Leveraging genome-wide association studies data to inform the biology behind the genetic risk for mitral valve prolapse

Mengyao Yu

UNIVERSITÉ PARIS DESCARTES

MEMBRE DE USPC Université Sorbonne Paris Cité

Inserm

Université Paris Descartes

ED562

Paris-Cardiovascular Research Center (PARCC)

Equipe : « Approches génétiques pour comprendre les maladies artérielles »

Exploitation des données issues d'études d'association pangénomiques pour caractériser les voies biologiques associées au risque génétique du prolapsus de la valve mitrale

Leveraging genome-wide association studies data to inform the biology behind the genetic risk for mitral valve prolapse

Par Mengyao Yu

Doctorat en Génétique

Dirigée par Nabila BOUATIA-NAJI

Présentée et soutenue publiquement le 18 Juin 2019

Devant un jury composé de :

Dr. Jean-Charles LAMBERT	Rapporteur
Dr. Stéphane ZAFFRAN	Rapporteur
Dr. Fabienne LESUEUR	Examinateur
Pr. Elisabeth TOURNIER-LASSERVE	Examinateur
Dr Nabila BOUATIA-NAJI	Directrice de thèse

Acknowledgment

Firstly, I would like to acknowledge my supervisor Dr. Nabila Bouatia-Naji who offered the precious opportunity to me to start an interesting topic in human genetics on cardiovascular diseases. I am lucky to have such a responsible, affinitive and brilliant researcher as my supervisor. During the Ph.D. study, she trained me with the target of being a good researcher with her broad vision and rich experience. She kept me in the right direction during my study with her great patience, by sharing many valuable ideas and placing her great trust in me. I appreciate all the chances that she has offered to learn how to think and behave like a scientist, such as supporting me to attend international conferences during my training, teaching me how to share my work with the experts in the field and guiding me to write a qualified research article. She has always been a great supporter of my research work and life.

I am grateful to Dr. Lambert Jean-Charles and Dr. Zaffran Stéphane for accepting to review my thesis and for carefully reading and commenting on the manuscript. I thank Fabienne Lesueur and Elisabeth Tournier-Lasserve for taking the time to examine my work.

I would like to thank Dr. Adrien Georges and Dr. Sergiy Kyryachenko, who helped me to overcome difficulties in my work with their valuable opinions. Thank many discussions with Pr. Xavier Jeunemaitre, Dr. Aurelie Morin, Ms. Judith Goncalves, Ms. Yujiao Zhang, my presentation skills got improved. I would like to thank Dr. Soto Romuald Kiando, who helped me to start the GWAS analysis at the beginning of my thesis. I would like to express my gratitude to the collaborators in the MVP project, especially Dr. Russell A. Norris and Dr. Francesca N. Delling, who provided me a great opportunity to participate in other MVP related works. I would like to thank Pr. Albert A. HAGEGE for providing me a recommendation letter. Of course, the success of my work in the Paris Cardiovascular Research Center (PARCC) is inseparable from the wonderful people I have met in the past four years. I would never forget how my friends helped me when I first came to French, I would remember every “soirée” after the work and “retrait”. All the valuable experiences that happened in PARCC are deeply printed in my memory and they will always be with me.

In the end, I would like to thank my husband for his meticulous care and support. Without him, I would not come to Paris, I would not have the opportunity to become a young researcher and would not encounter so many excellent people and things. It is him who makes our life wonderful. We are looking forward to a new chapter of our lives, but we will never forget the life in Paris.

Abstract

Mitral valve prolapse (MVP) is a common heart valve disease affecting nearly 1 in 40 individuals in the general population. It is the first indication for valve repair and/or replacement and moreover, a risk factor for mitral regurgitation, an established cause of endocarditis and sudden death. MVP is characterized by excess extracellular matrix secretion and cellular disorganization which leads to bulky valves that are unable to coapt correctly during ventricular systole. Even though several genes including *FLNA*, *DCHS1 TNS1*, and *LMCD1* were reported to be associated with MVP, these explain partially its heritability. However, understanding the biological mechanisms underlying the genetic susceptibility to MVP is necessary to characterize its triggering mechanisms.

In this thesis, I aimed 1) to characterize globally the biological mechanisms involved in the genetic risk for MVP in the context of genome-wide association studies (GWAS), and 2) improve the genotyping resolution using genetic imputation, which allowed the discovery of additional risk genes for MVP.

