE ADJUSTED	betaine	0.00111	0.216
WAIST CIRCUMFERENCE ADJUSTED	proline	0.00109	0.278
WAIST CIRCUMFERENCE ADJUSTED	myristate(14:0)	0.000566	0.165
WAIST CIRCUMFERENCE ADJUSTED	1-palmitoleoyl-2-linoleoyl-GPC(16:1/18:2)	0.00013	0.0641
WAIST-HIP RATIO	kynurenate	0.0038	0.321
WAIST-HIP RATIO	X-18249	0.00125	0.321

phenotype	metabolite	p-value	Q-value
WAIST-HIP RATIO	beta-alanine	0.00317	0.321
WAIST-HIP RATIO	X-12729	0.00305	0.321
WAIST-HIP RATIO	X-11299	0.00136	0.33
WAIST-HIP RATIO	hexanoylcarnitine	0.00154	0.33
WAIST-HIP RATIO	1-nonadecanoyl-GPC(19:0)	0.00258	0.321
WAIST-HIP RATIO	X-15492	0.00283	0.322
WAIST-HIP RATIO	X-21411	0.00214	0.322
WAIST-HIP RATIO	sphingomyelin(d18:1/17:0_d17:1/18:0_d19:1/16:0)	0.00171	0.322
WAIST-HIP RATIO	N6-acetyllysine	0.00316	0.321
WAIST-HIP RATIO	trigonelline(N'-methylnicotinate)	0.000186	0.117
WAIST-HIP RATIO	palmitate(16:0)	0.00336	0.321
WAIST-HIP RATIO	1-palmitoyl-2-arachidonoyl-GPE(16:0/20:4)	0.00141	0.322
WAIST-HIP RATIO ADJUSTED FOR	X-18249	0.00243	0.174
WAIST-HIP RATIO ADJUSTED FOR	palmitoyl-linoleoyl-glycerol(16:0/18:2)[2]	0.00286	0.184
WAIST-HIP RATIO ADJUSTED FOR	N-acetyltaurine	0.00102	0.319
WAIST-HIP RATIO ADJUSTED FOR	1-palmitoleoyl-2-linoleoyl-GPC(16:1/18:2)	0.00176	0.325
WAIST-HIP RATIO ADJUSTED FOR	orotidine	0.0035	0.319
WAIST-HIP RATIO ADJUSTED FOR	X-22816	0.0032	0.319
WAIST-HIP RATIO ADJUSTED FOR	X-18914	0.00167	0.155
WAIST-HIP RATIO ADJUSTED FOR	palmitoleate(16:1n7)	0.00188	0.319
WAIST-HIP RATIO ADJUSTED FOR	1-stearoyl-2-linoleoyl-GPC(18:0/18:2)	0.000541	0.155
WAIST-HIP RATIO ADJUSTED FOR	sphingomyelin(d18:1/17:0_d17:1/18:0_d19:1/16:0)	0.00133	0.155
WAIST-HIP RATIO ADJUSTED FOR	X-11441	0.00119	0.155
WAIST-HIP RATIO ADJUSTED FOR	citrulline	0.00175	0.155
WAIST-HIP RATIO ADJUSTED FOR	N-acetylneuraminate	0.00402	0.321
WAIST-HIP RATIO ADJUSTED FOR	X-23788	0.000295	0.19
WAIST-HIP RATIO ADJUSTED FOR	orotate	0.00179	0.155
WAIST-HIP RATIO ADJUSTED FOR	hexanoylcarnitine	0.00145	0.325
WAIST-HIP RATIO ADJUSTED FOR	X-21286	0.00202	0.325
WAIST-HIP RATIO ADJUSTED FOR	acetylcarnitine	0.00109	0.319
WAIST-HIP RATIO ADJUSTED FOR	gluconate	0.00295	0.319
WAIST-HIP RATIO ADJUSTED FOR	1-nonadecanoyl-GPC(19:0)	0.00463	0.271
WAIST-HIP RATIO ADJUSTED FOR	1-palmitoyl-2-linoleoyl-GPC(16:0/18:2)	0.0015	0.155
WAIST-HIP RATIO ADJUSTED FOR	gamma-glutamylglutamate	0.00192	0.155
WEIGHT	trigonelline(N'-methylnicotinate)	9.26E-05	0.045
WEIGHT	2-isopropylmalate	0.00122	0.297

Gene phenotype associations with Q-value threshold 0.333

phenotype	gene	p-value	Q-value
2HR GLUCOSE	SHISA4	3.53E-05	0.0648
2HR GLUCOSE	ATF7IP2	0.00198	0.323
2HR GLUCOSE	FKRP	0.0016	0.323
2HR GLUCOSE	CACNG3	0.000476	0.221
2HR GLUCOSE	TECR	9.68E-05	0.101
2HR GLUCOSE	SYT4	4.38E-06	0.0175
2HR GLUCOSE	CCNY	0.00184	0.323
2HR GLUCOSE	SV2A	0.00147	0.323
2HR GLUCOSE	HIGD1B	0.0017	0.323
2HR GLUCOSE	ELMO1	0.000389	0.209
2HR GLUCOSE	MCEE	0.00102	0.323
2HR GLUCOSE	NAA50	0.00245	0.332
2HR GLUCOSE	PYGB	0.00207	0.327
2HR GLUCOSE	STMN4	0.000644	0.265
2HR GLUCOSE	PTDSS2	8.00E-05	0.101
2HR GLUCOSE	NEFL	0.00148	0.323
2HR GLUCOSE	CHRM5	0.000581	0.248
2HR GLUCOSE	TFPI	0.00184	0.323
2HR GLUCOSE	ARSK	0.0022	0.332
2HR GLUCOSE	RPL4	0.000127	0.119
2HR GLUCOSE	MAPK8IP1	0.000108	0.107
2HR GLUCOSE	SDF4	0.000343	0.198
2HR GLUCOSE	ITPK1	3.41E-05	0.0648
2HR GLUCOSE	PCBP2	0.00161	0.323
2HR GLUCOSE	IGSF21	0.00145	0.323
2HR GLUCOSE	PPP1CA	0.000144	0.119
2HR GLUCOSE	ISOC2	0.00151	0.323
2HR GLUCOSE	SV2B	0.00122	0.323
2HR GLUCOSE	TMEM132A	0.00209	0.328
2HR GLUCOSE	INA	0.000781	0.292
2HR GLUCOSE	MOAP1	0.00158	0.323
2HR GLUCOSE	RSL24D1	0.00236	0.332
2HR GLUCOSE	KLHL25	0.00196	0.323
2HR GLUCOSE	APLP1	0.00235	0.332
2HR GLUCOSE	HS3ST2	0.00156	0.323
2HR GLUCOSE	SYNDIG1	0.00151	0.323
2HR GLUCOSE	APEH	0.00122	0.323
2HR GLUCOSE	SSRP1	0.0014	0.323
2HR GLUCOSE	RXRB	0.00173	0.323
2HR GLUCOSE	OLAH	0.00179	0.323
2HR GLUCOSE	SERPINI2	0.0019	0.323
2HR GLUCOSE	CYLC2	0.00155	0.323
2HR GLUCOSE	PDZD4	0.00014	0.119
2HR GLUCOSE	PPP2R5B	0.000858	0.304
2HR GLUCOSE	SENP3	0.000418	0.209
2HR GLUCOSE	VARS2	7.35E-05	0.101
2HR GLUCOSE	PTPN23	9.70E-05	0.101
2HR GLUCOSE	PRDM2	0.000319	0.198
2HR GLUCOSE	LHX2	4.22E-05	0.0672
2HR GLUCOSE	NME3	0.00196	0.323
2HR GLUCOSE	RALYL	0.000217	0.173
2HR GLUCOSE	RPL12	0.00231	0.332
2HR GLUCOSE	CHAC1	0.0017	0.323
2HR GLUCOSE	LRRD1	0.00159	0.323
2HR GLUCOSE	MAP7D1	0.000144	0.119

phenotype	gene	p-value	Q-value
2HR GLUCOSE	NOP16	0.00105	0.323
2HR GLUCOSE	ZNF280B	0.00137	0.323
2HR GLUCOSE	TMEM251	0.00133	0.323
2HR GLUCOSE	TLR5	7.17E-05	0.101
2HR GLUCOSE	GABBR2	0.000865	0.304
2HR GLUCOSE	CCND3	0.000482	0.221
2HR GLUCOSE	BAG6	3.94E-05	0.0672
2HR GLUCOSE	MCHR1	0.00176	0.323
2HR GLUCOSE	KRT17	0.00129	0.323
2HR GLUCOSE	JPH4	0.00163	0.323
2HR GLUCOSE	FGF1	0.00166	0.323
2HR GLUCOSE	BLOC1S2	0.000333	0.198
2HR GLUCOSE	SNCB	0.00168	0.323
2HR GLUCOSE	ALKBH7	7.89E-07	0.00629
2HR GLUCOSE	PRPF8	9.48E-05	0.101
2HR GLUCOSE	AATF	0.00229	0.332
2HR GLUCOSE	TNFAIP8L3	0.00158	0.323
2HR GLUCOSE	MAN2C1	0.000768	0.292
2HR GLUCOSE	RASAL1	0.000521	0.231
2HR GLUCOSE	PRPS1L1	0.00118	0.323
2HR GLUCOSE	PPFIA3	0.000552	0.24
2HR GLUCOSE	CDC25B	0.00227	0.332
2HR GLUCOSE	C12orf57	0.000342	0.198
2HR GLUCOSE	INO80E	0.000225	0.174
2HR GLUCOSE	PRPF31	0.00196	0.323
2HR GLUCOSE	NRGN	0.0019	0.323
2HR GLUCOSE	DUS1L	0.000251	0.182
2HR GLUCOSE	RPS6KA4	0.00247	0.332
2HR GLUCOSE	SLC25A43	0.00166	0.323
2HR GLUCOSE	TAGLN3	0.000427	0.209
2HR GLUCOSE	NUDC	2.91E-06	0.0139
2HR GLUCOSE	VAC14	0.00104	0.323
2HR GLUCOSE	EMP2	0.00242	0.332
2HR GLUCOSE	SERF2	0.000975	0.323
2HR GLUCOSE	UBR4	0.000363	0.201
2HR GLUCOSE	ZFAND2A	0.00177	0.323
2HR GLUCOSE	ASIC2	0.00204	0.327
2HR GLUCOSE	SCGB1D2	0.00237	0.332
2HR GLUCOSE	SSR2	3.19E-05	0.0648
2HR GLUCOSE	MAN1B1	0.000341	0.198
2HR GLUCOSE	TPD52L2	0.000348	0.198
2HR GLUCOSE	SAMD8	0.000413	0.209
2HR GLUCOSE	TPT1-AS1	0.0021	0.329
2HR GLUCOSE	LY6E	0.00158	0.323
2HR GLUCOSE	IQSEC3	0.00138	0.323
2HR GLUCOSE	MTFR1L	0.00196	0.323
2HR GLUCOSE	LPCAT4	3.47E-05	0.0648
2HR GLUCOSE	MFSD12	0.00243	0.332
2HR GLUCOSE	ENTPD6	0.00113	0.323
2HR GLUCOSE	DLGAP4	0.00183	0.323
2HR GLUCOSE	FAM3A	0.00145	0.323
2HR GLUCOSE	MAST3	0.00152	0.323
2HR GLUCOSE	EXOC8	0.00124	0.323
2HR GLUCOSE	CHRM1	0.00166	0.323
2HR GLUCOSE	ZNRF4	0.000994	0.323
2HR GLUCOSE	SAMD11	0.00228	0.332
2HR GLUCOSE	RPS19	0.00244	0.332
2HR GLUCOSE	MICAL1	0.00161	0.323

phenotype	gene	p-value	Q-value
2HR GLUCOSE	PRKCB	0.000393	0.209
2HR GLUCOSE	HDAC6	0.00247	0.332
2HR GLUCOSE	FTO	0.00213	0.331
2HR GLUCOSE	GANAB	0.000346	0.198
2HR GLUCOSE	ISCA2	0.00226	0.332
2HR GLUCOSE	DDB1	0.00123	0.323
2HR GLUCOSE	TJAP1	0.00183	0.323
2HR GLUCOSE	C7orf26	0.000343	0.198
2HR GLUCOSE	NRSN1	0.000711	0.283
2HR GLUCOSE	GBP4	0.00194	0.323
2HR GLUCOSE	RBM23	0.00162	0.323
2HR GLUCOSE	ZNF777	0.00164	0.323
2HR GLUCOSE	MLF2	0.0016	0.323
2HR GLUCOSE	CST3	0.00122	0.323
2HR GLUCOSE	CKMT1B	1.82E-07	0.00218
2HR GLUCOSE	DYNC2LI1	0.00199	0.323
2HR GLUCOSE	GAA	0.00231	0.332
2HR GLUCOSE	NGEF	0.00135	0.323
2HR GLUCOSE	FUNDC2	0.0022	0.332
2HR GLUCOSE	PPP1R1B	0.00237	0.332
2HR GLUCOSE	GPX4	1.51E-06	0.00904
2HR GLUCOSE	ELMOD3	0.00204	0.327
2HR GLUCOSE	RAP1GAP	0.00215	0.331
2HR GLUCOSE	NR1H2	0.000726	0.284
2HR GLUCOSE	ITGB4	0.00133	0.323
2HR GLUCOSE	CPNE1	0.00241	0.332
2HR GLUCOSE	SIPA1L3	0.00148	0.323
2HR GLUCOSE	WASH2P	0.000639	0.265
2HR GLUCOSE	ZCCHC11	0.000512	0.231
2HR GLUCOSE	CKAP2L	0.00123	0.323
2HR GLUCOSE	TUBB4A	0.000427	0.209
2HR GLUCOSE	SULT4A1	0.00206	0.327
2HR GLUCOSE	C19orf70	0.00174	0.323
2HR GLUCOSE	PPP1R15A	0.00077	0.292
2HR GLUCOSE	DPM2	2.55E-05	0.0648
2HR GLUCOSE	ICAM5	0.00187	0.323
2HR GLUCOSE	LAMB3	2.53E-05	0.0648
2HR GLUCOSE	ROGDI	0.00224	0.332
2HR GLUCOSE	P4HB	8.67E-05	0.101
2HR GLUCOSE	PPP6R1	0.00113	0.323
ACUTE MYELOID LEUKEMIA	LIPI	9.31E-06	0.283
AGE AT MENARCHE	BDNF	0.000137	0.26
AGE AT MENARCHE	SLC16A4	6.31E-05	0.22
AGE AT MENARCHE	EXOSC6	8.71E-05	0.248
AGE AT MENARCHE	RHOA	2.55E-06	0.0133
AGE AT MENARCHE	RBM6	0.000129	0.26
AGE AT MENARCHE	NOB1	0.000104	0.327
AGE AT MENARCHE	WWP2	0.000184	0.32
AGE AT MENARCHE	TMEM38B	8.79E-05	0.306
AGE AT MENARCHE	IP6K1	1.22E-05	0.048
AGE AT MENARCHE	SATB2	5.33E-07	0.00835
AGE AT MENARCHE	RNF123	9.72E-06	0.0436
AGE AT MENARCHE	SMAD3	0.000141	0.26
AGE AT MENARCHE	SERPINA2	7.56E-05	0.296
AGE AT MENARCHE	NFAT5	7.91E-05	0.248
AGE AT MENARCHE	DOCK3	0.000107	0.26
AGE RELATED MACULAR	LHX9	2.11E-05	0.0755
AGE RELATED MACULAR	TRIM4	0.00012	0.247

phenotype	gene	p-value	Q-value
AGE RELATED MACULAR	BTBD16	3.97E-05	0.109
AGE RELATED MACULAR	PLA2G12A	0.000131	0.247
AGE RELATED MACULAR	ZAN	0.000127	0.247
AGE RELATED MACULAR	CFHR1	4.75E-07	0.00382
AGE RELATED MACULAR	GABRR2	3.86E-05	0.109
AGE RELATED MACULAR	CASP6	4.07E-05	0.109
AGE RELATED MACULAR	SMU1	6.53E-06	0.0511
AGE RELATED MACULAR	CFHR4	1.00E-05	0.043
AGE RELATED MACULAR	CLK4	5.33E-05	0.132
AGE RELATED MACULAR	CFHR3	7.01E-12	2.25E-07
AGE RELATED MACULAR	KCNT2	1.21E-06	0.0189
AGE RELATED MACULAR	ZBTB41	7.47E-07	0.0048
AGE RELATED MACULAR	CFH	1.74E-07	0.00186
AGE RELATED MACULAR	ZNF766	9.67E-05	0.222
AGE RELATED MACULAR	PLEKHA1	3.15E-10	5.06E-06
AGE RELATED MACULAR	STAG3L5P	1.07E-05	0.043
ALCOHOLISM	EZH2	1.97E-05	0.305
ALZHEIMER	EHD2	0.000189	0.253
ALZHEIMER	CFHR3	6.68E-05	0.247
ALZHEIMER	OPA3	2.24E-06	0.0141
ALZHEIMER	KCTD5	5.49E-05	0.232
ALZHEIMER	FAM151A	2.27E-07	0.00404
ALZHEIMER	DCN	0.000158	0.23
ALZHEIMER	CC2D1A	7.87E-05	0.318
ALZHEIMER	PALM2	2.17E-05	0.214
ALZHEIMER	CR1	2.48E-05	0.0724
ALZHEIMER	L3HYPDH	6.90E-06	0.0681
ALZHEIMER	MARK4	5.12E-07	0.00557
ALZHEIMER	SETD5	8.13E-05	0.157
ALZHEIMER	ADAM23	0.000213	0.264
ALZHEIMER	GGCT	5.07E-05	0.27
ALZHEIMER	CPA5	2.76E-05	0.163
ALZHEIMER	CLASRP	1.24E-05	0.0442
ALZHEIMER	CLPTM1	8.83E-05	0.221
ALZHEIMER	TAS2R60	0.00015	0.229
ALZHEIMER	LINC00639	9.17E-05	0.271
ALZHEIMER	CEACAM19	4.61E-09	0.000149
ALZHEIMER	SV2B	4.85E-08	0.00155
ALZHEIMER	INTU	2.69E-05	0.214
ALZHEIMER	ZNF222	0.000112	0.234
ALZHEIMER	PVRIG	0.000114	0.234
ALZHEIMER	ZNF133	0.000183	0.257
ALZHEIMER	EML2	1.92E-06	0.0124
ALZHEIMER	PCDH19	2.06E-05	0.329
ALZHEIMER	PIGR	0.000103	0.174
ALZHEIMER	PPP1R37	7.16E-06	0.0288
ALZHEIMER	ZNF235	0.000122	0.234
ALZHEIMER	MDK	3.69E-05	0.236
ALZHEIMER	ELMO1	7.09E-05	0.318
ALZHEIMER	NFATC2IP	3.44E-05	0.17
ALZHEIMER	SYT4	3.77E-06	0.0302
ALZHEIMER	RIN1	0.000104	0.28
ALZHEIMER	FOSB	4.77E-05	0.141
ALZHEIMER	APOC1	6.43E-05	0.122
ALZHEIMER	JKAMP	1.52E-05	0.112
ALZHEIMER	ZNF225	2.59E-06	0.0141
ALZHEIMER	APOE	6.20E-09	9.96E-05
ALZHEIMER	ENKD1	8.95E-05	0.318

phenotype	gene	p-value	Q-value
ALZHEIMER	PTK2B	3.08E-06	0.0141
ALZHEIMER	MS4A6A	0.000174	0.244
ALZHEIMER	GPR135	2.73E-07	0.00404
ALZHEIMER	ERCC2	4.63E-05	0.0992
ALZHEIMER	EPHA1-AS1	8.25E-05	0.157
ALZHEIMER	ZNF226	4.52E-06	0.021
ALZHEIMER	ZNF417	8.79E-05	0.318
ALZHEIMER	GEMIN7	2.08E-06	0.0141
ALZHEIMER	NKPD1	3.24E-05	0.0771
ALZHEIMER	SEZ6L	3.35E-05	0.214
ALZHEIMER	KLC3	3.36E-05	0.0771
ALZHEIMER	BRMS1L	7.56E-05	0.249
ALZHEIMER	APOC1P1	0.000243	0.289
ALZHEIMER	AKAP2	8.53E-06	0.214
ANOREXIA NERVOSA	PP2D1	1.86E-06	0.0575
ASTHMA	ORMDL3	3.04E-05	0.189
ASTHMA	GSDMB	2.57E-06	0.0398
ASTHMA	CCDC117	4.31E-05	0.199
ASTHMA	PNMT	7.57E-06	0.0781
ASTHMA	CACNB1	2.46E-06	0.0398
ASTHMA	RAPGEFL1	2.72E-05	0.189
ASTHMA	ZPBP2	7.41E-05	0.232
ASTHMA	IKZF3	7.11E-05	0.232
AUTISM	DNM1P46	1.12E-05	0.253
AUTISM	WFDC6	8.77E-06	0.272
AUTISM	MAPT	5.20E-06	0.0805
AUTISM	UBE2H	3.41E-06	0.106
AUTISM	ZNRF2	3.47E-05	0.174
AUTISM	CLDN23	3.85E-05	0.174
AUTISM	IRX1	4.85E-05	0.192
AUTISM	EXTL3-AS1	6.25E-05	0.324
AUTISM	PDCL3	7.71E-05	0.298
AUTISM	INSL3	3.14E-05	0.324
AUTISM	OR5M10	9.25E-06	0.265
AUTISM	LEPR	1.28E-06	0.0398
AUTISM	NF2	0.000107	0.322
AUTISM	PPP1R3B	6.80E-06	0.108
AUTISM	GBP3	2.57E-05	0.265
AUTISM	LIMD2	4.35E-06	0.135
AUTISM	NEIL2	5.46E-05	0.185
AUTISM	SGK1	2.78E-06	0.0863
AUTISM	CITED1	5.85E-05	0.298
AUTISM	BPTF	1.22E-05	0.189
AUTISM	DND1	5.58E-05	0.323
AUTISM	RNF126	2.79E-05	0.324
AUTISM	UBE2Q2P1	2.61E-06	0.0805
AUTISM	SRC	3.79E-05	0.294
AUTISM	ROBO1	8.96E-06	0.278
AUTISM	RNF152	5.63E-05	0.298
AUTISM	FBXW4P1	6.97E-05	0.309
AUTISM	FAM41C	2.10E-05	0.217
AUTISM	ZNF641	4.77E-06	0.148
AUTISM	LRFN4	4.30E-06	0.133
AUTISM	TMEM14A	7.32E-05	0.324
AUTISM	EXOC4	1.44E-05	0.152
AUTISM	SLC35G5	8.82E-05	0.254
AUTISM	USP11	6.26E-05	0.323
AUTISM	AFF1	1.63E-05	0.253

phenotype	gene	p-value	Q-value
AUTISM	RIMKLA	4.70E-06	0.146
AUTISM	ZNF713	1.63E-05	0.127
AUTISM	PHTF1	1.18E-05	0.184
AUTISM	CLDN22	9.85E-06	0.305
AUTISM	MTMR9	5.20E-06	0.0825
AUTISM	NSUN7	1.51E-05	0.265
AUTISM	DPH7	8.69E-06	0.27
AUTISM	BNIPL	1.54E-06	0.0476
AUTISM	OR4D5	1.13E-05	0.175
AUTISM	TTBK1	3.07E-05	0.162
AUTISM	BID	9.89E-06	0.153
AUTISM	PPIH	1.61E-05	0.285
AUTISM	ARMC2	9.26E-05	0.324
AUTISM	RBM33	2.57E-05	0.265
AUTISM	GPR141	1.93E-05	0.299
AUTISM	LYAR	5.06E-05	0.324
AUTISM	BPESC1	4.12E-05	0.258
AUTISM	AQR	3.34E-05	0.207
AUTISM	NOXO1	6.02E-05	0.324
AUTISM	TMEM222	1.55E-05	0.127
AUTISM	IL20RB	2.54E-07	0.00789
AUTISM	RAD51	4.21E-05	0.324
AUTISM	DZANK1	2.63E-05	0.285
AUTISM	RRP12	7.43E-06	0.0769
AUTISM	SIGMAR1	1.61E-05	0.324
AUTISM	NPRL2	6.65E-05	0.251
AUTISM	C2orf88	3.19E-05	0.33
AUTISM	ENPP4	4.91E-08	0.00153
AUTISM	MSRA	5.83E-05	0.185
AUTISM	C17orf80	1.18E-05	0.183
AUTISM	RSPO1	8.25E-05	0.324
AUTISM	MTHFR	1.79E-06	0.0557
AUTISM	ACSF2	1.00E-05	0.31
AUTISM	FAM66A	1.62E-05	0.128
AUTISM	UBE2L3	1.08E-05	0.183
AUTISM	PCDHGC5	6.63E-05	0.324
AUTISM	5-Mar	1.87E-05	0.265
AUTISM	ENPP5	2.39E-05	0.187
BIPOLAR DISORDER	CDAN1	5.22E-06	0.16
BIPOLAR DISORDER	SLC7A6	1.58E-05	0.0992
BIPOLAR DISORDER	LMAN2L	1.00E-05	0.0992
BIPOLAR DISORDER	DDN	1.60E-06	0.0251
BIPOLAR DISORDER	ITIH4	1.37E-05	0.0992
BIPOLAR DISORDER	CACNB3	2.14E-08	0.000671
BIPOLAR DISORDER AND	BTN3A2	4.11E-11	1.29E-06
BIPOLAR DISORDER AND	NT5C2	6.29E-07	0.00492
BIPOLAR DISORDER AND	FRMD7	2.93E-05	0.0918
BIPOLAR DISORDER AND	BTN1A1	2.03E-07	0.00212
BIPOLAR DISORDER AND	SNF8	0.000101	0.293
BIPOLAR DISORDER AND	ARL17A	3.71E-07	0.0116
BIPOLAR DISORDER AND	CCDC36	1.95E-05	0.0873
BIPOLAR DISORDER AND	CACNB3	2.85E-08	0.000892
BIPOLAR DISORDER AND	HIST1H4C	5.60E-06	0.0292
BIPOLAR DISORDER AND	SPTBN2	2.75E-05	0.108
BIPOLAR DISORDER AND	LPAR3	4.65E-05	0.242
BIPOLAR DISORDER AND	ARL3	6.68E-07	0.0104
BIPOLAR DISORDER AND	MVP	1.26E-05	0.0873
BIPOLAR DISORDER AND	MUSTN1	7.84E-05	0.223

phenotype	gene	p-value	Q-value
BIPOLAR DISORDER AND	RCE1	6.35E-06	0.0663
BIPOLAR DISORDER AND	CTSF	1.60E-05	0.0873
BIPOLAR DISORDER AND	ITIH4	1.26E-05	0.0656
BIPOLAR DISORDER AND	PRKAG1	9.28E-05	0.242
BIPOLAR DISORDER AND	C11orf80	4.69E-05	0.133
BIPOLAR DISORDER AND	RNPS1	0.000103	0.293
BIPOLAR DISORDER AND	PSMD3	9.38E-05	0.293
BIPOLAR DISORDER AND	SFXN2	0.000105	0.234
BIPOLAR DISORDER AND	MCHR1	2.49E-05	0.0865
BIPOLAR DISORDER AND	DDN	5.11E-06	0.0663
BIPOLAR DISORDER AND	WFDC12	1.93E-05	0.0873
BIRTH LENGTH	GPR6	9.87E-06	0.153
BIRTH LENGTH	ZBTB7B	3.02E-06	0.0937
BIRTH WEIGHT	TIPARP	4.68E-05	0.323
BIRTH WEIGHT	POC1A	5.14E-05	0.323
BIRTH WEIGHT	RBBP9	6.12E-05	0.323
BIRTH WEIGHT	CHADL	6.67E-06	0.211
BIRTH WEIGHT	HMGA2	5.76E-05	0.323
BLOOD_TEST_INDICATOR	CWF19L1	1.33E-06	0.0435
BLOOD_TEST_INDICATOR	ABHD1	1.72E-05	0.282
BLOOD_TEST_INDICATOR	DOCK7	3.04E-06	0.0497
BLOOD_TEST_INDICATOR	OR2AE1	1.35E-05	0.147
BLOOD_TEST_INDICATOR	CNNM3	2.49E-05	0.203
BLOOD_TEST_INDICATOR	ALDH1A2	2.93E-06	0.096
BLOOD_TEST_INDICATOR	GATM	3.93E-05	0.305
BLOOD_TEST_INDICATOR	HFE	5.90E-05	0.305
BLOOD_TEST_INDICATOR	HNF1A	1.98E-06	0.0648
BLOOD_TEST_INDICATOR	MRC1	5.03E-06	0.0822
BLOOD_TEST_INDICATOR	PSTK	5.46E-05	0.305
BLOOD_TEST_INDICATOR	CYP3A7	2.44E-08	0.000586
BLOOD_TEST_INDICATOR	CCDC110	3.66E-05	0.238
BLOOD_TEST_INDICATOR	NR1H3	3.85E-06	0.126
BLOOD_TEST_INDICATOR	CSF3	1.52E-05	0.203
BLOOD_TEST_INDICATOR	MAN2A2	4.55E-06	0.149
BLOOD_TEST_INDICATOR	AP2M1	2.57E-06	0.0822
BLOOD_TEST_INDICATOR	BCAM	2.46E-05	0.269
BLOOD_TEST_INDICATOR	SIM1	1.92E-05	0.305
BLOOD_TEST_INDICATOR	NAP1L2	2.46E-05	0.203
BLOOD_TEST_INDICATOR	SIAH3	6.62E-05	0.308
BLOOD_TEST_INDICATOR	NBPF3	1.10E-05	0.119
BLOOD_TEST_INDICATOR	CDKN2D	1.67E-05	0.203
BLOOD_TEST_INDICATOR	SPATA5L1	2.34E-05	0.305
BLOOD_TEST_INDICATOR	C15orf48	9.13E-05	0.332
BLOOD_TEST_INDICATOR	PSMA5	2.44E-05	0.269
BLOOD_TEST_INDICATOR	ZNF280A	3.30E-05	0.305
BLOOD_TEST_INDICATOR	ANGPTL3	3.84E-07	0.0126
BODY MASS INDEX	FBXL19	2.01E-05	0.0443
BODY MASS INDEX	FAM180B	0.000112	0.187
BODY MASS INDEX	TM6SF2	4.06E-05	0.0709
BODY MASS INDEX	NPIPB9	0.000102	0.269
BODY MASS INDEX	TUFM	6.79E-06	0.0194
BODY MASS INDEX	BDNF	0.000133	0.321
BODY MASS INDEX	ANKDD1B	0.000127	0.181
BODY MASS INDEX	RPS10	9.72E-05	0.153
BODY MASS INDEX	SH3YL1	0.000204	0.246
BODY MASS INDEX	HCN4	3.02E-06	0.0479
BODY MASS INDEX	C1QL2	9.91E-06	0.0619
BODY MASS INDEX	SLC39A13	0.000123	0.193

phenotype	gene	p-value	Q-value
BODY MASS INDEX	MTCH2	0.000113	0.187
BODY MASS INDEX	FKBP1B	0.000133	0.182
BODY MASS INDEX	CENPO	7.25E-05	0.256
BODY MASS INDEX	MYBPC3	3.01E-05	0.0727
BODY MASS INDEX	C1QTNF4	5.47E-05	0.244
BODY MASS INDEX	RBL2	3.22E-05	0.0562
BODY MASS INDEX	NDUFS3	5.94E-06	0.0627
BODY MASS INDEX	UHRF1BP1	1.86E-05	0.0419
BODY MASS INDEX	FEM1B	0.000247	0.272
BODY MASS INDEX	ACP1	0.000129	0.283
BODY MASS INDEX	ADCY3	3.20E-06	0.05
BODY MASS INDEX	NPIPB6	7.32E-07	0.00583
BODY MASS INDEX	SH2B1	4.89E-05	0.0959
BODY MASS INDEX	TFAP2B	5.56E-06	0.0194
BODY MASS INDEX	DNAJC27-AS1	8.21E-05	0.256
BODY MASS INDEX	FNBP4	0.000134	0.281
BODY MASS INDEX	NUPR1	4.06E-06	0.0155
BODY MASS INDEX	SNRPC	0.000241	0.261
BODY MASS INDEX	OR14A16	0.000152	0.199
BODY MASS INDEX	FUBP1	2.80E-07	0.00509
BODY MASS INDEX	NFATC2IP	1.08E-05	0.0679
BODY MASS INDEX	BPIFB1	0.000283	0.329
BODY MASS INDEX	NEGR1	1.93E-07	0.00606
BODY MASS INDEX	DNMT3L	0.000257	0.322
BODY MASS INDEX	PMS2P3	2.85E-07	0.00751
BODY MASS INDEX	EIF3CL	2.48E-06	0.0259
BODY MASS INDEX	C18orf54	6.02E-05	0.21
BODY MASS INDEX	LGR4	2.98E-05	0.104
BREAST CANCER	PDE4A	0.00116	0.292
BREAST CANCER	TTYH1	0.000768	0.313
BREAST CANCER	WDR13	0.00273	0.332
BREAST CANCER	AP5M1	0.00108	0.313
BREAST CANCER	BLOC1S2	3.29E-07	0.00829
BREAST CANCER	FCN1	0.000849	0.292
BREAST CANCER	RXRB	0.00125	0.292
BREAST CANCER	ACTG1	0.000905	0.292
BREAST CANCER	CHCHD6	0.00221	0.319
BREAST CANCER	CORO2B	0.00164	0.297
BREAST CANCER	SSR2	0.00124	0.292
BREAST CANCER	GATAD2A	0.000272	0.3
BREAST CANCER	SLC2A1	0.002	0.319
BREAST CANCER	PHC2	0.000221	0.18
BREAST CANCER	SLC8A2	0.00245	0.329
BREAST CANCER	PRKCZ	0.000408	0.245
BREAST CANCER	PSMD6	4.29E-05	0.1
BREAST CANCER	CCDC14	0.0024	0.324
BREAST CANCER	WDR45B	0.00043	0.252
BREAST CANCER	SHISA4	4.80E-06	0.0242
BREAST CANCER	KEAP1	0.00194	0.311
BREAST CANCER	ABHD14A	0.000203	0.171
BREAST CANCER	DHPS	0.00202	0.312
BREAST CANCER	NSUN4	0.000274	0.3
BREAST CANCER	LSM7	0.00142	0.292
BREAST CANCER	RPS19	1.11E-05	0.0312
BREAST CANCER	OGDHL	0.00277	0.332
BREAST CANCER	CNTROB	0.000528	0.267
BREAST CANCER	NCAN	0.00137	0.292
BREAST CANCER	C1orf61	0.00163	0.314

phenotype	gene	p-value	Q-value
BREAST CANCER	C1orf35	0.00226	0.321
BREAST CANCER	LRP1	0.00198	0.311
BREAST CANCER	FOXJ2	5.79E-05	0.0841
BREAST CANCER	ZNF347	0.00187	0.314
BREAST CANCER	PI4KAP2	0.00157	0.314
BREAST CANCER	OR4D1	2.49E-05	0.074
BREAST CANCER	LHX6	0.00228	0.321
BREAST CANCER	PTGS2	0.000456	0.254
BREAST CANCER	SLITRK1	0.0019	0.315
BREAST CANCER	PTPN23	0.000127	0.133
BREAST CANCER	CDK16	0.00117	0.313
BREAST CANCER	SUPT16H	0.000641	0.289
BREAST CANCER	CCDC85B	0.0003	0.205
BREAST CANCER	SHARPIN	7.87E-06	0.0312
BREAST CANCER	COPE	0.00201	0.312
BREAST CANCER	ROGDI	5.68E-05	0.0841
BREAST CANCER	EFNB3	0.000825	0.313
BREAST CANCER	DPM2	0.000311	0.207
BREAST CANCER	FAM213B	0.000252	0.181
BREAST CANCER	TMEM237	0.000353	0.316
BREAST CANCER	BRINP2	0.00114	0.313
BREAST CANCER	ZNF777	0.00214	0.318
BREAST CANCER	DND1	0.000292	0.3
BREAST CANCER	C19orf66	0.00121	0.292
BREAST CANCER	SHISA5	8.58E-05	0.108
BREAST CANCER	KCNJ8	0.00176	0.314
BREAST CANCER	KCNN4	7.35E-05	0.136
BREAST CANCER	TESK2	0.000253	0.3
BREAST CANCER	AP2A2	0.00263	0.332
BREAST CANCER	KLHL25	0.000337	0.233
BREAST CANCER	SCGB1D2	0.000264	0.185
BREAST CANCER	HSCB	0.000293	0.3
BREAST CANCER	TNFSF9	0.00107	0.313
BREAST CANCER	GKN1	0.00169	0.299
BREAST CANCER	LYPD5	0.000193	0.263
BREAST CANCER	HCG27	3.20E-05	0.316
BREAST CANCER	LRR1	0.00154	0.297
BREAST CANCER	NSFL1C	0.000877	0.292
BREAST CANCER	TRIM41	0.000906	0.292
BREAST CANCER	CDC37	0.00142	0.292
BREAST CANCER	CCND3	0.00274	0.332
BREAST CANCER	EDF1	8.44E-05	0.0974
BREAST CANCER	HEXA-AS1	0.00136	0.292
BREAST CANCER	ALKBH7	0.000786	0.292
BREAST CANCER	DNM3	0.0012	0.292
BREAST CANCER	HDAC3	0.00186	0.311
BREAST CANCER	TMEM231	0.0022	0.319
BREAST CANCER	MLF2	0.00185	0.311
BREAST CANCER	KLC1	0.00205	0.315
BREAST CANCER	ARPC5	0.0015	0.313
BREAST CANCER	DDN	2.59E-05	0.0592
BREAST CANCER	PPP1CA	5.92E-05	0.0841
BREAST CANCER	PDAP1	0.00173	0.303
BREAST CANCER	EIF5A2	0.00235	0.325
BREAST CANCER	TRIM2	0.00197	0.311
BREAST CANCER	SLC22A17	0.00159	0.297
BREAST CANCER	ACOT11	0.000672	0.313
BREAST CANCER	NUDC	3.64E-06	0.0264