In the first part of my study, I applied pathway enrichment tools (i-GSEA4GWAS, DEPICT) to the GWAS data. I was able to show that genes at risk loci are involved in biological functions relevant to actin filament organization, cytoskeleton biology, and cardiac development. The enrichment for positive regulation of transcription, cell proliferation, and migration motivated the follow-up of *GLIS1*, a transcription factor that regulates Hedgehog signaling. I followed up the association with MVP in a dataset of cases and controls from the UK Biobank and, in combination with previously available data, I found a genome-wide significant association with MVP (OR=1.22, P=4.36 $\times 10^{-10}$). Through collaborative efforts, immunohistochemistry experiments in mouse indicated that *Glis1* is expressed during embryonic development predominantly in nuclei of endothelial and interstitial cells of mitral valves, while *Glis1* knockdown using morpholinos caused atrioventricular regurgitation in zebrafish.

In the second part of my work, I generated larger genotyping datasets using an imputation based on HRC and TOPMed, two large and highly dense imputation panels that were recently made available. I first compared the imputation accuracy between data using HRC and TopMED and found that both panels have low imputation accuracy for the rare allele (MAF<0.01). However, the imputation accuracy increased with the input sample size for common variants (MAF>0.05), especially when genotyping platforms were harmonized. I was able to fine map established loci (e.g Chr 2) and also able to identify six novel and promising associated loci. All new loci are driven by common variants that I confirmed as high profile regulatory variants through extensive computationally-based functional annotations at promising loci that pointed at several candidate genes for valve biology and development (e.g *PDGFD* and *ACTN4)*.

In summary, my Ph.D. work applied up-to-date high throughput genetic association methods and functional enrichment and annotation to GWAS data. My results provide novel insights into the genetics, molecular and cellular basis of valve disease. Further genetic confirmation through replication, but also through biological experiments are expected to consolidate these statistically and computationally supported results.

Keywords: GWAS, heart valve disease, valve development, mitral valve prolapse, enrichment analysis, imputation

Résumé

Le prolapsus de la valve mitrale (MVP) est une valvulopathie fréquente qui touche près de 1 personne sur 40 dans la population générale. De nombreux gènes comme *FLNA, DCHS1* pour les formes familiales et *TNS1* et *LMCD1* pour les formes sporadiques ont récemment été décrit comme associés au MVP. Cependant, les défauts génétiques touchant ces gènes n'expliquent pas tous les cas du MVP. De plus, les mécanismes biologiques expliquant la susceptibilité génétique au MVP, notamment pour les formes sporadiques les plus fréquentes restent mal compris.

Dans cette thèse, mon objectif était 1) de caractériser globalement les mécanismes biologiques impliqués dans le risque génétique du MVP dans le contexte des études d'association pangénomique (GWAS), et 2) d'améliorer la résolution du génotypage par l'imputation génétique et par l'addition d'une nouvelle étude cas témoins, (UKBioBank) afin de permettre la découverte de nouveaux loci de prédisposition.

Dans la première partie de mon travail de thèse, j'ai appliqué des outils d'enrichissement de voies biologiques ou sets de gènes aux données GWAS. J'ai pu montrer que les gènes présents autour des loci GWAS sont impliqués dans l'organisation des filaments d'actine, l'organisation du cytosquelette et le développement cardiaque, ce qui était attendu. Nous avons également décrits le gène *GLIS1*. Afin de confirmer le rôle de *GLIS1* dans l'association avec le MVP, j'ai réalisé une analyse génétique dans une étude case témoins (UKBiobank) et, en combinaison avec les données françaises, l'association a atteint le seuil de significativité génomique ($P = 4{,}36 \times 10^{-10}$). En collaboration avec des équipes américaines, des expériences d'immunohistochimie ont indiqué que Glis1, la protéine orthologue de la souris est exprimée au cours du développement embryonnaire. D'autre part, l'inactivation de *Glis1* à l'aide d'oligonucléotides de type *Morpholinos* ont été l'origine d'une régurgitation atrio-ventriculaire chez le poisson zèbre.

Dans la deuxième partie de mon travail, j'ai généré des données de génotypage plus dense à l'aide d'une imputation basée sur HRC et TOPMed, deux panels d'imputation ultra-dense récemment mis à disposition de la communauté internationale. J'ai d'abord comparé la précision d'imputation entre les données utilisant les différents panels et constaté qu'aucun panel n'atteignait une précision optimale pour les variants rares (MAF <0,01) dans nos échantillons. La précision d'imputation s'améliorait pour les variants fréquents (MAF> 0,05), en particulier pour les cohortes dont le génotypage étaient réalisé avec des puces identiques. J'ai pu ainsi cartographier avec plus de précision les loci déjà confirmés (ex. Chr 2 autour de *TNS1*). J'ai également identifié 6 nouveaux loci associés au MVP prometteurs. L'annotation fonctionnelle fine à l'aide de données publiques a indiqué leurs rôle potentiels dans la régulation transcriptionnelle de plusieurs gènes candidats en (ex. *PDGFD* et *ACTN4*).