phenotype	gene	p-value	Q-value
BREAST CANCER	NOSIP	0.000154	0.155
BREAST CANCER	SSNA1	0.00126	0.292
BREAST CANCER	DLEU2	0.00162	0.314
BREAST CANCER	TTLL1	0.00166	0.297
BREAST CANCER	RPLP1	0.000235	0.181
BREAST CANCER	ASH1L	0.00016	0.228
BREAST CANCER	MAGI2	0.00208	0.319
BREAST CANCER	FAM3A	6.74E-05	0.0894
BREAST CANCER	ENTPD8	0.00248	0.325
BREAST CANCER	VPS4A	8.50E-05	0.0974
BREAST CANCER	MRPL47	0.000112	0.122
BREAST CANCER	TYRO3	0.0015	0.297
BREAST CANCER	AGXT	0.000872	0.292
BREAST CANCER	PYGB	0.00269	0.332
BREAST CANCER	RUVBL1	0.000318	0.308
BREAST CANCER	PNPLA6	0.00218	0.321
BREAST CANCER	NRGN	0.000619	0.313
BREAST CANCER	VBP1	0.000998	0.292
BREAST CANCER	MAGED2	0.00104	0.313
BREAST CANCER	BEAN1	0.00171	0.314
BREAST CANCER	HIST2H2BA	9.69E-08	0.00159
BREAST CANCER	ZNF48	0.000725	0.313
BREAST CANCER	STRN4	0.00123	0.292
BREAST CANCER	TUBGCP2	0.00103	0.292
BREAST CANCER	POLR2A	0.000978	0.292
BREAST CANCER	B4GALNT4	0.000281	0.3
BREAST CANCER	APEH	0.000831	0.292
BREAST CANCER	BAG6	4.13E-05	0.0802
BREAST CANCER	ATG10	0.000358	0.316
BREAST CANCER	AARSD1	0.00123	0.292
BREAST CANCER	TMEM141	0.00242	0.324
BREAST CANCER	PPP6R1	0.000408	0.245
BREAST CANCER	MLC1	0.00178	0.314
BREAST CANCER	RAD51B	3.09E-05	0.0843
BREAST CANCER	FAM200B	0.00116	0.313
BREAST CANCER	SGCE	1.53E-05	0.0557
BREAST CANCER	NUDT5	0.00147	0.297
BREAST CANCER	GPX4	2.15E-05	0.0543
BREAST CANCER	HDGF	0.00149	0.313
BREAST CANCER	BTBD7	0.00177	0.314
BREAST CANCER	FUNDC2	0.00249	0.325
BREAST CANCER	VPS4B	0.00136	0.292
BREAST CANCER	CD83	5.01E-05	0.316
BREAST CANCER	TNFAIP8L3	0.00136	0.313
BREAST CANCER	SEC13	0.00234	0.32
BREAST CANCER	RPS6KB2	0.00255	0.328
BREAST CANCER	MICAL3	0.00225	0.319
BREAST CANCER	C11orf49	0.00232	0.32
BREAST CANCER	MAPK8IP1	2.74E-05	0.0624
BREAST CANCER	ST7	0.00199	0.311
BREAST CANCER	TAOK2	0.000752	0.292
BREAST CANCER	CYP3A5	0.00101	0.292
BREAST CANCER	GAS7	0.00122	0.313
BREAST CANCER	GABARAP	0.000334	0.212
BREAST CANCER	CRHR1	0.000142	0.211
BREAST CANCER	IGSF9B	0.000167	0.158
BREAST CANCER	HEXB	0.0016	0.297
BREAST CANCER	FAM13B	0.00126	0.313

phenotype	gene	p-value	Q-value
BREAST CANCER	OPTN	0.0025	0.325
BREAST CANCER	CAPNS1	0.00125	0.313
BREAST CANCER	RPL7A	0.00276	0.332
BREAST CANCER	DDIT3	0.0017	0.314
BREAST CANCER	SUPT5H	0.00198	0.311
BREAST CANCER	MTFR1L	3.72E-06	0.0264
BREAST CANCER	RPL4	0.00192	0.311
BREAST CANCER	ELMO1	0.0013	0.292
BREAST CANCER	ESYT2	0.000753	0.292
BREAST CANCER	KXD1	0.000599	0.285
BREAST CANCER	DUS1L	5.52E-05	0.0841
BREAST CANCER	PDIA3	0.000176	0.158
BREAST CANCER	SLC9A3R1	0.00108	0.292
BREAST CANCER	CRIP2	0.000157	0.152
BREAST CANCER	FDXACB1	1.66E-05	0.0557
BREAST CANCER	LHX2	9.62E-06	0.0346
BREAST CANCER	RPL26L1	0.00148	0.297
BREAST CANCER	GGA3	0.00268	0.332
BREAST CANCER	C7orf26	0.00124	0.292
BREAST CANCER	GANAB	0.00136	0.292
BREAST CANCER	GNG10	0.00203	0.319
BREAST CANCER	RBM23	0.00134	0.292
BREAST CANCER	CRAT	0.000646	0.289
BREAST CANCER	MAP7D1	0.000937	0.292
BREAST CANCER	EIF3G	0.00192	0.311
BREAST CANCER	ARF4	0.00181	0.311
BREAST CANCER	KCNJ4	0.00239	0.327
BREAST CANCER	WBP2	0.00142	0.292
BREAST CANCER	TPD52L2	0.000337	0.212
BREAST CANCER	STIP1	0.000677	0.289
BREAST CANCER	P4HA1	0.00234	0.32
BREAST CANCER	TRAPPC12	0.00106	0.292
BREAST CANCER	YDJC	0.00111	0.292
BREAST CANCER	RNF220	0.00137	0.292
BREAST CANCER	CTNNA3	0.000729	0.292
BREAST CANCER	DLGAP4	0.00211	0.318
BREAST CANCER	SLC4A3	0.00209	0.319
BREAST CANCER	LRRC37A2	0.000292	0.3
BREAST CANCER	TMED9	0.000961	0.292
BREAST CANCER	NR1H2	0.00182	0.311
BREAST CANCER	RIIAD1	0.00241	0.324
BREAST CANCER	FLII	0.000514	0.267
BREAST CANCER	CMC2	7.39E-05	0.0931
BREAST CANCER	UBXN6	0.000896	0.292
BREAST CANCER	SENP3	0.000949	0.292
BREAST CANCER	SDF4	0.00247	0.325
BREAST CANCER	BAD	0.00111	0.292
BREAST CANCER	P4HB	0.000249	0.181
BREAST CANCER	GAA	0.000738	0.292
BREAST CANCER	ENTPD6	6.01E-05	0.0841
BREAST CANCER	LCA5	0.00182	0.311
BREAST CANCER	PSMF1	0.00161	0.297
BREAST CANCER	FAP	0.00195	0.311
BREAST CANCER	ITPK1	0.00122	0.292
BREAST CANCER	TET2-AS1	0.000122	0.19
BREAST CANCER	TUBB2B	0.00204	0.319
BREAST CANCER	TAF15	0.000615	0.287
BREAST CANCER	DDB1	0.00264	0.332

phenotype	gene	p-value	Q-value
BREAST CANCER	PHB2	0.000821	0.292
BREAST CANCER	NLGN2	0.00104	0.292
BREAST CANCER	SLC7A4	0.00228	0.319
BREAST CANCER	RDM1	5.81E-05	0.127
BREAST CANCER	PAK1	0.00225	0.321
BREAST CANCER	TMEM132A	0.00155	0.297
BREAST CANCER	PPP2R5B	0.000249	0.181
BREAST CANCER	ADRA2C	0.00205	0.319
BREAST CANCER	NIT2	0.000345	0.316
BREAST CANCER	CST3	0.000859	0.292
BREAST CANCER	FURIN	0.00223	0.319
BREAST CANCER	AKT1	0.000982	0.292
BREAST CANCER	LRRCC1	0.00187	0.314
BREAST CANCER	MAN2C1	0.000882	0.292
BREAST CANCER	9-Sep	0.00116	0.292
BREAST CANCER	RPS2	0.00243	0.324
BREAST CANCER	PSAP	0.00124	0.292
BREAST CANCER	COX6B1	0.00172	0.303
CHILDHOOD INTELLIGENCE	SDHC	5.36E-06	0.163
CHOLESTEROL	SDC1	9.55E-05	0.131
CHOLESTEROL	RBKS	0.000113	0.127
CHOLESTEROL	MPV17	1.40E-06	0.00507
CHOLESTEROL	FER1L4	5.58E-05	0.0979
CHOLESTEROL	AMIGO1	0.000171	0.175
CHOLESTEROL	TRIP11	0.000372	0.31
CHOLESTEROL	FNDC4	1.62E-05	0.0348
CHOLESTEROL	KIAA1324	0.000316	0.261
CHOLESTEROL	APOB	4.36E-05	0.0957
CHOLESTEROL	TAF6L	0.000224	0.245
CHOLESTEROL	MAFB	0.000156	0.17
CHOLESTEROL	PTPN13	5.41E-05	0.0901
CHOLESTEROL	ANGPTL3	2.15E-06	0.00618
CHOLESTEROL	MAU2	0.000292	0.254
CHOLESTEROL	YIPF2	0.000488	0.325
CHOLESTEROL	NEIL2	1.90E-05	0.0376
CHOLESTEROL	MLEC	1.76E-05	0.0558
CHOLESTEROL	HP	0.000165	0.167
CHOLESTEROL	IFT172	4.96E-08	0.000392
CHOLESTEROL	CGREF1	9.47E-07	0.00499
CHOLESTEROL	GDF5	3.00E-05	0.079
CHOLESTEROL	PLS3	2.31E-05	0.0664
CHOLESTEROL	SLC4A1AP	7.94E-05	0.114
CHOLESTEROL	FKBP7	0.000412	0.301
CHOLESTEROL	RPAP2	0.00046	0.32
CHOLESTEROL	POLK	2.67E-05	0.0597
CHOLESTEROL	WTAP	4.89E-07	0.00766
CHOLESTEROL	BCAM	5.17E-05	0.0979
CHOLESTEROL	AFF1	2.57E-05	0.0463
CHOLESTEROL	ATP13A1	1.99E-06	0.0104
CHOLESTEROL	HS1BP3-IT1	4.54E-05	0.0957
CHOLESTEROL	C6orf106	0.00023	0.218
CHOLESTEROL	KHK	1.61E-05	0.0348
CHOLESTEROL	FADS3	9.65E-05	0.133
CHOLESTEROL	ZNF441	0.000111	0.147
CHOLESTEROL	DOCK7	4.67E-08	0.000392
CHOLESTEROL	VAPA	0.00018	0.211
CHOLESTEROL	ADAL	0.000131	0.166
CHOLESTEROL	ATMIN	1.65E-05	0.0348

phenotype	gene	p-value	Q-value
CHOLESTEROL	MLXIPL	1.42E-08	0.000225
CHOLESTEROL	SLC44A2	5.31E-06	0.0177
CHOLESTEROL	LPIN3	0.00013	0.136
CHOLESTEROL	GSTM2	0.000333	0.261
CHOLESTEROL	POC5	4.62E-05	0.0804
CHOLESTEROL	PPP1R3B	0.000258	0.263
CHOLESTEROL	NCAN	0.000413	0.301
CHOLESTEROL	C10orf67	0.00028	0.25
CHOLESTEROL	SIDT2	2.41E-09	7.63E-05
CHOLESTEROL	MKRN2	1.91E-06	0.0104
CHOLESTEROL	PCSK7	0.000208	0.235
CHOLESTEROL	PAFAH1B2	6.18E-05	0.0976
CHOLESTEROL	GSTM4	1.16E-06	0.0123
CHOLESTEROL	CSGALNACT1	2.64E-05	0.0463
CHOLESTEROL	IRF7	0.000452	0.32
CHOLESTEROL	CPNE8	0.000319	0.273
CHOLESTEROL	TMEM258	9.53E-05	0.115
CHOLESTEROL	TAGLN	9.95E-05	0.115
CHOLESTEROL	SORT1	9.26E-05	0.115
CHOLESTEROL	APOE	1.59E-05	0.0558
CHOLESTEROL	TM6SF2	0.000329	0.261
CHOLESTEROL	TMEM31	0.000127	0.166
CHOLESTEROL	PSRC1	5.81E-09	0.000184
CHOLESTEROL	COPZ1	0.000292	0.289
CHOLESTEROL	GSTM2P1	0.000315	0.273
CHOLESTEROL	PSMA5	1.58E-05	0.0558
CHRONIC KIDNEY DISEASE	CXCL11	3.67E-05	0.283
CHRONIC KIDNEY DISEASE	AES	5.74E-05	0.296
CHRONIC KIDNEY DISEASE	FAM47E-STBD1	6.80E-05	0.283
CHRONIC KIDNEY DISEASE	NUPR1	8.39E-05	0.296
CHRONIC KIDNEY DISEASE	PPCS	0.000124	0.303
CHRONIC KIDNEY DISEASE	LPGAT1	0.00012	0.303
CHRONIC KIDNEY DISEASE	SFXN2	0.000116	0.303
CHRONIC KIDNEY DISEASE	RTF1	7.14E-05	0.283
CHRONIC KIDNEY DISEASE	ACBD3	1.45E-05	0.23
CHRONIC KIDNEY DISEASE	RSBN1L	5.63E-05	0.283
CHRONIC KIDNEY DISEASE	SLC30A4	6.04E-05	0.283
CHRONIC KIDNEY DISEASE	SLC28A2	1.88E-05	0.198
CHRONIC KIDNEY DISEASE	NPIPB6	3.34E-05	0.265
CHRONIC KIDNEY DISEASE	SPATA5L1	1.67E-06	0.0265
CHRONIC KIDNEY DISEASE	NLRP2	0.000139	0.314
COGNITIVE PERFORMANCE	BAK1	0.000141	0.239
COGNITIVE PERFORMANCE	ATRIP	4.58E-05	0.125
COGNITIVE PERFORMANCE	DOCK3	0.000121	0.229
COGNITIVE PERFORMANCE	TANK	0.000169	0.243
COGNITIVE PERFORMANCE	PRR22	0.000169	0.243
COGNITIVE PERFORMANCE	DNAJB2	0.000219	0.296
COGNITIVE PERFORMANCE	MST1	2.23E-05	0.0866
COGNITIVE PERFORMANCE	WDR6	0.000115	0.229
COGNITIVE PERFORMANCE	TUFM	6.60E-06	0.0606
COGNITIVE PERFORMANCE	NPIPB6	9.77E-06	0.0606
COGNITIVE PERFORMANCE	KCNIP2	3.79E-05	0.117
COGNITIVE PERFORMANCE	TNKS2	4.83E-05	0.125
COGNITIVE PERFORMANCE	APEH	1.77E-05	0.0866
COGNITIVE PERFORMANCE	RNF123	1.30E-06	0.0404
COGNITIVE PERFORMANCE	IRF2BP2	0.000172	0.243
COGNITIVE PERFORMANCE	KCNIP2-AS1	8.37E-05	0.185
CORONARY ARTERY DISEASE	MTAP	9.06E-06	0.139

phenotype	gene	p-value	Q-value
CORONARY ARTERY DISEASE	PSMA5	8.42E-07	0.0258
CORONARY ARTERY DISEASE	PSRC1	1.03E-05	0.323
CROHN DISEASE	AIRE	0.000193	0.242
CROHN DISEASE	ZNF300P1	5.54E-06	0.0566
CROHN DISEASE	NUPR1	1.02E-05	0.0311
CROHN DISEASE	PTGER4	1.07E-07	0.00326
CROHN DISEASE	LIPG	0.000213	0.258
CROHN DISEASE	SPINK4	5.89E-05	0.16
CROHN DISEASE	GSTCD	0.000271	0.306
CROHN DISEASE	SOHLH2	2.09E-05	0.0675
CROHN DISEASE	UBE2L3	0.000362	0.326
CROHN DISEASE	ZNF169	0.000185	0.242
CROHN DISEASE	SCAMP3	0.000241	0.262
CROHN DISEASE	NPIPB6	2.56E-06	0.0156
CROHN DISEASE	UBA7	2.16E-05	0.0506
CROHN DISEASE	CCDC116	0.000362	0.326
CROHN DISEASE	FGFR1OP	6.70E-06	0.0255
CROHN DISEASE	RNF123	1.06E-06	0.0108
CROHN DISEASE	RBM6	3.31E-05	0.0681
CROHN DISEASE	TUFM	0.000168	0.197
CROHN DISEASE	NFATC2IP	4.56E-05	0.0731
CROHN DISEASE	ADAM15	0.000105	0.18
CROHN DISEASE	CAMKV	1.20E-05	0.112
CROHN DISEASE	PRRC1	5.19E-05	0.0833
CROHN DISEASE	CLEC1B	0.000221	0.258
CROHN DISEASE	EIF3C	3.31E-05	0.112
CROHN DISEASE	ZPBP2	2.02E-05	0.0506
CROHN DISEASE	EDC3	7.67E-05	0.117
CROHN DISEASE	GPX1	0.00012	0.197
CROHN DISEASE	SEMA3B-AS1	7.78E-05	0.17
CROHN DISEASE	PKLR	8.47E-06	0.043
CROHN DISEASE	RNASET2	3.37E-05	0.0642
CROHN DISEASE	NKD1	3.70E-05	0.112
CROHN DISEASE	ASH1L	1.55E-05	0.0591
CROHN DISEASE	LYRM7	3.22E-05	0.112
CROHN DISEASE	P4HA2	2.98E-05	0.0681
CROHN DISEASE	IL1RL1	0.000291	0.306
CROHN DISEASE	HNRNPA1L2	0.000149	0.207
CROHN DISEASE	IL18R1	0.000322	0.316
CROHN DISEASE	IRGM	4.45E-05	0.231
CROHN DISEASE	HCN3	3.76E-05	0.0681
CROHN DISEASE	CUL2	9.54E-05	0.18
CROHN DISEASE	AMT	2.51E-05	0.112
CROHN DISEASE	SLC22A5	3.05E-05	0.0681
CROHN DISEASE	EGR2	2.24E-05	0.171
CROHN DISEASE	NPIPB9	0.000216	0.258
CROHN DISEASE	GBA	3.59E-05	0.0681
CROHN DISEASE	FAM69A	0.000338	0.312
CROHN DISEASE	FAM212A	7.25E-07	0.011
CROHN DISEASE	CYLD	5.56E-06	0.0566
CROHN DISEASE	C3orf62	3.13E-05	0.112
CROHN DISEASE	KTN1	0.000291	0.306
CROHN DISEASE	SLC22A4	9.92E-05	0.18
CROHN DISEASE	CARD6	6.31E-05	0.0874
CROHN DISEASE	IP6K1	3.27E-05	0.0642
CROHN DISEASE	DAP3	0.000331	0.312
DEPRESSIVE	TP53	6.93E-06	0.217
DIABETES	PYY2	5.02E-05	0.315