En résumé, mes travaux de thèse ont apporté des résultats génétiques originaux mettant en lumière de nouveaux mécanismes biologiques en rapport avec la biologie et le développement de la valve. Mes travaux ont également été renforcés par des validations dans des modèles animaux en collaboration. Il sera nécessaire toutefois de confirmer par réplication, et potentiellement par des expériences biologiques, les résultats nouveaux issus des travaux d'imputation haute densité afin de déclarer ces nouveaux gènes de prédispositions au MVP.

Mots-clés: GWAS, maladie de la valve cardiaque, développement de la valve, prolapsus de la valve mitrale, analyse d'enrichissement, imputation

Table of Contents

Publications

***Joint first authors.**

1. **Mengyao Yu**, Adrien Georges, Nathan R Tucker, Sergiy Kyryachenko, Katelyn Toomer, Jean-Jacques Schott, Francesca N Delling, Patrick T Ellinor, Robert A Levine, Susan A Slaugenhaupt, Albert A Hagege, Christian Dina, Xavier Jeunemaitre, David J Milan, Russell A Norris, Nabila Bouatia-Naji. **GWAS-driven Pathway Analyses and Functional Validation Suggest GLIS1 as a Susceptibility Gene for Mitral Valve Prolapse.** *Circulation: Genomic and Precision Medicine*; *accepted to published online 23rd April* (https://www.ahajournals.org/doi/10.1161/CIRCGEN.119.002497) Other publication: bioRxiv 433268; doi: https://doi.org/10.1101/433268

2. Katelynn Toomer*, **Mengyao Yu***, Diana Fulmer*, Lilong Guo, Kelsey Moore, Reece Moore, Ka'la Drayton, Janiece Glover, Neal Peterson, Sandra Ramos-Ortiz, Alex Drohan, Breiona J. Catching, Rebecca Stairley, Andy Wessels, Joshua H. Lipschutz, Francesca N. Delling, Xavier Jeunemaitre, Christian Dina, Ryan L. Collins, Harrison Brand, Michael E. Talkowski, Federica del Monte, Rupak Mukherjee, Alexander Awgulewitsch, Simon Body, Gary Hardiman, Starr E. Hazard, Willian Da Silveira, Baolin Wang, Maire Leyne, Ronen Durst, Roger Markwald, Solena Le Scouarnec, Albert Hagege, Thierry Le Tourneau, Peter Kohl, Eva Rog-Zielinska, Patrick T. Ellinor, Robert A. Levine, David Milan, Jean-Jacques Schott, Nabila Bouatia-Naji, Susan Slaugenhaupt, Russell A. Norris. **Primary Cilia Defects Cause Mitral Valve Prolapse**; Science Translational Medicine; accepted in principle.

3. Hafid Ait-Oufella*, **Mengyao Yu*,** Adrien Georges, Jeremie Joffre, Maire Vandestienne, Amir Boufenzer, Sébastien Gibot; Marc Derive, Nicolas Danchin, Tabassome Simon, Nabila-Bouatia-Naji. **Genetic determinants of soluble TREM-1 plasma levels in major adverse cardiovascular events prediction: FAST-MI 2010 study,** *in preparation to Circulation: Genomic and Precision Medicine*

Communications

Invited presentation

- The European Society of Cardiology (ESC) Congress. 25-29 August 2018. Munich, Germany

  Title: **Genetics of Mitral Valve Prolapse**

Oral and poster presentations

- **M. Yu**, C. Dina, N. Tucker, F. Delling, S.A. Slaugenhaupt, R.A. Levine, A.A. Hagège, J. Schott, X. Jeunemaitre, D. Milan, R. Norris, N. Bouatia-Naji: GWAS-driven Pathway Analyses and Functional Validation Reveals GLIS1 Predisposes to Mitral Valve Prolapse. American Heart Association (AHA) Scientific Sessions, New Orleans, Louisiana, United States, 2016 (**Oral Presentation**)

- **M. Yu**, C. Dina, N. Tucker, F. Delling, S.A. Slaugenhaupt, R.A. Levine, A.A. Hagège, J. Schott, X. Jeunemaitre, D. Milan, R. Norris, N. Bouatia-Naji: GWAS-driven pathway analyses and functional studies reveals GLIS1 to predispose to mitral valve prolapse. 27th European Days of the French Cardiology Society, Paris, France, 2016 (**Poster**)

- **M. Yu**, C. Dina, N. Tucker, F. Delling, S. Slaugenhaupt, R. A. Levine, A. A. Hagège, J. Schott, X. Jeunemaitre1, D. Milan, R. Norris, N. Bouatia-Naji: GWAS-driven pathway analyses and functional studies reveal GLIS1 to predispose to mitral valve prolapsed. The American Society of Human Genetics (ASHG) 2017 Annual Meeting, Orlando, United States, 2017 (**Poster**)

- **M. Yu**, Z. Alsalman, J. Schott, A. Hagege, X. Jeunemaitre, C. Dina, N. Bouatia-Naji: Up-dated Genome-Wide Association Study and Functional Annotation Reveal New Risk Loci for Mitral Valve Prolapse. The American Society of Human Genetics (ASHG) 2017 Annual Meeting, Orlando, United States, 2017 (**Poster**)

- **Mengyao Yu**, Adrien Georges, Nathan R. Tucker, Sergiy Kyryachenko, Patrick T. Ellinor, David J. Milan, Russell (Chip) Norris, Nabila Bouatia-Naji. GWAS-driven Pathway Analyses and Functional Validation Suggest GLIS1 as a Susceptibility Gene for Mitral Valve Prolapse. Heart Valve Society (HVS) Annual Meeting, Sitges(Barcelona), Spain, 2019 (**Oral Presentation**)

List of Abbreviations and Acronyms

DNA	Deoxyribonucleic acid
SNP	Single nucleotide polymorphisms
GWAS	Genome-wide association analysis
HRC	Haplotype Reference Consortium
TOPMed	Trans-Omics for Precision Medicine
NGS	Next generation sequencing
WGS	Whole genome sequencing
LD	Linkage disequilibrium
OR	Odds ratio
CI	Confidence interval
MAF	Minor allele frequency
HWE	Hardy–weinberg equilibrium
QC	Quality control
eQTLs	Expression quantitative trait loci
AS	Aortic valve
MV	Mitral valve
RVHDs	Right-sided valvular heart diseases
LVHDs	Left-sided valvular heart diseases
AVS or AS	Aortic valve stenosis
MR	Mitral regurgitation
MVP	Mitral valve prolapse
BAV	Bicuspid aortic valve

CAVS	Calcific aortic valve stenosis
ECM	Extracellular matrix
EMT	Endothelial to mesenchymal transition
VICs	Valvular interstitial cells
MFS	Marfan syndrome
IBD	Identical by descent
TF	Transcription factor

List of Figures

List of Tables

Part I – General introduction

1 Chapter 1 – The study of Human genetics

In this chapter, the human genome and related sequencing projects are first summarized with an emphasis on the Human Genome Project and to give an overview of human genetic variants. The sequencing technologies are outlined which were developed for revealing the DNA structure. Then, the SNP genotyping, an efficient and economical approach to determine the genetic variations of SNPs among individuals, is introduced. The basic concepts in population genetics are also presented. the two important analyses in genetic epidemiology, linkage analysis, and association analysis, are summarized. I put more stress on the introduction of genome-wide association analysis, as it is the approach that we used to study non-syndromic mitral valve prolapse. Finally, the functional strategies to extract biological functions from GWAS findings, including in silico annotations, pathway-based analysis using GWAS summary data, and network analysis for GWAS data, are summarized.

1.1 Detecting human genetic variability

1.1.1 The human genome and related sequencing projects

The human genome contains 23 pairs of chromosomes, in which approximately 3 billion of these base pairs reside. According to recent database, 20,805 and 20,418 estimated protein-coding genes were identified in GRCh37.p13 and GRCh38.p12, respectively. Coding DNA sequences are defined as the sequences that can be transcribed into mRNA and then translated into proteins. Coding DNA sequences account for only a very small fraction of the genome (< 2%). More than 98% of human genomes is composed of noncoding DNA sequences (Lander et al. 2001), which are not found within protein-coding exons and are never represented within the amino acid sequence of expressed proteins. Various classes of noncoding DNA have been identified, e.g., genes for noncoding RNA (e.g. tRNA and rRNA), pseudogenes, introns, untranslated regions of mRNA, regulatory DNA sequences, repetitive DNA sequences, and sequences related to mobile genetic elements.

To determine DNA sequences and map all of the genes of the entire human genome from a physical and functional point of view, the Human Genome Project (HGP) has made the effort to reveal about 20,500 human genes (Lander et al. 2001). Having the complete sequence of the human genome is essentially similar to having a manual making a human body. The challenge to researchers now is to determine how to interpret the manual to discover the genetic basis for health and of the pathology of human diseases. Based on this project, some online tools, such as UCSC Genome Browser (Kent et al. 2002) and Ensembl (Zerbino et al. 2018) which have added additional data, provide a powerful function to visualize and search in the human genome. Taking advantage of this project, the sequence of DNA helps us to study genetic variants in the human genome which enables a better understanding of human diseases and personalized precision medicine.