phenotype	gene	p-value	Q-value
DIABETES	CDKN2B	4.81E-05	0.315
DIABETES	FCHSD1	2.04E-06	0.0373
DIABETES	NINJ2	2.96E-05	0.31
DIABETES	KIF11	4.42E-06	0.14
DIASTOLIC BLOOD PRESSURE	CLIC4	1.95E-05	0.178
DIASTOLIC BLOOD PRESSURE	CSK	1.80E-05	0.178
DIASTOLIC BLOOD PRESSURE	OXER1	2.27E-05	0.178
DIASTOLIC BLOOD PRESSURE	LMAN1L	1.20E-05	0.178
EARLY GROWTH	NDUFB4	1.50E-05	0.232
EARLY GROWTH	LIN28B	1.34E-05	0.232
EDUYEARS	RNF123	1.83E-05	0.0568
EDUYEARS	MST1	7.11E-05	0.179
EDUYEARS	TANK	0.0002	0.274
EDUYEARS	NPIPB9	0.000104	0.179
EDUYEARS	PRMT5-AS1	0.000323	0.3
EDUYEARS	NPIPB6	1.61E-06	0.0167
EDUYEARS	IL27	0.000321	0.3
EDUYEARS	TUFM	1.32E-06	0.0167
EDUYEARS	CDK2AP1	0.000343	0.305
EDUYEARS	HCG11	0.000128	0.21
EDUYEARS	EPHA2	0.000356	0.307
EDUYEARS	DGUOK	0.000318	0.3
EDUYEARS	TNKS2	1.67E-05	0.0568
EDUYEARS	WDR6	8.58E-05	0.179
EDUYEARS	NUPR1	9.85E-05	0.179
EDUYEARS	DNAJB2	9.51E-05	0.179
EDUYEARS	PRR22	0.000203	0.274
EDUYEARS	SEMA6D	0.000227	0.284
EDUYEARS	IRF2BP2	9.98E-06	0.0443
EDUYEARS	KCNIP2	0.000298	0.3
EDUYEARS	APEH	0.000176	0.26
EDUYEARS	KCNIP2-AS1	0.000228	0.284
EXTRAVERSION	LIMD2	3.87E-05	0.208
EXTRAVERSION	CEP55	1.95E-05	0.208
EXTREME_BMI	ADCY3	4.18E-06	0.0644
EXTREME_BMI	NEGR1	3.20E-06	0.0644
EXTREME_HEIGHT	E2F1	0.00013	0.211
EXTREME_HEIGHT	MYBPC3	5.00E-05	0.13
EXTREME_HEIGHT	ADAMTSL3	0.000167	0.244
EXTREME_HEIGHT	NPR3	0.00021	0.294
EXTREME_HEIGHT	MYH7B	3.67E-05	0.114
EXTREME_HEIGHT	PUS1	0.000275	0.294
EXTREME_HEIGHT	KCTD18	3.20E-05	0.114
EXTREME_HEIGHT	AKTIP	0.000114	0.211
EXTREME_HEIGHT	PLCL1	0.000229	0.294
EXTREME_HEIGHT	VTA1	1.24E-05	0.0838
EXTREME_HEIGHT	RASA2	0.000167	0.244
EXTREME_HEIGHT	MAP3K3	1.34E-05	0.0838
EXTREME_HEIGHT	P4HA2	0.000328	0.296
EXTREME_HEIGHT	ZBTB38	0.000325	0.296
EXTREME_HEIGHT	AMZ1	8.78E-05	0.193
EXTREME_HEIGHT	FER1L4	2.77E-05	0.114
EXTREME_HEIGHT	PROCR	0.000336	0.296
EXTREME_HEIGHT	PXMP4	3.13E-05	0.114
EXTREME_HEIGHT	GDF5	4.21E-06	0.0838
EXTREME_HEIGHT	HIST1H3B	5.07E-05	0.13
EXTREME_HEIGHT	HIST1H2BF	0.000272	0.294
EXTREME_HEIGHT	CDK6	1.33E-05	0.0838

phenotype	gene	p-value	Q-value
EXTREME_HEIGHT	ASIP	0.000129	0.211
EXTREME_HEIGHT	DHX16	0.000259	0.294
EXTREME_HEIGHT	CEP250	5.94E-05	0.141
EXTREME_HEIGHT	C6orf106	0.000104	0.211
EXTREME_HEIGHT	HMGA2	0.000121	0.211
EXTREME_HEIGHT	SLC22A5	0.000298	0.296
FASTING GLUCOSE	IFT172	1.64E-07	0.00376
FASTING GLUCOSE	FNDC4	2.63E-06	0.0142
FASTING GLUCOSE	ZSCAN25	0.000124	0.33
FASTING GLUCOSE	RGS8	3.13E-06	0.0142
FASTING GLUCOSE	KHK	3.00E-06	0.0142
FASTING GLUCOSE	CRY2	9.33E-05	0.328
FASTING GLUCOSE	BBIP1	1.56E-05	0.198
FASTING GLUCOSE	G6PC2	5.31E-07	0.00681
FASTING GLUCOSE	P4HA2	2.76E-05	0.217
FASTING GLUCOSE	CGREF1	4.94E-07	0.00681
FASTING GLUCOSE	GPN1	3.03E-05	0.181
FASTING GLUCOSE	KRTCAP3	4.13E-05	0.187
FASTING GLUCOSE	SNX17	8.84E-05	0.302
FASTING INSULIN	STARD9	0.00196	0.318
FASTING INSULIN	NGEF	0.00159	0.318
FASTING INSULIN	NDUFS3	2.09E-07	0.00195
FASTING INSULIN	PCBP2	0.00174	0.318
FASTING INSULIN	NRSN1	0.00115	0.318
FASTING INSULIN	RASAL1	4.90E-05	0.0902
FASTING INSULIN	ZCCHC11	0.000659	0.267
FASTING INSULIN	ZNRF4	0.000993	0.318
FASTING INSULIN	VAC14	0.00115	0.318
FASTING INSULIN	DLGAP4	0.00212	0.327
FASTING INSULIN	CA10	0.00183	0.318
FASTING INSULIN	SYNDIG1	0.00163	0.318
FASTING INSULIN	BRINP2	0.00236	0.327
FASTING INSULIN	CACNG3	0.000652	0.267
FASTING INSULIN	FUNDC2	0.00242	0.331
FASTING INSULIN	SMDT1	0.000138	0.311
FASTING INSULIN	KCNJ4	0.0016	0.318
FASTING INSULIN	COL26A1	1.79E-05	0.281
FASTING INSULIN	PDZD4	0.000121	0.131
FASTING INSULIN	CDC25B	0.00231	0.327
FASTING INSULIN	FKRP	0.00164	0.318
FASTING INSULIN	TPT1-AS1	0.00233	0.327
FASTING INSULIN	URM1	0.00167	0.318
FASTING INSULIN	ITGB4	0.00146	0.318
FASTING INSULIN	UBR4	0.000519	0.253
FASTING INSULIN	CLPTM1	0.00144	0.318
FASTING INSULIN	LPCAT4	7.36E-05	0.126
FASTING INSULIN	TLR5	0.000176	0.15
FASTING INSULIN	DDB1	0.00196	0.318
FASTING INSULIN	EXOC8	0.00125	0.318
FASTING INSULIN	CELF6	0.00158	0.318
FASTING INSULIN	FAM180B	1.98E-07	0.00195
FASTING INSULIN	HS3ST2	0.00155	0.318
FASTING INSULIN	SERPINH1	0.00207	0.324
FASTING INSULIN	DUS1L	0.000317	0.181
FASTING INSULIN	ASIC2	0.00235	0.327
FASTING INSULIN	PPP6R1	0.00184	0.318
FASTING INSULIN	C7orf26	0.000495	0.247
FASTING INSULIN	NUDC	2.59E-06	0.0124

phenotype	gene	p-value	Q-value
FASTING INSULIN	PRPF8	0.000233	0.154
FASTING INSULIN	MTFR1L	0.00217	0.327
FASTING INSULIN	PTDSS2	0.000198	0.154
FASTING INSULIN	SNCB	0.00157	0.318
FASTING INSULIN	ALKBH7	1.15E-06	0.00811
FASTING INSULIN	TUBB4A	0.000992	0.318
FASTING INSULIN	CDC37	0.00185	0.318
FASTING INSULIN	DYNC2LI1	0.00232	0.327
FASTING INSULIN	GAA	0.00237	0.327
FASTING INSULIN	TMEM132A	0.00228	0.327
FASTING INSULIN	WBP2	0.00158	0.318
FASTING INSULIN	PPP1R15A	0.00127	0.318
FASTING INSULIN	SSRP1	0.00171	0.318
FASTING INSULIN	IGSF21	0.00244	0.332
FASTING INSULIN	C1QTNF4	4.05E-05	0.128
FASTING INSULIN	DPM2	8.02E-05	0.128
FASTING INSULIN	ATP6V0B	0.00219	0.327
FASTING INSULIN	GPX4	1.36E-06	0.00811
FASTING INSULIN	CHRM1	0.00135	0.318
FASTING INSULIN	ICAM5	0.000229	0.154
FASTING INSULIN	SHISA4	2.94E-05	0.0704
FASTING INSULIN	NR1H2	0.000822	0.296
FASTING INSULIN	SIPA1L3	0.00154	0.318
FASTING INSULIN	C19orf70	0.00187	0.318
FASTING INSULIN	PRKCB	0.000621	0.267
FASTING INSULIN	STAT1	0.00122	0.318
FASTING INSULIN	WASH2P	0.00131	0.318
FASTING INSULIN	SULT4A1	0.00216	0.327
FASTING INSULIN	TFPI	0.00186	0.318
FASTING INSULIN	FNBP4	5.09E-07	0.00229
FASTING INSULIN	SLC25A43	0.00189	0.318
FASTING INSULIN	MYBPC3	0.000102	0.247
FASTING INSULIN	FGF1	0.00166	0.318
FASTING INSULIN	ARHGAP1	2.96E-07	0.00195
FASTING INSULIN	AGBL2	6.83E-05	0.179
FASTING INSULIN	SV2A	0.00153	0.318
FASTING INSULIN	CKMT1B	6.21E-07	0.00743
FASTING INSULIN	IQSEC3	0.00142	0.318
FASTING INSULIN	MAN1B1	0.000279	0.167
FASTING INSULIN	CDK16	0.00015	0.139
FASTING INSULIN	KRT17	0.00179	0.318
FASTING INSULIN	SCGB1D2	0.00196	0.318
FASTING INSULIN	BAG6	9.40E-06	0.0363
FASTING INSULIN	MOAP1	0.00193	0.318
FASTING INSULIN	SGTA	0.00234	0.327
FASTING INSULIN	RPL12	0.00236	0.327
FASTING INSULIN	HIGD1B	0.00175	0.318
FASTING INSULIN	P4HB	0.000119	0.131
FASTING INSULIN	MICAL1	0.00204	0.32
FASTING INSULIN	IFT172	1.06E-05	0.152
FASTING INSULIN	SENP3	0.000745	0.282
FASTING INSULIN	ARAP1	3.17E-05	0.111
FASTING INSULIN	SDF4	0.000712	0.279
FASTING INSULIN	ADAMTS19	0.000361	0.196
FASTING INSULIN	GBP4	7.12E-07	0.0225
FASTING INSULIN	SLC39A13	2.89E-07	0.00195
FASTING INSULIN	TNFAIP8L3	0.00026	0.16
FASTING INSULIN	ISOC2	0.00107	0.318

phenotype	gene	p-value	Q-value
FASTING INSULIN	MLF2	0.00185	0.318
FASTING INSULIN	CST3	0.00149	0.318
FASTING INSULIN	TECR	0.000245	0.154
FASTING INSULIN	JPH4	0.00223	0.327
FASTING INSULIN	TUBGCP2	0.00107	0.318
FASTING INSULIN	LAMB3	3.25E-05	0.0708
FASTING INSULIN	PTPN23	0.000116	0.131
FASTING INSULIN	CHRM5	0.000581	0.263
FASTING INSULIN	NOP16	0.00135	0.318
FASTING INSULIN	TXNL4B	0.00191	0.318
FASTING INSULIN	ARSK	0.0023	0.327
FASTING INSULIN	MAN2C1	0.000583	0.263
FASTING INSULIN	TMEM251	0.00149	0.318
FASTING INSULIN	PRPS1L1	0.00135	0.318
FASTING INSULIN	PYGB	0.00165	0.318
FASTING INSULIN	RBM23	0.00166	0.318
FASTING INSULIN	PPP1CA	0.00013	0.131
FASTING INSULIN	OLAH	0.00202	0.32
FASTING INSULIN	GABBR2	9.64E-05	0.128
FASTING INSULIN	VARS2	0.000142	0.136
FASTING INSULIN	CKAP2L	0.00123	0.318
FASTING INSULIN	LHX2	1.21E-05	0.0363
FASTING INSULIN	MAPK8IP1	9.25E-05	0.128
FASTING INSULIN	POLR2A	0.0011	0.318
FASTING INSULIN	NEFL	0.00198	0.318
FASTING INSULIN	ZNF280B	0.00187	0.318
FASTING INSULIN	INA	0.00172	0.318
FASTING INSULIN	RALYL	0.000446	0.227
FASTING INSULIN	CHAC1	0.00177	0.318
FASTING INSULIN	SAMD8	0.000411	0.214
FASTING INSULIN	SERPINI2	0.00209	0.325
FASTING INSULIN	PPP2R5B	0.000568	0.263
FASTING INSULIN	STMN4	0.00144	0.318
FASTING INSULIN	INO80E	0.000243	0.154
FASTING INSULIN	CCNY	0.00221	0.327
FASTING INSULIN	PRDM2	0.000671	0.268
FASTING INSULIN	RPL4	0.000192	0.154
FASTING INSULIN	SYT4	1.16E-05	0.0363
FASTING INSULIN	MTCH2	3.71E-07	0.00195
FASTING INSULIN	ITPK1	9.58E-05	0.128
FASTING INSULIN	NRGN	0.000216	0.154
FASTING INSULIN	TPD52L2	0.000888	0.313
FASTING INSULIN	CYLC2	0.00175	0.318
FASTING INSULIN	CCND3	0.00064	0.267
FASTING INSULIN	LY6E	0.00177	0.318
FASTING INSULIN	BLOC1S2	0.000131	0.131
FASTING INSULIN	ELMO1	0.000412	0.214
FASTING INSULIN	SOD1	0.0019	0.318
FASTING INSULIN	DDN	0.000747	0.282
FASTING INSULIN	RXRB	0.00126	0.318
FASTING INSULIN	MAST3	0.00162	0.318
FASTING INSULIN	C12orf57	0.000548	0.262
FASTING INSULIN	SSR2	3.66E-05	0.073
FASTING INSULIN	ASXL3	0.00139	0.318
FASTING INSULIN	GANAB	0.000753	0.282
FASTING INSULIN	MFSD12	0.000346	0.193
FASTING INSULIN	MCEE	0.00102	0.318
FASTING INSULIN	TAF15	0.00142	0.318

phenotype	gene	p-value	Q-value
FASTING INSULIN	LRRD1	0.00159	0.318
FASTING INSULIN	FOXS1	1.27E-05	0.152
FASTING INSULIN	PDE2A	3.34E-07	0.00195
FASTING INSULIN	TAGLN3	0.000992	0.318
FASTING INSULIN	MCHR1	0.00144	0.318
FASTING INSULIN	PPFIA3	0.000108	0.131
FASTING INSULIN	MON1A	3.14E-06	0.0986
FASTING INSULIN	SERF2	0.00135	0.318
FASTING INSULIN	MAP7D1	0.000158	0.14
FASTING INSULIN	APEH	0.00128	0.318
FIVE PSYCHIATRIC DISEASES	ITIH4	3.16E-07	0.00338
FIVE PSYCHIATRIC DISEASES	MUSTN1	0.000157	0.289
FIVE PSYCHIATRIC DISEASES	GNL3	0.000194	0.294
FIVE PSYCHIATRIC DISEASES	PBRM1	2.02E-05	0.0703
FIVE PSYCHIATRIC DISEASES	FAM219A	0.000145	0.284
FIVE PSYCHIATRIC DISEASES	SFXN2	1.62E-05	0.0703
FIVE PSYCHIATRIC DISEASES	NEK4	0.00017	0.294
FIVE PSYCHIATRIC DISEASES	ARL3	4.49E-06	0.0351
FIVE PSYCHIATRIC DISEASES	GLT8D1	7.02E-05	0.165
FIVE PSYCHIATRIC DISEASES	BCAP29	3.07E-05	0.0873
FIVE PSYCHIATRIC DISEASES	NT5C2	3.24E-07	0.00338
FIVE PSYCHIATRIC DISEASES	SPCS1	2.42E-05	0.0756
HEAD CIRCUMFERENCE	SBNO1	1.88E-05	0.0738
HEAD CIRCUMFERENCE	OGFOD2	2.22E-06	0.0348
HEAD CIRCUMFERENCE	DCAKD	1.55E-05	0.0738
HEAD CIRCUMFERENCE	PXMP4	1.34E-05	0.0738
HEAD CIRCUMFERENCE	CDK2AP1	0.000104	0.272
HEAD CIRCUMFERENCE	ADORA3	1.65E-05	0.0738
HEAD CIRCUMFERENCE	ZNFX1	3.07E-07	0.00963
HEAD CIRCUMFERENCE	SLC45A3	5.43E-06	0.0568
HEAD CIRCUMFERENCE	ZNF32	1.80E-05	0.0738
HEIGHT	ANKRD24	0.00214	0.292
HEIGHT	ST3GAL6-AS1	0.00132	0.26
HEIGHT	S100A3	0.000938	0.281
HEIGHT	NCOA6	0.0025	0.3
HEIGHT	USP12	0.00206	0.29
HEIGHT	ERVK13-1	9.28E-05	0.122
HEIGHT	CBFA2T2	0.000612	0.226
HEIGHT	HEMK1	0.000204	0.137
HEIGHT	ARHGAP35	0.00213	0.292
HEIGHT	ZNF277	0.0033	0.329
HEIGHT	PXDN	0.000382	0.177
HEIGHT	MEOX1	0.00261	0.301
HEIGHT	MAP3K3	3.66E-05	0.104
HEIGHT	DCDC5	0.00197	0.289
HEIGHT	NFATC4	0.00107	0.281
HEIGHT	ADCY4	0.00211	0.291
HEIGHT	RAB24	0.00248	0.3
HEIGHT	PTPRJ	0.000241	0.139
HEIGHT	PIEZO1	0.00327	0.328
HEIGHT	OGFOD2	0.00204	0.29
HEIGHT	OMG	0.000313	0.154
HEIGHT	ASB3	0.00305	0.317
HEIGHT	CDC40	0.000934	0.281
HEIGHT	TACO1	0.000376	0.176
HEIGHT	USP25	0.00151	0.289
HEIGHT	P4HA2	0.00111	0.281
HEIGHT	SLC11A1	0.000411	0.185