In DNA sequences, there are many different types of genetic variants which can be classified in three forms in general, 1) single nucleotide polymorphisms (SNPs), 2) structural variation (indel (insertion or deletion), duplications, copy number variation, inversions) and 3) chromosome abnormality (Burton et al. 2005). Structural variation is used to describe the genetic variation that occurs over a larger DNA sequence compared to SNP which represents a variation in a single nucleotide. The number of currently known SNPs in the human genome is 655,379,774 (dbSNP Build 152). According to whether SNPs modify the aminoacidic sequence in the gene product, they are classified into non-synonymous and synonymous (Burton et al. 2005). Non-synonymous SNPs in the protein-coding regions can be called coding SNPs (Burton et al. 2005) and those SNPs are more likely to affect the function of a protein (Carlson et al. 2004). However, disease-associated variants with modest effects should be distributed proportionately between noncoding and coding sequences (Carlson et al. 2004) and all kinds of SNPs including synonymous SNPs can cause diseases through different ways, e.g., through changing the regulation of transcription (Burton et al. 2005).

To address most genetic variants, the 1000 genome project (Consortium et al. 2015) provides highly accurate assignment of the genotypes from 2,504 samples, which contains 84.4 million variants in 26 populations (Phase 3). On the other hand, to determine the haplotype map of the human genome and the LD patterns of DNA sequence variation, the HapMap project

(Consortium 2003) genotyped 5000,000 "tagSNPs" and provides a powerful source for studying the genetic variations. HapMap might be replaced by the 1,000 genomes, as the latter has become a research standard for population genetics and genomics gradually with an increasing number of novel variants.

1.1.2 Next-generation sequencing and its applications

As was started for discovering DNA structure, sequencing is widely used to understand the complexity and diversity of human genomes (Goodwin et al. 2016). Compared to the Sanger sequencing, the "next generation" sequencing (NGS) technologies enable rapid generation of a large amount of data by sequencing enormous amounts of DNA or RNA samples in parallel using diverse methodologies with lower costs (Rizzo et al. 2012).

NGS sequencing can be divided into short-read NGS and long-read NGS. SOLiD, Complete Genomics platforms, Illumina, Qiagen GeneReader, Ion Torrent, and Roche 454 platforms use short-read sequencing approaches (Tomkinson et al. 2006) (Goodwin et al. 2016). Both SOLiD and Complete Genomics platforms offer a very high accuracy (~99.99%) (Liu et al. 2012), but the maximum read length is only 75bp for SOLiD and 28-100bp for Complete Genomics, which limits their use for genome assembly and structural variant detection applications (Goodwin et al. 2016). Illumina has a series of platforms and dominates the short-read sequencing industry. HiSeq X as the newest and highest throughput instrument is still limited to whole genome sequencing (WGS) and whole-genome bisulfite sequencing. Qiagen GeneReader focuses on cancer gene panels (Goodwin et al. 2016). Both the Ion Torrent systems and the 454 platforms have longer read lengths, up to an average of 400bp and 700bp, respectively. Unfortunately, the 454 platforms had been discontinued in 2016 as they were unable to adapt to the rapid development in the NGS field.

The complexity of genomes lies in their long repetitive elements, structural variations and copy number variants (CNVs), which are too long to be sequenced by short-read paired-end technologies. Long-read sequencing enables a single continuous read, which is useful for complex or repetitive regions and transcriptomic (entire mRNA transcripts) research (Goodwin

et al. 2016). Several platforms can be used to sequence long read, such as Pacific Biosciences (PacBio) (Eid et al. 2009), MinION (nanopore sequencer) and Illumina platforms. The PacBio RS II can generate single polymerase reads over 50kb with average read lengths of 10-15kb for a long-insert library. However, the single-pass error is as high as 15%, especially indel errors (~14%) with few apparent miscalls (~1%) (Carneiro et al. 2012). Even though this limitation can be overcome by increasing the coverage (Koren et al. 2012), the high cost (around $1,000 per Gb) and limited throughput still hinder the wide use of PacBio RS II. MinION can directly detect the DNA composition of a native ssDNA molecule, but it has a large error rate (up to 30% for a 1D read) (Goodwin et al. 2016). Nevertheless, the accuracy limitation may be overcome by using good calling algorithms (Jain et al. 2015). The Illumina synthetic long-read sequencing platform and the 10X Genomics emulsion-based platform are now available for generating more than 100kb long-reads (Goodwin et al. 2016).

A wide range of applications based on NGS suite includes genome sequencing through WGS, whole-exome and targeted sequencing (Hodges et al. 2007), epigenomics applications through chromatin immunoprecipitation followed by sequencing (CHIP-seq) (Park 2009), ATAC–seq (assay for transposase-accessible chromatin using sequencing) (Buenrostro et al. 2013), DNA methylation sequencing (methylseq) (Brunner et al. 2009), and transcriptomics applications through RNA sequencing (RNA-Seq) (Wang et al. 2009). Long-read sequencing approaches are more effective in identifying clinically relevant structural variation than short-read approaches but are less used.

Using NGS technology, a large number of sequencing data were generated. Web-accessible databases like Encyclopedia of DNA Elements (ENCODE) (Consortium 2012) and UCSC (Kent et al. 2002), integrating the approaches and sequencing data to provide a comprehensive annotation of functional elements in the human genome and offer a visualization framework. With the increasing number of sequencing data, one challenge now is how to analyze and store this wealth of data (Pop et al. 2008; Goodwin et al. 2016). The transformation of this vast genetic database into a biological context associated with genomic information remains challenging. One important difficulty is that, as the cost per sample is still high, the sequencing technology cannot be applied for large scale population studies, such as GWAS.

1.1.3 SNPs genotyping

SNPs, as a type of genetic markers, defined as a DNA sequence that is sufficiently heterogeneous to identify differences between individuals and between homologous chromosomes in an individual, exist in the human genome and were found to be the contributors to the etiology of many human diseases. SNP genotyping is the process to determine the genetic variations of SNPs among individuals. A variety of approaches can be used to perform SNP genotyping, such as SNP array and sequencing.

SNP array, a type of DNA microarray, is used to detect the SNPs within a population. In high-density oligonucleotide, SNP arrays, by generalizing to a very large number of immobilized probes on a small chip, can be used to detect a large number of SNPs simultaneously. The most successful and commonly used SNPs genotyping arrays mostly come from Affymetrix and Illumina (Bumgarner 2013). Nowadays, SNP arrays can detect more than 1 million different human SNPs. Call rates and reproducibility of SNP calls exceed 99.5% (Bumgarner 2013).

One important application of SNP arrays is to determine disease susceptibility loci/genes, using the genome-wide association analysis studies (GWAS). The genome-wide SNP genotyping using fix SNP panel provides high data quality, accurate genotyping results to detect markers for genetic diseases with complex traits in a cost-effective way. However, SNP genotyping arrays are not designed for detecting rare variants. When there are no suitable assays, target sequencing can be used for SNP analysis and detection. Moreover, genome-wide SNP genotyping array only contains a small fraction of human genetic variants. The ungenotyped variants can be detected by genetic imputation.

Sequencing can be carried out for SNP genotyping as well, such as WGS and restriction-site-associated DNA sequencing (RAD-seq) (Davey et al. 2010). WGS is an efficient approach to detect SNPs, but the cost is overwhelming as a huge amount of data is generated. RAD-seq can identify genetic variants in a short fragmental region of the genome using a restriction nuclease and an NGS platform. It is useful to detect SNP genotypes in a specific region than sequencing large regions, but not applicable to do genome-wide SNP genotyping.

1.2 Population genetics

Population genetics is the study of genetic variation within populations. It aims to understand the changes in the genetic composition by detecting changes in the frequencies of genes and alleles in populations. In an ideal situation, the genotypic proportions are constant from generation to generation. For example, a single locus with two alleles denotes A and a with frequencies f(A) = p and f(a) = q. The expected genotype frequencies for AA, Aa and aa should be f(AA) = p^2, f(aa) = q^2, and f(Aa) = $2pq$, respectively. After several generations, the frequencies of allele A and a should still be p and q, which suggests that the subsequent generations have reached the equilibrium, called Hardy-Weinberg equilibrium (HWE) (Wigginton et al. 2005). However, the genetic composition can be changed by various factors, including natural selection, mutation, random genetic drift, and migration into or out of the population. The individuals composing a given generation have unequal chances of transmitting their gametes to the next generation and some genotypic variants are more likely to survive or reproduce than others. This natural selection may be due to differences in mortality before reproductive age, or reduced fertility. Mutations in the population keep the population from homogeneous. It is the source of genetic variation and ensures that the allele will not be fixed in the population and a degree of polymorphism will be maintained. When genetic drift happens, allele frequencies in a population will be changed. This change can make genetic variants to disappear or more common depending on their reproductive success. The genetic composition of the population can be changed as well when migration occurs between or among different populations (Okasha et al.2016).