phenotype	gene	p-value	Q-value
HEIGHT	HIST1H2BD	0.000257	0.139
HEIGHT	PCTP	0.00156	0.294
HEIGHT	NME6	0.00196	0.289
HEIGHT	TMX4	0.00298	0.317
HEIGHT	HHIP-AS1	0.000326	0.168
HEIGHT	OMP	0.00145	0.263
HEIGHT	RPL23	0.0017	0.271
HEIGHT	LGALS4	0.00104	0.281
HEIGHT	RASA2	6.53E-05	0.143
HEIGHT	CENPO	9.82E-05	0.148
HEIGHT	C14orf39	0.0021	0.291
HEIGHT	TBX2	0.00214	0.324
HEIGHT	SPATA3-AS1	0.000846	0.279
HEIGHT	CA11	0.00186	0.31
HEIGHT	FCER2	0.000888	0.281
HEIGHT	LINC00928	0.00109	0.281
HEIGHT	ZBTB4	0.00231	0.295
HEIGHT	OR2T29	0.000841	0.23
HEIGHT	CAMSAP2	0.00154	0.263
HEIGHT	TTLL6	0.00121	0.258
HEIGHT	IHH	0.000186	0.137
HEIGHT	MAP1LC3A	2.66E-05	0.0928
HEIGHT	CDH16	0.0003	0.157
HEIGHT	CDK6	2.32E-05	0.0669
HEIGHT	PLCL1	0.000173	0.21
HEIGHT	CD99L2	0.000107	0.282
HEIGHT	PNPLA1	0.000869	0.232
HEIGHT	BCKDHB	0.002	0.29
HEIGHT	EFEMP1	0.0018	0.308
HEIGHT	C6orf106	1.11E-06	0.0133
HEIGHT	PNPLA8	0.00236	0.296
HEIGHT	PPARD	1.73E-05	0.0907
HEIGHT	C1QTNF4	0.00162	0.267
HEIGHT	AMZ1	0.000148	0.113
HEIGHT	DHX16	0.00121	0.258
HEIGHT	NPR3	4.72E-05	0.104
HEIGHT	EBPL	0.000926	0.239
HEIGHT	DDX19A	0.00238	0.298
HEIGHT	E2F1	1.18E-05	0.0489
HEIGHT	PTPMT1	0.00296	0.317
HEIGHT	APOBEC3C	0.001	0.281
HEIGHT	PIGU	0.000223	0.138
HEIGHT	IGHMBP2	0.00319	0.321
HEIGHT	GPR182	0.00301	0.317
HEIGHT	HMGA2	9.32E-05	0.148
HEIGHT	FRYL	0.00222	0.294
HEIGHT	ZBTB24	0.00095	0.241
HEIGHT	DNAJC27-AS1	0.00163	0.297
HEIGHT	ITIH4	0.000396	0.181
HEIGHT	TSEN15	8.07E-05	0.148
HEIGHT	SMG9	0.00289	0.315
HEIGHT	GRINA	0.00071	0.217
HEIGHT	KIR3DX1	0.00168	0.301
HEIGHT	CALCOCO1	0.00245	0.3
HEIGHT	ACTL10	0.00196	0.289
HEIGHT	SUGP2	0.00231	0.295
HEIGHT	PFN4	0.000305	0.293
HEIGHT	RSPH9	0.00233	0.295

phenotype	gene	p-value	Q-value
HEIGHT	MST1P2	0.00139	0.289
HEIGHT	FAM180B	9.26E-05	0.122
HEIGHT	SNORA75	0.000542	0.196
HEIGHT	U2AF1L4	0.00165	0.297
HEIGHT	TTTY13	0.000493	0.192
HEIGHT	KCNH2	0.00048	0.192
HEIGHT	IDH3A	0.00159	0.267
HEIGHT	MED28	0.000338	0.296
HEIGHT	BAG2	0.00136	0.261
HEIGHT	ATAD1	0.000793	0.226
HEIGHT	GFER	0.00248	0.3
HEIGHT	TRPC4AP	0.000367	0.298
HEIGHT	SNX32	0.000169	0.137
HEIGHT	CEP250	8.68E-05	0.102
HEIGHT	KCNK9	0.000405	0.3
HEIGHT	DENND6A	4.90E-05	0.104
HEIGHT	OR7E156P	0.00267	0.304
HEIGHT	PRRC1	0.000296	0.153
HEIGHT	GDF5	4.39E-07	0.0139
HEIGHT	NPY2R	0.000633	0.213
HEIGHT	SBNO1	0.00152	0.263
HEIGHT	JKAMP	0.00115	0.281
HEIGHT	SOCS5	0.00154	0.263
HEIGHT	CCDC126	0.000438	0.189
HEIGHT	HIST1H2BF	5.64E-05	0.224
HEIGHT	LINC00471	0.00118	0.257
HEIGHT	PGP	0.000534	0.196
HEIGHT	MAFF	0.00139	0.261
HEIGHT	ASIP	0.000351	0.172
HEIGHT	KCNG1	0.00137	0.289
HEIGHT	CDK2AP1	0.000701	0.241
HEIGHT	NEK3	0.00181	0.308
HEIGHT	C2orf69	0.00118	0.281
HEIGHT	WDR6	0.000447	0.192
HEIGHT	ATRNL1	0.003	0.317
HEIGHT	SMIM23	0.0016	0.294
HEIGHT	TTC23	0.00142	0.263
HEIGHT	PEX1	0.000818	0.226
HEIGHT	RAPSN	0.00119	0.283
HEIGHT	VTA1	1.40E-05	0.0492
HEIGHT	CD8B	0.00113	0.257
HEIGHT	STAG2	0.00257	0.301
HEIGHT	SERAC1	0.00294	0.317
HEIGHT	AP5B1	0.000489	0.192
HEIGHT	RIOK3	0.00292	0.315
HEIGHT	PXMP4	0.00015	0.113
HEIGHT	FASTKD1	0.00155	0.293
HEIGHT	ITIH4-AS1	0.00225	0.294
HEIGHT	SLC25A45	0.00195	0.31
HEIGHT	SLC39A13	0.000704	0.217
HEIGHT	SCUBE3	0.000146	0.186
HEIGHT	BAAT	0.00226	0.294
HEIGHT	CCDC91	8.71E-05	0.148
HEIGHT	HAPLN4	0.00244	0.3
HEIGHT	KPNA6	0.000809	0.226
HEIGHT	NCF2	0.00221	0.294
HEIGHT	ITPR3	0.000456	0.193
HEIGHT	PCDH19	0.00138	0.261

phenotype	gene	p-value	Q-value
HEIGHT	BIN2	0.00262	0.301
HEIGHT	LOXL1-AS1	0.00148	0.263
HEIGHT	CDKN1A	0.0014	0.289
HEIGHT	RNASE4	0.00114	0.281
HEIGHT	SERPINF2	0.00206	0.29
HEIGHT	GGT7	0.000908	0.281
HEIGHT	KRT6C	0.000134	0.112
HEIGHT	LGALS8	1.14E-06	0.0357
HEIGHT	PCIF1	0.00231	0.295
HEIGHT	THUMPD1	0.000232	0.139
HEIGHT	TM7SF3	0.00159	0.267
HEIGHT	LGALS3	0.00222	0.294
HEIGHT	CNNM2	0.00098	0.244
HEIGHT	HAS2	0.00335	0.332
HEIGHT	LYSMD1	0.000871	0.232
HEIGHT	GNA12	0.00127	0.289
HEIGHT	BCS1L	0.00193	0.289
HEIGHT	SLC9A3R2	0.000214	0.137
HEIGHT	FER1L4	5.79E-07	0.0182
HEIGHT	ZNFX1	0.000199	0.13
HEIGHT	STAG1	0.00136	0.261
HEIGHT	TCP11	7.76E-05	0.102
HEIGHT	AGBL2	0.00188	0.286
HEIGHT	LHFPL3-AS1	0.00147	0.289
HEIGHT	ZNF688	0.000864	0.232
HEIGHT	STXBP3	0.000251	0.266
HEIGHT	FUT4	0.00108	0.281
HEIGHT	DVL3	0.00172	0.271
HEIGHT	FRRS1	0.00271	0.304
HEIGHT	CHRNB1	0.000752	0.222
HEIGHT	RGS11	0.00111	0.257
HEIGHT	DMRTA2	0.00108	0.281
HEIGHT	RDH8	0.00112	0.281
HEIGHT	CLPSL1	0.00197	0.289
HEIGHT	ABCE1	2.91E-05	0.148
HEIGHT	C17orf53	0.00127	0.26
HEIGHT	SMOC2	0.00259	0.301
HEIGHT	KCTD18	8.77E-05	0.102
HEIGHT	NCL	0.00258	0.301
HEIGHT	HIST1H3B	9.01E-05	0.148
HEIGHT	PNKD	0.000816	0.226
HEIGHT	G6PC3	0.000104	0.127
HEIGHT	ADCY3	1.66E-06	0.0181
HEIGHT	CAMKMT	0.000649	0.213
HEIGHT	PIGV	0.00139	0.261
HEIGHT	ZCCHC24	0.00162	0.267
HEIGHT	USP26	0.00136	0.289
HEIGHT	ACTR1A	0.00206	0.29
HEIGHT	TTTY5	0.00075	0.253
HEIGHT	ARFGEF2	0.0018	0.282
HEIGHT	NDUFAF1	0.00284	0.312
HEIGHT	DPAGT1	0.00226	0.294
HEIGHT	PROCR	1.23E-05	0.13
HEIGHT	TRPM7	0.00135	0.261
HEIGHT	CRNN	0.000636	0.232
HEIGHT	AKTIP	0.000334	0.162
HEIGHT	GABRR1	0.00216	0.292
HEIGHT	LINC00933	0.000289	0.286

phenotype	gene	p-value	Q-value
HEIGHT	GPR152	1.03E-05	0.0489
HEIGHT	CA5BP1	0.0017	0.302
HEIGHT	UBQLN2	0.000202	0.137
HEIGHT	TRIM8	0.000251	0.139
HEIGHT	PHF20	0.000517	0.201
HEIGHT	NCOA1	0.00149	0.289
HEIGHT	NDUFS3	0.00127	0.26
HEIGHT	GOLGA6L5P	0.00194	0.31
HEIGHT	RAB9A	0.00191	0.31
HEIGHT	ZFAS1	6.63E-05	0.11
HEIGHT	SNX12	0.00241	0.299
HEIGHT	NFAT5	0.0011	0.257
HEIGHT	BTN2A2	9.12E-05	0.122
HEIGHT	HCN2	0.00305	0.317
HEIGHT	CCDC177	0.00107	0.256
HEIGHT	HIST1H4E	0.000255	0.149
HEIGHT	OR1N2	0.000405	0.3
HEIGHT	CDC42SE2	0.000214	0.137
HEIGHT	DCPS	0.00107	0.256
HEIGHT	CPNE1	2.62E-05	0.0928
HEIGHT	RAPGEF6	0.000512	0.196
HEIGHT	ARF6	0.0015	0.263
HEIGHT	ANG	0.00192	0.31
HEIGHT	DCAF16	0.00027	0.151
HEIGHT	NKAIN1	0.00152	0.289
HEIGHT	OXCT1	0.00136	0.261
HEIGHT	PLEC	0.00301	0.317
HEIGHT	RAB33A	0.00215	0.292
HEIGHT	PADI2	0.00132	0.26
HEIGHT	MYBPC3	0.000242	0.148
HEIGHT	SNF8	0.00152	0.263
HEIGHT	UBL5	0.0021	0.291
HEIGHT	PPM1M	0.00303	0.317
HEIGHT	RBMXL1	0.000527	0.196
HEIGHT	SLC22A4	5.03E-05	0.104
HEIGHT	ZBTB38	2.11E-05	0.148
HEIGHT	RABEP1	0.00283	0.312
HEIGHT	SURF4	0.00159	0.267
HEIGHT	PRRT3	2.34E-05	0.0928
HEIGHT	ERGIC3	1.27E-06	0.0133
HEIGHT	CNRIP1	0.00149	0.263
HEIGHT	HIST1H2AC	5.57E-05	0.104
HEIGHT	RAD9A	0.000201	0.137
HEIGHT	ZNF76	0.00149	0.289
HEIGHT	DLEU1	0.000163	0.137
HEIGHT	C12orf65	0.000896	0.281
HEIGHT	PSMC5	0.00149	0.289
HEIGHT	MS4A12	0.00133	0.289
HEIGHT	KRTAP19-3	0.00201	0.29
HEIGHT	RCBTB1	0.00266	0.304
HEIGHT	SPIN1	0.00226	0.294
HEIGHT	FUBP1	5.63E-05	0.104
HEIGHT	HCG11	0.00149	0.263
HEIGHT	DBF4B	0.000126	0.111
HEIGHT	FAM208A	0.000755	0.222
HEIGHT	PUS1	0.000258	0.149
HEIGHT	MYH7B	2.68E-06	0.0425
HEIGHT	MOB4	0.00207	0.29

phenotype	gene	p-value	Q-value
HEIGHT	ADAMTSL3	2.85E-06	0.0181
HEIGHT	NMT1	0.000648	0.213
HEIGHT	HIST1H4B	1.37E-05	0.086
HEIGHT	CWH43	0.00208	0.29
HEIGHT	FASLG	0.00212	0.323
HEIGHT	ZRANB1	0.00291	0.315
HEIGHT	LRBA	0.00248	0.3
HEIGHT	GOLGA2P7	0.000993	0.244
HEIGHT	RASA3	0.00139	0.289
HEIGHT	UHRF1BP1	0.00112	0.257
HEIGHT	IL20RB	0.00302	0.317
HEIGHT	JMJD4	6.82E-05	0.143
HEIGHT	EIF1AD	0.000222	0.138
HEIGHT	EIF6	1.87E-05	0.0585
HEIGHT	KATNBL1	0.000526	0.201
HEIGHT	CYSTM1	0.00318	0.321
HEIGHT	RNF25	0.00195	0.289
HEIGHT	HIST1H1C	0.000329	0.296
HEIGHT	SPAG4	0.00171	0.302
HEIGHT	BICD2	0.00143	0.289
HEIGHT	PRLH	0.00151	0.289
HEIGHT	DEF6	6.36E-05	0.101
HEIGHT	RYBP	0.000628	0.213
HEIGHT	INE1	0.000261	0.149
HEIGHT	ARL6IP4	0.00126	0.26
HEIGHT	KCND2	4.89E-05	0.104
HEIGHT	SLC35G6	0.00187	0.286
HEIGHT	HHIP	6.79E-05	0.101
HEIGHT	SLC22A5	0.000734	0.22
HEIGHT	PRKAG3	0.00318	0.321
HEIGHT	RAB40A	0.00164	0.297
HEIGHT	CTPS2	0.00299	0.317
HEIGHT	PRKXP1	0.00118	0.257
HEIGHT	KDM1A	0.000886	0.232
HEIGHT	FANCE	0.00281	0.311
HEIGHT	TMEM82	0.00208	0.323
HIGH-DENSITY LIPOPROTEIN	AFF1	0.000177	0.18
HIGH-DENSITY LIPOPROTEIN	MYBPC3	4.39E-07	0.0029
HIGH-DENSITY LIPOPROTEIN	CSGALNACT1	0.000203	0.163
HIGH-DENSITY LIPOPROTEIN	IRF7	4.40E-05	0.0696
HIGH-DENSITY LIPOPROTEIN	SLC39A13	5.78E-05	0.0695
HIGH-DENSITY LIPOPROTEIN	MT1X	0.000372	0.259
HIGH-DENSITY LIPOPROTEIN	CTCF	0.000427	0.284
HIGH-DENSITY LIPOPROTEIN	MEP1A	0.000118	0.156
HIGH-DENSITY LIPOPROTEIN	PSKH1	6.14E-06	0.0388
HIGH-DENSITY LIPOPROTEIN	ERBB2	1.28E-05	0.0463
HIGH-DENSITY LIPOPROTEIN	REEP1	0.00013	0.158
HIGH-DENSITY LIPOPROTEIN	PCSK7	0.000183	0.154
HIGH-DENSITY LIPOPROTEIN	LRP4	3.13E-05	0.0466
HIGH-DENSITY LIPOPROTEIN	CETP	1.71E-05	0.0539
HIGH-DENSITY LIPOPROTEIN	PTPMT1	5.85E-06	0.0166
HIGH-DENSITY LIPOPROTEIN	C1QTNF4	4.69E-07	0.0029
HIGH-DENSITY LIPOPROTEIN	APH1B	0.000269	0.239
HIGH-DENSITY LIPOPROTEIN	RPL17	0.00017	0.148
HIGH-DENSITY LIPOPROTEIN	RAPSN	3.78E-05	0.0513
HIGH-DENSITY LIPOPROTEIN	SPIB	0.000154	0.18
HIGH-DENSITY LIPOPROTEIN	GFOD2	5.84E-06	0.0166
HIGH-DENSITY LIPOPROTEIN	UBE3B	0.000184	0.182