If any of these conditions happened during the evolution, the HWE will be broken which causes deviations from the expectations. HWE test is usually used to test the difference between observed genotype frequencies obtained from the data and the expected genotype frequencies obtained. The deviation from HWE at particular markers may suggest an association between the marker and susceptibility disease (Wigginton et al. 2005).

1.3 Genetic epidemiology

Genetic epidemiology is a discipline closely related to traditional epidemiology, focusing on the genetic determinants of human diseases, and the combined effects of genes and non-genetic determinants (Burton et al. 2005). Before the boost of genomic sequencing, a great breakthrough was made on monogenetic diseases. These diseases are affected by a mutation in a single gene and they are recurrent in the familial following the laws of Mendelian inheritance (Burton et al. 2005). However, with a better understanding of human genomes and the development of sequencing and genotyping technologies, we are able to detect multiple genetic determinants in a family or in the population that is associated with the disease or trait, referred as complex disease.

In this part, we introduce the genetic linkage analysis, which estimates whether a given marker or a marker close to a causal variant is shared within a family. Then, we focus on association analysis, which looks for the association between variants and the disease or trait of interest in the general population.

1.3.1 Linkage analysis

Genetic linkage analysis is used to map genetic loci for the Mendelian and complex traits with familial aggregation (Dawn Teare et al. 2005; Bush et al. 2010). Genetic linkage analysis aims to detect a chromosomal region that is likely to contain disease/trait-related genes. In the genome, especially in the same chromosome, the closet loci are more likely to segregate together through pedigrees, offering the opportunity to map genetic markers and infer the position within a locus.

Linkage analysis can be either parametric or non-parametric. Parametric linkage analysis is a traditional approach used when an inheritance model for the trait locus (e.g. dominant or recessive), allele frequencies, mutation rate, and penetrance is specified (Dawn Teare et al. 2005; Bush et al. 2010). The linkage can be measured by the logarithm of the odds (LOD) score, developed by Newton Morton (Morton 1955). This score mostly depends on the recombination fraction (θ), which measures the probability of recombination of the gametes by parents. If the gametes are transmitted independently, $\theta=0.5$. Otherwise, the parenteral gametes are more likely

to be transmitted preferentially to the recombined gametes, then $\theta < 0.5$ in which case those two loci are in the linkage. With the null hypothesis that there is no linkage between a putative disease locus and the marker locus (a single or multiple markers), the LOD score ranges from 3 to -2. Higher LOD score represents greater linkage. A score of -2 means there is no linkage between the disease locus and the maker.

Without accurate parameter estimates, the setting of different parameters for linkage studies will lead to confusion. It is hard to decide which set of parameters provides better results than another. In addition, false positive results may appear in multiple tests without adjustment (Bush et al. 2010).

For polygenic diseases, the inheritance model is not that clear anymore. Model-free (non-parametric) linkage analysis was developed to deal with these unspecified disease models. Affected relatives always share haplotypes that are identical by descent (IBD) in the region of a disease-causing gene without the mode of inheritance (Dawn Teare et al. 2005). Under the null hypothesis of no linkage, various methods were employed to test whether IBD sharing at a locus is greater than expected. The study of the affected sibling pairs is helpful to solve the problem caused by incomplete penetrance by using definitively affected individuals. But the distinction of IBD and identical by descent (IBS) is not easy when the markers are not sufficiently polymorphic, and the proper sibling with a disease is hard to find, which limits the application of this approach (Dawn Teare et al. 2005). Alternative methods have been developed to analyze families with a large number of affected relatives. Estimated IBD sharing score can be calculated by pairwise comparisons, then compared to the expected score to get the linkage relationship. The IBD sharing can be more accurately calculated by more multipoint analyses. Using several thousand SNPs, whole genome screens were created for a wide range of complex diseases (John et al. 2004) and MERLIN, a wide and popular software can be used to handle such data easily (Abecasis et al. 2002).

Linkage analysis is limited to complex disease studies so far due to the small sample size, and hence it cannot detect modest effect genes. Besides, whether identify a region linked to a disease gene or discover a shared region between affected individuals, only a region that shows linkage with a disease, a causal gene or loci cannot be detected (Dawn Teare et al. 2005; Hirschhorn et al.

2005). However, detecting a particular region of the genome using linkage analysis can be a proper prior step for target sequencing (Teare et al. 2014). Additionally, genes discovered by linkage analysis usually explain only a small part of the heritability of the diseases, even though linkage analysis might be more productive by using dense markers, larger pedigree and larger sample sizes (Hirschhorn et al. 2005). Recently, linkage analysis coupled with filtering approaches turned to be a powerful method dealing with WGS data, especially for rare variants in a complex trait with high penetrance (Ott et al. 2015). Nevertheless, common variants with modest effects on diseases are difficult to identify by linkage study. Those limitations may give us additional clues to association analysis.