phenotype	gene	p-value	Q-value
HIGH-DENSITY LIPOPROTEIN	PARD6A	1.96E-05	0.0361
HIGH-DENSITY LIPOPROTEIN	DUS2	3.08E-05	0.0628
HIGH-DENSITY LIPOPROTEIN	SSH1	0.000331	0.261
HIGH-DENSITY LIPOPROTEIN	DPEP3	1.66E-06	0.0263
HIGH-DENSITY LIPOPROTEIN	NDUFS3	1.56E-06	0.00696
HIGH-DENSITY LIPOPROTEIN	ZNF408	1.78E-05	0.0361
HIGH-DENSITY LIPOPROTEIN	C10orf67	0.000191	0.157
HIGH-DENSITY LIPOPROTEIN	NR1H3	9.31E-08	0.0029
HIGH-DENSITY LIPOPROTEIN	RNF214	0.000105	0.0968
HIGH-DENSITY LIPOPROTEIN	ACP2	1.62E-07	0.00512
HIGH-DENSITY LIPOPROTEIN	CACNB1	2.39E-05	0.0404
HIGH-DENSITY LIPOPROTEIN	LCAT	2.91E-06	0.0307
HIGH-DENSITY LIPOPROTEIN	TSNAXIP1	8.04E-05	0.084
HIGH-DENSITY LIPOPROTEIN	TCEANC2	2.15E-05	0.0567
HIGH-DENSITY LIPOPROTEIN	CCDC92	2.01E-05	0.0567
HIGH-DENSITY LIPOPROTEIN	PSMA5	8.06E-05	0.084
HIGH-DENSITY LIPOPROTEIN	NFATC3	0.000212	0.163
HIGH-DENSITY LIPOPROTEIN	WDR17	6.87E-05	0.0795
HIGH-DENSITY LIPOPROTEIN	ESRP2	5.21E-07	0.0029
HIGH-DENSITY LIPOPROTEIN	CKAP5	4.98E-07	0.0029
HIGH-DENSITY LIPOPROTEIN	PRKAR1B	0.000167	0.18
HIGH-DENSITY LIPOPROTEIN	PSMD3	0.000452	0.293
HIGH-DENSITY LIPOPROTEIN	C11orf49	1.11E-05	0.0289
HIGH-DENSITY LIPOPROTEIN	HPS3	0.000284	0.239
HIGH-DENSITY LIPOPROTEIN	DMRTC1	3.21E-05	0.0628
HIGH-DENSITY LIPOPROTEIN	TAGLN	0.000208	0.194
HIGH-DENSITY LIPOPROTEIN	FAM180B	4.67E-06	0.0369
HIGH-DENSITY LIPOPROTEIN	C6orf106	9.75E-05	0.0923
HIGH-DENSITY LIPOPROTEIN	ACD	1.21E-05	0.0463
HIGH-DENSITY LIPOPROTEIN	AGBL2	2.90E-05	0.0454
HIGH-DENSITY LIPOPROTEIN	RBM6	0.000223	0.166
HIP	NEGR1	3.31E-06	0.0116
HIP	FTO	7.56E-05	0.234
HIP	KCNJ13	0.000165	0.261
HIP	C6orf106	4.86E-05	0.0848
HIP	TMEM219	9.19E-05	0.144
HIP	GDF5	1.56E-05	0.0967
HIP	BPIFB1	0.00022	0.288
HIP	RBL2	8.24E-06	0.0199
HIP	SH2B1	3.13E-05	0.161
HIP	VWA7	0.000202	0.253
HIP	HCN4	4.39E-05	0.298
HIP	FER1L4	6.69E-07	0.00517
HIP	SPN	6.52E-05	0.136
HIP	SBK1	0.000161	0.282
HIP	HIST1H2BF	4.72E-05	0.0848
HIP	LRP8	0.000131	0.211
HIP	FUBP1	1.06E-08	0.000334
HIP	NPIPB9	0.000107	0.255
HIP	IL27	2.43E-05	0.0586
HIP	NFATC2IP	4.48E-06	0.0156
HIP	ACP1	0.000237	0.306
HIP	HIST1H4B	0.000143	0.204
HIP	TFAP2B	4.46E-06	0.0127
HIP	NPIPB8	0.000159	0.211
HIP	LGR4	1.53E-05	0.0344
HIP	NPIPB6	6.28E-08	0.000648
HIP	EIF3CL	1.43E-06	0.00697

phenotype	gene	p-value	Q-value
HIP	RABEP2	1.42E-05	0.039
HIP	ZBTB38	6.28E-05	0.298
HIP	RYBP	6.84E-06	0.0214
HIP	TUFM	3.17E-06	0.0111
HIP	PSG7	0.00025	0.328
HIP	FKBP1B	0.000207	0.269
HIP	NUPR1	5.93E-08	0.000648
HIP ADJUSTED BMI	ABCB9	1.75E-05	0.0685
HIP ADJUSTED BMI	KRT6C	0.000305	0.239
HIP ADJUSTED BMI	ADAMTSL3	1.22E-05	0.0579
HIP ADJUSTED BMI	EPB41L1	2.50E-05	0.117
HIP ADJUSTED BMI	CRNN	0.000591	0.284
HIP ADJUSTED BMI	CROCC	3.53E-05	0.0993
HIP ADJUSTED BMI	AKTIP	0.000373	0.268
HIP ADJUSTED BMI	SLC38A2	0.000172	0.279
HIP ADJUSTED BMI	RAB43	0.000449	0.281
HIP ADJUSTED BMI	CDK6	1.53E-05	0.0776
HIP ADJUSTED BMI	SNX32	7.24E-05	0.121
HIP ADJUSTED BMI	HIST1H2AC	5.37E-06	0.0389
HIP ADJUSTED BMI	THRA	0.000375	0.268
HIP ADJUSTED BMI	SUGP2	0.000112	0.153
HIP ADJUSTED BMI	FBXL13	1.02E-05	0.287
HIP ADJUSTED BMI	MAP3K3	0.000115	0.158
HIP ADJUSTED BMI	BRI3	0.000718	0.293
HIP ADJUSTED BMI	HIST1H4E	0.000676	0.29
HIP ADJUSTED BMI	PROCR	0.00014	0.175
HIP ADJUSTED BMI	ORMDL3	4.99E-05	0.105
HIP ADJUSTED BMI	SLC25A45	0.000232	0.316
HIP ADJUSTED BMI	PLCL1	8.93E-05	0.155
HIP ADJUSTED BMI	CDH16	0.00038	0.27
HIP ADJUSTED BMI	HHIP	2.58E-05	0.0841
HIP ADJUSTED BMI	AP5B1	0.000682	0.29
HIP ADJUSTED BMI	GPR152	3.50E-06	0.0315
HIP ADJUSTED BMI	HIST1H4B	7.24E-06	0.0379
HIP ADJUSTED BMI	CDK2AP1	0.000142	0.247
HIP ADJUSTED BMI	HHIP-AS1	4.75E-05	0.0994
HIP ADJUSTED BMI	GDF5	2.77E-06	0.0286
HIP ADJUSTED BMI	KLF14	0.000206	0.281
HIP ADJUSTED BMI	ABCE1	0.000248	0.226
HIP ADJUSTED BMI	LGI4	9.40E-05	0.155
HIP ADJUSTED BMI	FAM208A	2.09E-05	0.0688
HIP ADJUSTED BMI	VPS53	1.40E-05	0.0776
HIP ADJUSTED BMI	TRIM38	0.000199	0.254
HIP ADJUSTED BMI	ALDH1L1	0.000162	0.26
HIP ADJUSTED BMI	BTN2A2	2.07E-05	0.0651
HIP ADJUSTED BMI	SSBP2	0.000153	0.279
HIP ADJUSTED BMI	MST1P2	0.0007	0.292
HIP ADJUSTED BMI	HMGA2	2.17E-05	0.0809
HIP ADJUSTED BMI	THUMPD1	0.000153	0.172
HIP ADJUSTED BMI	RIOK3	0.000671	0.29
HIP ADJUSTED BMI	C1orf54	0.000423	0.281
HIP ADJUSTED BMI	BTN3A3	0.000176	0.273
HIP ADJUSTED BMI	ITPR3	0.000687	0.297
HIP ADJUSTED BMI	EIF2B1	6.61E-05	0.147
HIP ADJUSTED BMI	VTA1	3.10E-05	0.0809
HIP ADJUSTED BMI	MGARP	0.000622	0.284
HIP ADJUSTED BMI	SPIN1	0.000575	0.284
HIP ADJUSTED BMI	ZBTB38	3.91E-05	0.309

phenotype	gene	p-value	Q-value
HIP ADJUSTED BMI	HIST1H2BD	9.43E-06	0.0382
HIP ADJUSTED BMI	BHLHE41	0.000223	0.213
HIP ADJUSTED BMI	EBPL	4.80E-05	0.116
HIP ADJUSTED BMI	ESPNP	0.000178	0.279
HIP ADJUSTED BMI	HIST1H1C	2.51E-05	0.0718
HIP ADJUSTED BMI	GOLGA6L5P	0.000114	0.21
HIP ADJUSTED BMI	GKN2	0.000451	0.281
HIP ADJUSTED BMI	EHBP1L1	0.000257	0.231
HIP ADJUSTED BMI	BVES	0.000136	0.251
HIP ADJUSTED BMI	DBF4B	0.00035	0.262
HIP ADJUSTED BMI	TMEM117	0.000264	0.316
HIP ADJUSTED BMI	RASA2	0.000272	0.228
HIP ADJUSTED BMI	AMZ1	7.24E-05	0.172
HIP ADJUSTED BMI	OTUD4	1.89E-05	0.287
HIP ADJUSTED BMI	TMCC1-AS1	0.000182	0.26
HIP ADJUSTED BMI	MYH7B	0.000775	0.312
HIP ADJUSTED BMI	CEP250	0.000335	0.256
HIP ADJUSTED BMI	HIST1H3B	5.73E-06	0.036
HIP ADJUSTED BMI	FER1L4	7.95E-07	0.0222
HIP ADJUSTED BMI	HIST1H2BF	5.80E-05	0.122
HIP ADJUSTED BMI	EFEMP1	0.000237	0.225
INFLAMMATORY BOWEL DISEASE	BCAS3	0.000218	0.224
INFLAMMATORY BOWEL DISEASE	ADAM15	0.000114	0.163
INFLAMMATORY BOWEL DISEASE	NRN1	0.000474	0.308
INFLAMMATORY BOWEL DISEASE	TREX1	0.000186	0.218
INFLAMMATORY BOWEL DISEASE	CYLD	3.78E-05	0.138
INFLAMMATORY BOWEL DISEASE	USP4	0.000129	0.177
INFLAMMATORY BOWEL DISEASE	SLC26A6	0.000691	0.324
INFLAMMATORY BOWEL DISEASE	USP19	8.69E-05	0.142
INFLAMMATORY BOWEL DISEASE	KRTAP22-2	0.000578	0.308
INFLAMMATORY BOWEL DISEASE	CACNB1	8.36E-05	0.142
INFLAMMATORY BOWEL DISEASE	PDLIM4	0.00063	0.313
INFLAMMATORY BOWEL DISEASE	SALL4	0.000501	0.308
INFLAMMATORY BOWEL DISEASE	MEIS2	3.93E-05	0.138
INFLAMMATORY BOWEL DISEASE	NDUFS2	5.88E-05	0.138
INFLAMMATORY BOWEL DISEASE	TNFSF15	0.000439	0.308
INFLAMMATORY BOWEL DISEASE	PMPCA	9.07E-05	0.142
INFLAMMATORY BOWEL DISEASE	SLC22A5	6.42E-05	0.138
INFLAMMATORY BOWEL DISEASE	PRRC1	0.000412	0.308
INFLAMMATORY BOWEL DISEASE	DLD	0.000637	0.313
INFLAMMATORY BOWEL DISEASE	IP6K1	0.000238	0.237
INFLAMMATORY BOWEL DISEASE	NPIPB6	0.000557	0.308
INFLAMMATORY BOWEL DISEASE	GSDMB	0.000161	0.196
INFLAMMATORY BOWEL DISEASE	SPINK4	0.000302	0.276
INFLAMMATORY BOWEL DISEASE	RBM6	6.51E-05	0.138
INFLAMMATORY BOWEL DISEASE	KIF25	0.000475	0.308
INFLAMMATORY BOWEL DISEASE	LIPG	4.85E-05	0.138
INFLAMMATORY BOWEL DISEASE	GPX1	9.67E-05	0.144
INFLAMMATORY BOWEL DISEASE	UBE2L3	0.0006	0.313
INFLAMMATORY BOWEL DISEASE	IP6K2	0.000524	0.308
INFLAMMATORY BOWEL DISEASE	ORMDL3	0.000438	0.308
INFLAMMATORY BOWEL DISEASE	MST1	0.000298	0.276
INFLAMMATORY BOWEL DISEASE	RNF123	6.06E-05	0.138
INFLAMMATORY BOWEL DISEASE	SPANXN2	0.000706	0.326
INFLAMMATORY BOWEL DISEASE	CUL2	1.14E-05	0.0939
INFLAMMATORY BOWEL DISEASE	AMT	1.98E-06	0.0487
INFLAMMATORY BOWEL DISEASE	ETS2	0.000733	0.331
INFLAMMATORY BOWEL DISEASE	EIF3C	7.13E-05	0.138

phenotype	gene	p-value	Q-value
INFLAMMATORY BOWEL DISEASE	ERBB2	8.12E-05	0.142
INFLAMMATORY BOWEL DISEASE	MED13L	0.000566	0.308
INFLAMMATORY BOWEL DISEASE	ZPBP2	8.00E-06	0.0875
INFLAMMATORY BOWEL DISEASE	INSL6	0.000213	0.224
INFLAMMATORY BOWEL DISEASE	TRAIP	3.43E-05	0.138
INFLAMMATORY BOWEL DISEASE	SEMA3B-AS1	0.000136	0.178
INFLAMMATORY BOWEL DISEASE	P4HA2	0.000648	0.313
INFLAMMATORY BOWEL DISEASE	LYRM7	0.000256	0.248
INFLAMMATORY BOWEL DISEASE	CAMKV	2.97E-06	0.0487
INFLAMMATORY BOWEL DISEASE	SOHLH2	1.81E-05	0.103
INSULIN SECRETION	HNF4A	8.39E-06	0.132
INSULIN SECRETION	MARK2P9	1.38E-05	0.144
INSULIN SECRETION	SRSF10	7.42E-05	0.303
INSULIN SECRETION	SYMPK	4.00E-07	0.0125
INSULIN SECRETION	NRP2	3.86E-05	0.303
INSULIN SECRETION	KIF11	6.27E-06	0.112
INSULIN SECRETION	KPNA1	6.74E-05	0.303
INSULIN SECRETION	LRRC15	2.48E-05	0.302
INSULIN SECRETION	CCDC150	3.85E-05	0.302
INSULIN SECRETION	SLC36A4	3.26E-05	0.255
INSULIN SECRETION	PNRC2	5.04E-05	0.303
INSULIN SECRETION	HSPA6	9.87E-07	0.031
INSULIN SECRETION	MEP1B	2.50E-05	0.2
INSULIN SECRETION	CNKSR2	6.68E-05	0.303
INSULIN SECRETION	HHEX	1.98E-06	0.0622
INSULIN SECRETION	FAM131A	5.98E-05	0.303
INSULIN SECRETION	LPXN	2.47E-05	0.194
LONGEVITY	ZNF649	3.79E-06	0.117
LONGEVITY	PSMA5	9.71E-07	0.0264
LONGEVITY	PSRC1	1.64E-06	0.0264
LONGEVITY	SORT1	9.71E-06	0.0675
LONGEVITY	CRIP1	6.29E-05	0.254
LONGEVITY	LCAT	3.39E-05	0.267
LONGEVITY	FPR3	1.72E-05	0.266
LONGEVITY	CD81	1.63E-05	0.331
LONGEVITY	CELSR2	7.94E-05	0.271
LONGEVITY	SLC22A3	5.07E-05	0.254
LONGEVITY	MGA	2.28E-05	0.267
LONGEVITY	PSMA4	5.91E-06	0.0636
LOW-DENSITY LIPOPROTEIN	GSTM2	2.40E-05	0.0499
LOW-DENSITY LIPOPROTEIN	TMEM258	7.30E-05	0.104
LOW-DENSITY LIPOPROTEIN	RHD	7.00E-05	0.104
LOW-DENSITY LIPOPROTEIN	SLC24A4	0.000173	0.248
LOW-DENSITY LIPOPROTEIN	WTAP	5.56E-07	0.00435
LOW-DENSITY LIPOPROTEIN	GTF3C4	9.19E-05	0.12
LOW-DENSITY LIPOPROTEIN	BCAM	3.09E-05	0.0751
LOW-DENSITY LIPOPROTEIN	POLK	1.63E-05	0.0464
LOW-DENSITY LIPOPROTEIN	GDF7	7.85E-05	0.138
LOW-DENSITY LIPOPROTEIN	MKRN2	5.21E-05	0.096
LOW-DENSITY LIPOPROTEIN	KIAA1324	8.91E-05	0.12
LOW-DENSITY LIPOPROTEIN	SORT1	6.08E-05	0.0986
LOW-DENSITY LIPOPROTEIN	HS1BP3-IT1	0.000302	0.317
LOW-DENSITY LIPOPROTEIN	CEACAM19	0.000181	0.196
LOW-DENSITY LIPOPROTEIN	MAU2	0.000172	0.193
LOW-DENSITY LIPOPROTEIN	RHCE	0.000199	0.262
LOW-DENSITY LIPOPROTEIN	ST3GAL4	0.000319	0.317
LOW-DENSITY LIPOPROTEIN	SMARCA4	0.000159	0.185
LOW-DENSITY LIPOPROTEIN	ANGPTL3	1.92E-06	0.0141

phenotype	gene	p-value	Q-value
LOW-DENSITY LIPOPROTEIN	ATP13A1	2.48E-05	0.0499
LOW-DENSITY LIPOPROTEIN	TMEM50A	2.05E-05	0.0495
LOW-DENSITY LIPOPROTEIN	PSMA5	2.67E-06	0.0141
LOW-DENSITY LIPOPROTEIN	MAFB	0.000331	0.317
LOW-DENSITY LIPOPROTEIN	ATXN7L2	0.000305	0.299
LOW-DENSITY LIPOPROTEIN	MLEC	9.35E-06	0.0369
LOW-DENSITY LIPOPROTEIN	POC5	2.55E-05	0.0499
LOW-DENSITY LIPOPROTEIN	ABCG5	0.000198	0.262
LOW-DENSITY LIPOPROTEIN	DOCK7	8.85E-07	0.00554
LOW-DENSITY LIPOPROTEIN	SLC44A2	1.04E-07	0.00133
LOW-DENSITY LIPOPROTEIN	TMEM25	0.000294	0.317
LOW-DENSITY LIPOPROTEIN	GSTM4	9.53E-08	0.00133
LOW-DENSITY LIPOPROTEIN	APOE	1.19E-05	0.0374
LOW-DENSITY LIPOPROTEIN	APOB	5.63E-05	0.111
LOW-DENSITY LIPOPROTEIN	AMIGO1	5.95E-05	0.111
LOW-DENSITY LIPOPROTEIN	HNF1A	0.000126	0.199
LOW-DENSITY LIPOPROTEIN	PSRC1	3.18E-09	0.0001
LOW-DENSITY LIPOPROTEIN	ZNF441	4.54E-05	0.0956
LOW-DENSITY LIPOPROTEIN	SYPL2	0.000237	0.288
LUNG CANCER	SOCS1	3.88E-06	0.054
LUNG CANCER	MICAL1	6.53E-06	0.054
LUNG CANCER	TERT	1.43E-06	0.0244
MULTIPLE SCLEROSIS	HLA-DRB5	1.21E-05	0.132
MULTIPLE SCLEROSIS	WDR46	4.65E-05	0.202
MULTIPLE SCLEROSIS	GPR65	3.64E-06	0.0277
MULTIPLE SCLEROSIS	RNFT1	2.20E-05	0.112
MULTIPLE SCLEROSIS	PSMB9	1.58E-06	0.016
MULTIPLE SCLEROSIS	HLA-DOA	1.61E-07	0.00491
MULTIPLE SCLEROSIS	TAP2	4.22E-07	0.00643
MULTIPLE SCLEROSIS	GOLGB1	5.52E-05	0.324
MULTIPLE SCLEROSIS	SP110	1.40E-05	0.0854
MULTIPLE SCLEROSIS	RAD21L1	8.32E-06	0.255
NEUROBLASTOMA	MTFMT	2.23E-05	0.257
NEUROBLASTOMA	SHOX2	3.65E-05	0.257
NEUROBLASTOMA	OPRM1	4.23E-05	0.257
NEUROMYELITIS OPTICA	MIR3150A	9.12E-06	0.295
NEUROMYELITIS OPTICA	BTN3A2	1.82E-07	0.00294
NEUROMYELITIS OPTICA	ZCWPW1	2.73E-05	0.293
NEUROMYELITIS OPTICA	TRIM26	8.27E-06	0.133
NEUROMYELITIS OPTICA	FAM187B	9.84E-07	0.0317
NEUROTICISM	KANSL1	9.50E-05	0.0804
NEUROTICISM	FAM66D	1.65E-06	0.00332
NEUROTICISM	NMT1	0.000477	0.229
NEUROTICISM	FAM86B3P	1.31E-05	0.0156
NEUROTICISM	DEFB134	0.000752	0.295
NEUROTICISM	FAM167A	2.12E-05	0.0235
NEUROTICISM	SLC35G5	3.63E-08	0.00013
NEUROTICISM	ABCB6	0.000406	0.205
NEUROTICISM	LRRC37A	0.00015	0.103
NEUROTICISM	MSRA	3.92E-12	2.10E-08
NEUROTICISM	SERTAD1	0.000519	0.239
NEUROTICISM	PPP1R3B	1.90E-14	6.09E-10
NEUROTICISM	HCG18	4.66E-06	0.145
NEUROTICISM	BOLA2	3.24E-05	0.0347
NEUROTICISM	CCDC178	0.000585	0.258
NEUROTICISM	PLEKHM1	0.000577	0.258
NEUROTICISM	CLDN23	2.83E-12	1.82E-08
NEUROTICISM	LGALS3	0.00052	0.239

phenotype	gene	p-value	Q-value
NEUROTICISM	XKR6	1.90E-07	0.000554
NEUROTICISM	OR8B2	1.90E-06	0.00358
NEUROTICISM	CTSB	2.55E-07	0.000684
NEUROTICISM	LINC00208	0.000704	0.284
NEUROTICISM	CYSTM1	0.000392	0.205
NEUROTICISM	THAP2	0.000102	0.0837
NEUROTICISM	TNP2	1.13E-05	0.014
NEUROTICISM	ACVR2A	0.000193	0.126
NEUROTICISM	ACSL4	4.19E-05	0.0414
NEUROTICISM	BLK	4.25E-05	0.0414
NEUROTICISM	ZNF502	0.000614	0.267
NEUROTICISM	PARP8	0.000147	0.103
NEUROTICISM	ZNF671	0.000149	0.103
NEUROTICISM	ERI1	3.45E-06	0.00528
NEUROTICISM	DND1	2.18E-06	0.00389
NEUROTICISM	MFHAS1	3.41E-11	1.37E-07
NEUROTICISM	ARHGAP27	0.000113	0.0887
NEUROTICISM	ADM	0.000109	0.0879
NEUROTICISM	FAM66A	0.00044	0.216
NEUROTICISM	NSFP1	0.000777	0.301
NEUROTICISM	FDFT1	4.91E-14	7.90E-10
NEUROTICISM	TESC	0.000681	0.284
NEUROTICISM	DSC2	0.000223	0.135
NEUROTICISM	ARL17B	0.000693	0.284
NEUROTICISM	MTMR9	1.31E-13	1.28E-09
NEUROTICISM	SOX7	0.000816	0.312
NEUROTICISM	TDH	5.75E-05	0.0544
NEUROTICISM	NEIL2	1.60E-13	1.28E-09
NEUROTICISM	SLC25A23	0.000127	0.097
OBESITY	C1QL2	8.48E-05	0.276
OBESITY	LGR4	8.96E-05	0.276
OBESITY	BPIFB1	0.000189	0.306
OBESITY	FBXL19	0.000225	0.306
OBESITY	TUFM	0.000196	0.306
OBESITY	ADCY3	0.000133	0.276
OBESITY	EIF3CL	1.63E-05	0.17
OBESITY	SMAD3	0.000129	0.276
OBESITY	NEGR1	1.35E-06	0.0421
OBESITY	BDNF	8.24E-05	0.276
OBESITY	FTO	0.000138	0.276
OBESITY	ACP1	7.43E-05	0.276
OBESITY	RBL2	4.83E-05	0.276
OBESITY	FUBP1	2.51E-06	0.0392
OBESITY	NFATC2IP	3.51E-05	0.274
OBESITY	NPIPB9	0.000224	0.306
OSTEOPOROSIS	HOXC9	1.03E-06	0.0322
OSTEOPOROSIS	SEC11A	6.50E-06	0.212
OSTEOPOROSIS	NUDT11	6.65E-07	0.0209
OSTEOPOROSIS	GRN	8.29E-06	0.152
OSTEOPOROSIS	HTR1D	1.00E-05	0.152
OVERWEIGHT	C1QL2	5.35E-05	0.304
OVERWEIGHT	BDNF	5.82E-05	0.304
OVERWEIGHT	GALNT1	3.19E-05	0.304
OVERWEIGHT	EIF3CL	4.39E-05	0.304
OVERWEIGHT	NEGR1	5.98E-07	0.0187
PANCREATIC CANCER	COMMD5	2.80E-05	0.286
PANCREATIC CANCER	TSPYL2	4.22E-06	0.129
PARKINSON	PANX1	8.55E-05	0.162

phenotype	gene	p-value	Q-value
PARKINSON	PLEKHM1	2.61E-05	0.0879
PARKINSON	BRD8	0.000261	0.234
PARKINSON	MAPT-AS1	5.39E-05	0.119
PARKINSON	FAM83F	5.68E-06	0.161
PARKINSON	RDM1	7.69E-07	0.00583
PARKINSON	PIGG	0.00028	0.236
PARKINSON	ARL17B	1.76E-09	5.32E-05
PARKINSON	LUM	0.000233	0.228
PARKINSON	FUT3	0.000262	0.234
PARKINSON	MAPK8IP1	0.000393	0.298
PARKINSON	SLC2A13	0.000273	0.236
PARKINSON	BPTF	0.000114	0.169
PARKINSON	MAT2A	0.000113	0.169
PARKINSON	NMT1	0.000251	0.234
PARKINSON	ITGAL	4.17E-05	0.115
PARKINSON	MMRN1	0.000215	0.228
PARKINSON	KANSL1-AS1	3.51E-09	5.32E-05
PARKINSON	BCAP29	0.00011	0.169
PARKINSON	DND1	3.44E-07	0.00347
PARKINSON	MAPT	0.000125	0.173
PARKINSON	LRRC37A2	0.00016	0.194
PARKINSON	PDCD11	0.000317	0.253
PROSTATE CANCER	GIMAP5	8.83E-06	0.268
PUBERTAL GROWTH LATE	AMZ1	1.74E-05	0.214
PUBERTAL GROWTH LATE	ADAMTSL3	5.00E-05	0.274
PUBERTAL GROWTH LATE	FER1L4	2.76E-05	0.214
PUBERTAL GROWTH LATE	LRMP	5.30E-05	0.274
PUBERTAL GROWTH LATE	PFN4	1.85E-05	0.197
PUBERTAL GROWTH LATE	NDOR1	3.55E-05	0.296
PUBERTAL GROWTH LATE	NDUFAF1	1.66E-05	0.214
PUBERTAL GROWTH LATE	RPRD1A	1.35E-05	0.197
PUBERTAL GROWTH LATE	WDYHV1	8.48E-05	0.329
PUBERTAL GROWTH LATE	ZBTB38	1.10E-06	0.0341
PUBERTAL GROWTH LATE	ZW10	2.52E-05	0.197
PUBERTAL GROWTH LATE	DNAJC27-AS1	6.04E-05	0.314
PUBERTAL GROWTH LATE	GPR152	1.17E-05	0.168
PUBERTAL GROWTH LATE	AVPI1	7.20E-06	0.225
PUBERTAL GROWTH LATE	EFEMP1	1.62E-05	0.168
PUBERTAL GROWTH LATE	GNA12	8.15E-05	0.329
PUBERTAL GROWTH LATE	LIN28B	1.29E-05	0.201
PUBERTAL GROWTH LATE	ADCY3	7.97E-07	0.0249
PUBERTAL GROWTH LATE	LCE2C	2.11E-05	0.214
PUBERTAL GROWTH LATE	CENPO	1.33E-05	0.138
PUBERTAL GROWTH LATE	USP27X	4.20E-05	0.225
RHEUMATOID ARTHRITIS	TUT1	0.000229	0.247
RHEUMATOID ARTHRITIS	HLA-DPA1	3.42E-12	5.37E-08
RHEUMATOID ARTHRITIS	COL11A2	3.99E-05	0.0893
RHEUMATOID ARTHRITIS	HLA-DPB2	2.68E-14	8.43E-10
RHEUMATOID ARTHRITIS	SUOX	0.00017	0.118
RHEUMATOID ARTHRITIS	HSD17B8	5.95E-05	0.0655
RHEUMATOID ARTHRITIS	RSBN1	0.000134	0.119
RHEUMATOID ARTHRITIS	CUL5	1.56E-05	0.0246
RHEUMATOID ARTHRITIS	TMEM38B	0.000257	0.252
RHEUMATOID ARTHRITIS	BTN3A1	0.000176	0.239
RHEUMATOID ARTHRITIS	PPARD	0.000281	0.283
RHEUMATOID ARTHRITIS	TRIM9	5.59E-05	0.109
RHEUMATOID ARTHRITIS	TNFSF11	2.09E-06	0.00393
RHEUMATOID ARTHRITIS	TSFM	0.00012	0.0895

phenotype	gene	p-value	Q-value
RHEUMATOID ARTHRITIS	HSPA12B	7.24E-05	0.0723
RHEUMATOID ARTHRITIS	WDR86	0.000238	0.277
RHEUMATOID ARTHRITIS	ZBTB22	1.44E-06	0.00308
RHEUMATOID ARTHRITIS	RXRB	3.36E-07	0.00132
RHEUMATOID ARTHRITIS	HIST1H4B	0.000193	0.252
RHEUMATOID ARTHRITIS	ERBB2	7.33E-05	0.0711
RHEUMATOID ARTHRITIS	GGNBP1	0.00036	0.29
RHEUMATOID ARTHRITIS	RPS18	0.000111	0.0895
RHEUMATOID ARTHRITIS	RGL2	3.87E-07	0.00124
RHEUMATOID ARTHRITIS	TAP2	2.59E-06	0.0046
RHEUMATOID ARTHRITIS	DCLRE1B	1.96E-05	0.0294
RHEUMATOID ARTHRITIS	HIST1H2BD	1.21E-07	0.00175
RHEUMATOID ARTHRITIS	HMGN4	5.50E-06	0.0251
RHEUMATOID ARTHRITIS	HLA-DQA1	8.68E-10	5.61E-06
RHEUMATOID ARTHRITIS	PSMB9	1.27E-06	0.00304
RHEUMATOID ARTHRITIS	SMAD3	0.000154	0.128
RHEUMATOID ARTHRITIS	VPS52	0.000118	0.0895
RHEUMATOID ARTHRITIS	TBC1D22B	1.96E-05	0.0294
RHEUMATOID ARTHRITIS	BAK1	1.80E-08	6.39E-05
RHEUMATOID ARTHRITIS	HLA-DRB1	0.000125	0.152
RHEUMATOID ARTHRITIS	TUBB6	0.000415	0.246
RHEUMATOID ARTHRITIS	CACNB1	5.82E-05	0.0621
RHEUMATOID ARTHRITIS	BLK	0.000342	0.219
RHEUMATOID ARTHRITIS	IP6K3	7.21E-06	0.0251
RHEUMATOID ARTHRITIS	MEGF8	6.94E-05	0.0719
RHEUMATOID ARTHRITIS	CLEC16A	0.000379	0.233
RHEUMATOID ARTHRITIS	DAXX	4.47E-09	1.92E-05
RHEUMATOID ARTHRITIS	HIPK1	1.04E-05	0.0193
RHEUMATOID ARTHRITIS	HLA-DRB5	8.77E-12	2.80E-07
RHEUMATOID ARTHRITIS	ZNF322	4.52E-05	0.0944
RHEUMATOID ARTHRITIS	GSDMB	1.71E-05	0.0246
RHEUMATOID ARTHRITIS	MAGI3	1.41E-05	0.0246
RHEUMATOID ARTHRITIS	HIST1H4A	0.000164	0.234
RHEUMATOID ARTHRITIS	HLA-DOA	2.16E-07	0.000972
RHEUMATOID ARTHRITIS	HIST1H2AC	3.74E-05	0.0893
RHEUMATOID ARTHRITIS	WDR46	9.17E-06	0.018
RHEUMATOID ARTHRITIS	TAPBP	4.81E-09	1.92E-05
RHEUMATOID ARTHRITIS	H1FX	0.000106	0.0992
RHEUMATOID ARTHRITIS	HLA-DMA	7.22E-06	0.0251
RHEUMATOID ARTHRITIS	MFAP2	5.41E-05	0.0598
RHEUMATOID ARTHRITIS	HLA-DQB2	2.21E-05	0.0272
RHEUMATOID ARTHRITIS	HLA-DOB	1.21E-06	0.00759
RHEUMATOID ARTHRITIS	PHF1	7.64E-07	0.00267
RHEUMATOID ARTHRITIS	AP4B1-AS1	0.000307	0.261
RHEUMATOID ARTHRITIS	MLN	0.000126	0.152
RHEUMATOID ARTHRITIS	COCH	0.000203	0.236
RHEUMATOID ARTHRITIS	HMGA1	1.64E-05	0.0246
RHEUMATOID ARTHRITIS	CUTA	2.50E-06	0.00535
RHEUMATOID ARTHRITIS	ITPR3	2.27E-11	3.62E-07
RHEUMATOID ARTHRITIS	RNASET2	6.78E-07	0.00531
RHEUMATOID ARTHRITIS	HLA-DPB1	8.78E-10	5.61E-06
RHEUMATOID ARTHRITIS	PHF19	0.00034	0.281
RHEUMATOID ARTHRITIS	HCG24	1.30E-06	0.00304
RHEUMATOID ARTHRITIS	HLA-DQB1	4.84E-05	0.0552
SCHIZOPHRENIA	TRPC4	0.000175	0.196
SCHIZOPHRENIA	NT5C2	9.75E-09	0.000152
SCHIZOPHRENIA	BCAP29	5.34E-05	0.159
SCHIZOPHRENIA	HLA-DPB1	0.000358	0.29

phenotype	gene	p-value	Q-value
SCHIZOPHRENIA	TNKS	0.000266	0.269
SCHIZOPHRENIA	BTN3A1	0.000145	0.267
SCHIZOPHRENIA	MCHR1	1.54E-05	0.0536
SCHIZOPHRENIA	CSH2	0.000173	0.3
SCHIZOPHRENIA	KCNJ13	0.000102	0.137
SCHIZOPHRENIA	CREB1	9.38E-05	0.209
SCHIZOPHRENIA	ASAH1	0.00013	0.282
SCHIZOPHRENIA	GNL1	8.94E-05	0.208
SCHIZOPHRENIA	C16orf96	0.000198	0.31
SCHIZOPHRENIA	SFXN2	0.000129	0.235
SCHIZOPHRENIA	BTN3A2	1.60E-11	5.01E-07
SCHIZOPHRENIA	ZSCAN16	0.000147	0.282
SCHIZOPHRENIA	VPS29	0.000351	0.29
SCHIZOPHRENIA	MPHOSPH9	0.000226	0.244
SCHIZOPHRENIA	ZFP57	0.00014	0.235
SCHIZOPHRENIA	CHRNB4	4.41E-05	0.11
SCHIZOPHRENIA	HLA-DQA1	9.78E-06	0.0528
SCHIZOPHRENIA	DPYD	0.000177	0.247
SCHIZOPHRENIA	KLC1	7.72E-05	0.131
SCHIZOPHRENIA	TM9SF1	0.000157	0.237
SCHIZOPHRENIA	SLC17A4	7.86E-05	0.205
SCHIZOPHRENIA	HCG11	7.64E-05	0.131
SCHIZOPHRENIA	FRMD7	0.000129	0.282
SCHIZOPHRENIA	BTN3A3	0.00014	0.167
SCHIZOPHRENIA	LEMD2	1.74E-05	0.0564
SCHIZOPHRENIA	C11orf24	8.23E-08	0.000858
SCHIZOPHRENIA	CNNM2	2.20E-05	0.0762
SCHIZOPHRENIA	PDCD11	0.000339	0.321
SCHIZOPHRENIA	AK2	3.96E-05	0.103
SCHIZOPHRENIA	GTF2A1L	2.72E-05	0.0852
SCHIZOPHRENIA	PSMB9	0.000355	0.29
SCHIZOPHRENIA	HIST1H2AC	0.000503	0.33
SCHIZOPHRENIA	ZKSCAN4	3.27E-05	0.0928
SCHIZOPHRENIA	HLA-DPB2	0.00051	0.33
SCHIZOPHRENIA	UPB1	0.000455	0.307
SCHIZOPHRENIA	WDR82	1.71E-05	0.0762
SCHIZOPHRENIA	BTN1A1	1.64E-07	0.00128
SCHIZOPHRENIA	HFE	1.56E-05	0.061
SCHIZOPHRENIA	HIST1H4C	3.34E-06	0.0261
SCHIZOPHRENIA	CCDC107	5.38E-05	0.168
SCHIZOPHRENIA	ANKRD44	0.000157	0.181
SCHIZOPHRENIA	PPP1R3B	1.29E-05	0.0541
SCHIZOPHRENIA	ABCB9	2.97E-05	0.0875
SCHIZOPHRENIA	HLA-DQA2	3.82E-06	0.0268
SCHIZOPHRENIA	GATAD2A	0.000346	0.29
SCHIZOPHRENIA	TAP2	0.000318	0.29
SCHIZOPHRENIA	WBP1L	1.24E-05	0.0578
SCHIZOPHRENIA	ARL3	9.56E-07	0.00886
SCHIZOPHRENIA	HIST1H4B	0.000248	0.258
SMOKING CIGAR PER DAY	ZNF665	3.99E-06	0.125
SMOKING EVER SMOKE	RTEL1	2.00E-05	0.314
SMOKING EVER SMOKE	TXN	4.41E-06	0.138
STOKE	ICA1L	7.34E-05	0.302
STOKE	LRRC18	3.98E-05	0.253
STOKE	RTKN	3.81E-05	0.326
STOKE	PIGA	5.04E-05	0.267
STOKE	ACLY	2.50E-06	0.0809
STOKE	CRIP1	4.02E-05	0.326

phenotype	gene	p-value	Q-value
STOKE	NPEPPS	9.49E-05	0.302
STOKE	ALDH2	1.44E-07	0.00456
STOKE	AMT	0.000113	0.302
STOKE	CDC42EP5	9.87E-05	0.302
STOKE	CENPQ	8.99E-05	0.302
STOKE	LRCH1	3.08E-05	0.326
STOKE	AS3MT	3.36E-05	0.253
STOKE	KRTAP19-7	1.86E-06	0.0296
SUBJECTIVE WELL BEING	ZFAS1	3.09E-08	0.000917
SUBJECTIVE WELL BEING	ZNFX1	1.62E-06	0.024
SYSTOLIC BLOOD PRESSURE	AGO1	0.000115	0.328
SYSTOLIC BLOOD PRESSURE	LMAN1L	4.49E-05	0.235
SYSTOLIC BLOOD PRESSURE	OXER1	5.62E-05	0.252
SYSTOLIC BLOOD PRESSURE	SFXN2	6.31E-06	0.198
SYSTOLIC BLOOD PRESSURE	SLC39A12-AS1	9.77E-05	0.328
SYSTOLIC BLOOD PRESSURE	ITGA3	2.91E-05	0.229
SYSTOLIC BLOOD PRESSURE	CSK	1.97E-05	0.207
SYSTOLIC BLOOD PRESSURE	ADM	4.41E-05	0.235
TRIGLYCERIDES	DOCK7	1.60E-08	0.000167
TRIGLYCERIDES	OR8B2	8.57E-05	0.112
TRIGLYCERIDES	LINC00599	0.000272	0.251
TRIGLYCERIDES	FNDC4	1.86E-06	0.00973
TRIGLYCERIDES	ANGPTL3	4.13E-06	0.0144
TRIGLYCERIDES	ADAL	5.42E-05	0.0772
TRIGLYCERIDES	SLC5A6	5.74E-05	0.0782
TRIGLYCERIDES	KHK	2.88E-05	0.0477
TRIGLYCERIDES	GCKR	2.19E-05	0.0412
TRIGLYCERIDES	FADS3	0.000155	0.173
TRIGLYCERIDES	PPP1R3B	0.000249	0.245
TRIGLYCERIDES	MPV17	1.25E-05	0.0308
TRIGLYCERIDES	ATP13A1	4.73E-05	0.0706
TRIGLYCERIDES	RNF214	0.00025	0.245
TRIGLYCERIDES	TAGLN	1.45E-09	3.07E-05
TRIGLYCERIDES	MLXIPL	3.18E-08	0.000199
TRIGLYCERIDES	SNX17	2.89E-05	0.0477
TRIGLYCERIDES	AFF1	9.94E-05	0.125
TRIGLYCERIDES	TMEM31	0.000211	0.228
TRIGLYCERIDES	OST4	0.000245	0.245
TRIGLYCERIDES	CGREF1	6.10E-06	0.0191
TRIGLYCERIDES	CSGALNACT1	3.90E-05	0.061
TRIGLYCERIDES	RBKS	1.49E-05	0.0332
TRIGLYCERIDES	BLK	0.000126	0.146
TRIGLYCERIDES	PCSK7	1.28E-05	0.0308
TRIGLYCERIDES	SIDT2	1.96E-09	3.07E-05
TRIGLYCERIDES	IFT172	0.000108	0.13
TRIGLYCERIDES	NEIL2	1.95E-05	0.0407
TYPE 2 DIABETES	ZNF239	9.79E-06	0.145
ULCERATIVE COLITIS	GALNT9	1.98E-05	0.0328
ULCERATIVE COLITIS	MED24	0.000799	0.294
ULCERATIVE COLITIS	ZGLP1	0.00105	0.32
ULCERATIVE COLITIS	ZPBP2	3.22E-05	0.214
ULCERATIVE COLITIS	CYTH1	0.000601	0.242
ULCERATIVE COLITIS	ACAP1	1.75E-05	0.0319
ULCERATIVE COLITIS	CDH18	2.98E-05	0.0408
ULCERATIVE COLITIS	DGCR9	0.000302	0.173
ULCERATIVE COLITIS	ELAVL4	7.89E-05	0.0709
ULCERATIVE COLITIS	CPNE9	7.80E-05	0.0709
ULCERATIVE COLITIS	ISG15	0.000443	0.208

phenotype	gene	p-value	Q-value
ULCERATIVE COLITIS	GPR88	4.56E-05	0.0548
ULCERATIVE COLITIS	ST8SIA5	6.06E-06	0.0176
ULCERATIVE COLITIS	LRMP	0.00101	0.32
ULCERATIVE COLITIS	CASQ1	0.000205	0.134
ULCERATIVE COLITIS	DLD	6.75E-06	0.0177
ULCERATIVE COLITIS	CDHR4	0.000923	0.312
ULCERATIVE COLITIS	S100G	9.26E-06	0.0208
ULCERATIVE COLITIS	BSPRY	0.000287	0.167
ULCERATIVE COLITIS	ADORA2A	0.000119	0.0947
ULCERATIVE COLITIS	WRB	3.88E-06	0.0153
ULCERATIVE COLITIS	USP4	2.22E-05	0.0333
ULCERATIVE COLITIS	TCERG1L	1.82E-05	0.0319
ULCERATIVE COLITIS	IGSF1	4.51E-06	0.0158
ULCERATIVE COLITIS	UBA7	0.000121	0.0947
ULCERATIVE COLITIS	CAMKV	1.02E-05	0.214
ULCERATIVE COLITIS	C20orf27	9.07E-06	0.0208
ULCERATIVE COLITIS	TGM6	0.00022	0.27
ULCERATIVE COLITIS	IP6K1	1.30E-05	0.0269
ULCERATIVE COLITIS	DAG1	5.47E-05	0.0573
ULCERATIVE COLITIS	KRTAP22-2	0.000402	0.198
ULCERATIVE COLITIS	DOCK3	0.000127	0.247
ULCERATIVE COLITIS	STX1A	6.14E-06	0.0176
ULCERATIVE COLITIS	RGS14	2.39E-13	7.51E-09
ULCERATIVE COLITIS	MEIS2	4.00E-05	0.214
ULCERATIVE COLITIS	SLC30A3	0.00054	0.235
ULCERATIVE COLITIS	NKX2-3	6.60E-05	0.214
ULCERATIVE COLITIS	TNFRSF14	0.000624	0.248
ULCERATIVE COLITIS	RHOA	1.37E-05	0.0269
ULCERATIVE COLITIS	CCDC88C	2.21E-05	0.0333
ULCERATIVE COLITIS	CBWD1	0.000875	0.309
ULCERATIVE COLITIS	NEUROD2	3.24E-05	0.0424
ULCERATIVE COLITIS	NXPE1	7.17E-05	0.214
ULCERATIVE COLITIS	HERC2	0.000194	0.27
ULCERATIVE COLITIS	OLAH	6.06E-05	0.0615
ULCERATIVE COLITIS	RTN4R	0.000146	0.112
ULCERATIVE COLITIS	C1orf194	0.000324	0.182
ULCERATIVE COLITIS	SOHLH2	0.00102	0.32
ULCERATIVE COLITIS	HTR4	0.000203	0.134
ULCERATIVE COLITIS	ARHGEF25	0.000443	0.208
ULCERATIVE COLITIS	GSDMB	0.000217	0.27
ULCERATIVE COLITIS	KCNJ6	0.000221	0.14
ULCERATIVE COLITIS	PASK	2.46E-06	0.0111
ULCERATIVE COLITIS	PCDHGC3	0.000797	0.294
ULCERATIVE COLITIS	ERBB2	0.000423	0.205
ULCERATIVE COLITIS	APEH	0.000114	0.247
ULCERATIVE COLITIS	TMA7	0.000151	0.247
ULCERATIVE COLITIS	RNF123	6.11E-07	0.0048
ULCERATIVE COLITIS	IFIT3	0.000164	0.123
ULCERATIVE COLITIS	MST1	0.000105	0.247
ULCERATIVE COLITIS	CACNB1	8.74E-05	0.0743
ULCERATIVE COLITIS	FAM212A	0.000587	0.24
ULCERATIVE COLITIS	ORMDL3	4.70E-05	0.0548
ULCERATIVE COLITIS	NPTX1	0.000222	0.14
ULCERATIVE COLITIS	CHRM3	4.10E-07	0.0043
ULCERATIVE COLITIS	GSDMA	0.000819	0.296
ULCERATIVE COLITIS	RBM6	1.45E-06	0.00762
ULCERATIVE COLITIS	HPCAL1	2.62E-05	0.0374
ULCERATIVE COLITIS	AMT	7.74E-05	0.0709

phenotype	gene	p-value	Q-value
ULCERATIVE COLITIS	RAPGEFL1	5.16E-05	0.0565
ULCERATIVE COLITIS	ETS2	0.000222	0.27
ULCERATIVE COLITIS	WDR66	1.20E-06	0.00752
ULCERATIVE COLITIS	PSMD3	0.000373	0.192
ULCERATIVE COLITIS	SLC7A6OS	0.000514	0.229
ULCERATIVE COLITIS	FCER1G	9.59E-05	0.247
ULCERATIVE COLITIS	ZMYND10	3.81E-05	0.0479
ULCERATIVE COLITIS	CAMSAP2	7.21E-05	0.0709
UPPER GASTROINTESTINAL	SEMG2	1.89E-06	0.0286
UPPER GASTROINTESTINAL	PLCE1	1.83E-06	0.0286
UPPER GASTROINTESTINAL	PLCE1-AS1	2.40E-05	0.182
WAIST CIRCUMFERENCE	SBK1	9.80E-05	0.209
WAIST CIRCUMFERENCE	RAG1	4.31E-05	0.149
WAIST CIRCUMFERENCE	ANKDD1B	8.74E-05	0.209
WAIST CIRCUMFERENCE	NFATC2IP	4.79E-06	0.0419
WAIST CIRCUMFERENCE	HCN4	4.35E-06	0.0419
WAIST CIRCUMFERENCE	TMEM219	1.76E-05	0.0426
WAIST CIRCUMFERENCE	EIF3CL	1.97E-06	0.0137
WAIST CIRCUMFERENCE	FTO	9.40E-05	0.265
WAIST CIRCUMFERENCE	IL27	2.53E-05	0.0874
WAIST CIRCUMFERENCE	LGR4	5.34E-05	0.112
WAIST CIRCUMFERENCE	CRY1	0.000186	0.253
WAIST CIRCUMFERENCE	TUFM	6.58E-07	0.00414
WAIST CIRCUMFERENCE	RBL2	1.60E-05	0.062
WAIST CIRCUMFERENCE	VWA7	0.000195	0.255
WAIST CIRCUMFERENCE	OTX1	5.91E-05	0.154
WAIST CIRCUMFERENCE	NPIPB9	0.000118	0.311
WAIST CIRCUMFERENCE	NEGR1	3.88E-07	0.00305
WAIST CIRCUMFERENCE	NPIPB6	2.19E-07	0.00279
WAIST CIRCUMFERENCE	OAZ1	0.000106	0.277
WAIST CIRCUMFERENCE	RABEP2	0.000144	0.325
WAIST CIRCUMFERENCE	UPP1	0.000231	0.289
WAIST CIRCUMFERENCE	TFAP2B	1.01E-06	0.00452
WAIST CIRCUMFERENCE	NUPR1	1.08E-06	0.00422
WAIST CIRCUMFERENCE	FUBP1	2.30E-06	0.00804
WAIST CIRCUMFERENCE	DNMT3L	0.000159	0.324
WAIST CIRCUMFERENCE	EIF3C	7.31E-05	0.232
WAIST CIRCUMFERENCE ADJUSTED	GOLGA6L5P	0.000181	0.33
WAIST CIRCUMFERENCE ADJUSTED	HIST1H2BK	0.000124	0.313
WAIST CIRCUMFERENCE ADJUSTED	KRT6C	0.000159	0.329
WAIST CIRCUMFERENCE ADJUSTED	DYNLT1	0.000159	0.331
WAIST CIRCUMFERENCE ADJUSTED	CTCF	6.12E-05	0.274
WAIST CIRCUMFERENCE ADJUSTED	CEP63	2.08E-05	0.147
WAIST CIRCUMFERENCE ADJUSTED	ANAPC13	1.11E-05	0.117
WAIST CIRCUMFERENCE ADJUSTED	AMZ1	0.000102	0.288
WAIST CIRCUMFERENCE ADJUSTED	SLC38A11	3.52E-06	0.0716
WAIST CIRCUMFERENCE ADJUSTED	CCDC92	1.41E-06	0.0442
WAIST CIRCUMFERENCE ADJUSTED	ZNF606	0.000139	0.328
WAIST CIRCUMFERENCE ADJUSTED	C5orf47	5.21E-05	0.233
WAIST CIRCUMFERENCE ADJUSTED	STXBP3	9.14E-05	0.284
WAIST CIRCUMFERENCE ADJUSTED	HIST1H3B	6.86E-06	0.106
WAIST CIRCUMFERENCE ADJUSTED	ZNF385D	2.37E-05	0.151
WAIST CIRCUMFERENCE ADJUSTED	LOXL1-AS1	1.03E-05	0.151
WAIST CIRCUMFERENCE ADJUSTED	TRIM8	3.42E-05	0.241
WAIST CIRCUMFERENCE ADJUSTED	FOXS1	4.62E-05	0.241
WAIST CIRCUMFERENCE ADJUSTED	GPR152	4.68E-06	0.101
WAIST CIRCUMFERENCE ADJUSTED	KCTD19	0.000143	0.324
WAIST CIRCUMFERENCE ADJUSTED	ADAMTSL3	4.52E-06	0.0716

phenotype	gene	p-value	Q-value
WAIST CIRCUMFERENCE ADJUSTED	CPAMD8	0.000179	0.331
WAIST CIRCUMFERENCE ADJUSTED	CROCC	3.29E-05	0.17
WAIST CIRCUMFERENCE ADJUSTED	HAS2	2.63E-05	0.147
WAIST CIRCUMFERENCE ADJUSTED	HIST1H2AC	9.80E-06	0.101
WAIST CIRCUMFERENCE ADJUSTED	RIOK3	0.000155	0.324
WAIST CIRCUMFERENCE ADJUSTED	HIST1H2BD	3.63E-05	0.182
WAIST CIRCUMFERENCE ADJUSTED	WDR82	9.89E-05	0.31
WAIST CIRCUMFERENCE ADJUSTED	TUBD1	2.78E-05	0.147
WAIST-HIP RATIO	SLC38A11	6.83E-06	0.217
WAIST-HIP RATIO	HCN4	9.00E-06	0.143
WAIST-HIP RATIO ADJUSTED FOR	CCDC92	5.34E-07	0.0168
WAIST-HIP RATIO ADJUSTED FOR	GDF5	2.66E-06	0.0827
WAIST-HIP RATIO ADJUSTED FOR	RAF1	1.90E-05	0.199
WAIST-HIP RATIO ADJUSTED FOR	SEMA3G	6.57E-06	0.103
WAIST-HIP RATIO ADJUSTED FOR	ZBTB7B	4.20E-05	0.33
WAIST-HIP RATIO ADJUSTED FOR	CAPNS1	2.36E-05	0.243
WAIST-HIP RATIO ADJUSTED FOR	CTCF	6.29E-05	0.268
WAIST-HIP RATIO ADJUSTED FOR	MKRN2	1.12E-05	0.176
WAIST-HIP RATIO ADJUSTED FOR	FER1L4	1.03E-05	0.159
WAIST-HIP RATIO ADJUSTED FOR	MYH7B	2.36E-05	0.249
WAIST-HIP RATIO ADJUSTED FOR	SLC38A11	1.88E-05	0.228
WAIST-HIP RATIO ADJUSTED FOR	ADAM15	4.20E-05	0.33
WAIST-HIP RATIO ADJUSTED FOR	SCN2A	2.16E-05	0.228
WAIST-HIP RATIO ADJUSTED FOR	KRT28	2.02E-05	0.249
WAIST-HIP RATIO ADJUSTED FOR	FAM110B	9.44E-06	0.228
WAIST-HIP RATIO ADJUSTED FOR	CHKB	2.91E-05	0.228
WEIGHT	NEGR1	7.77E-07	0.0123
WEIGHT	FUBP1	7.63E-07	0.0123
WEIGHT	HCN4	2.32E-06	0.0245

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

6/7/2019

Date