1.3.2 Genetic association analysis -- Population studies

Extensive genetic variants in large, population-based samples detected in human genome sequencing give us the opportunity to understand the pathogenesis of common diseases. Differ from linkage analysis, association analysis studies the relationships between one or more common genetic variants and diseases (Cordell et al. 2005). Rather than other types of genetic polymorphism, SNP is widely studied in the association analysis. There are more than 84.7 million SNPs in the human genome. At least 99% of SNP variants have a frequency greater than 1% (Consortium et al. 2015).

Association study can be used as candidate genes association study if the associated region is well predicted. This study heavily relies on biological literature or the location provided by the linkage study. An alternative association study is GWAS, which is, in theory, expected to identify the causal genetic variants by testing the association between a large number of variants and the trait (Hirschhorn et al. 2005). In GWAS, compared to linkage analysis, there is no need to guess the location of the potential risk loci which may cause bias. Chances to discover the risk variants and genes are also increased as a few hundred thousand positions are tested.

1.3.2.1 Select SNPs for genotyping

The tested association can be classified into direct and indirect association (Cordell et al. 2005) (**Figure 1**). The direct association is powerful but difficult to determine the candidate SNP (**Figure 1a**). The selected SNPs for direct testing for the association are always functionally known by previous knowledge. When an association happens in the codon region, it may alter the DNA sequence directly and affect the candidate gene, such as missense SNPs (Cordell et al. 2005; Hirschhorn et al. 2005). However, most of the association are located in the non-coding region. Variants in the non-coding region may be responsible for the heritability of common complex disorders as well as the coding region (Cordell et al. 2005). Thus, the direct association has a limited potential to discover genetic causes in GWAS.

The second type of association is indirect association (**Figure 1b**), which has no causal role but is linked to a nearby causal variant. It is difficult to analyze the indirect association because they are less powerful than direct associations (Cordell et al. 2005). A large number of SNPs can be genotyped to increase the possibility to find the indirect association. However, the high cost in genotyping lots of SNP kept researchers away (Halldórsson et al. 2004). Fortunately, based on the special haplotypic structure in the human genome, only a small number of SNPs need to be genotyped (Flint 2013).

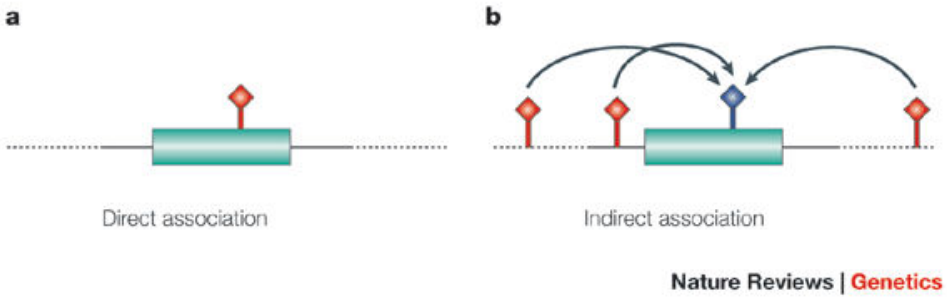

Figure 1. Direct and indirect association. *a) candidate SNP (red) is directly tested for association with a disease phenotype. b) The SNPs to be genotyped (red) are in the linkage disequilibrium (LD) with the tested SNPs (blue). Adapted from (Hirschhorn et al. 2005).*

In the population, alleles at different loci are more likely to be inherited together than by chance. The set of the special allele called haplotype and alleles inherited together in the haplotype is in the LD (Consortium 2003) (**Figure 2a,b**). In the same chromosome or different chromosomes, the level of LD between alleles mostly depends on the recombination rate (Slatkin 2008). The distance between genes is proportional to recombination rate but is inversely proportional to the linkage score. Of course, when the two loci are in a complete linkage state, there is a strong LD between the alleles (Slatkin 2008). LD can be measured by calculating D' (Lewontin 1964) and r2 (Hill et al. 1968). When LD happens in alleles or genes, it suggests that they are in a potential association. The strong association can represent measurable haplotypes in many chromosome regions which can explain most of the variation in those regions (Daly et al. 2001; Gabriel et al. 2002). By using this strong association, a set of carefully selected SNPs, 'tag' SNPs, contains most information for that region. Therefore, genotyping a few 'tag' SNPs in those regions is sufficient to identify their correspondent haplotypes (**Figure 2c**).

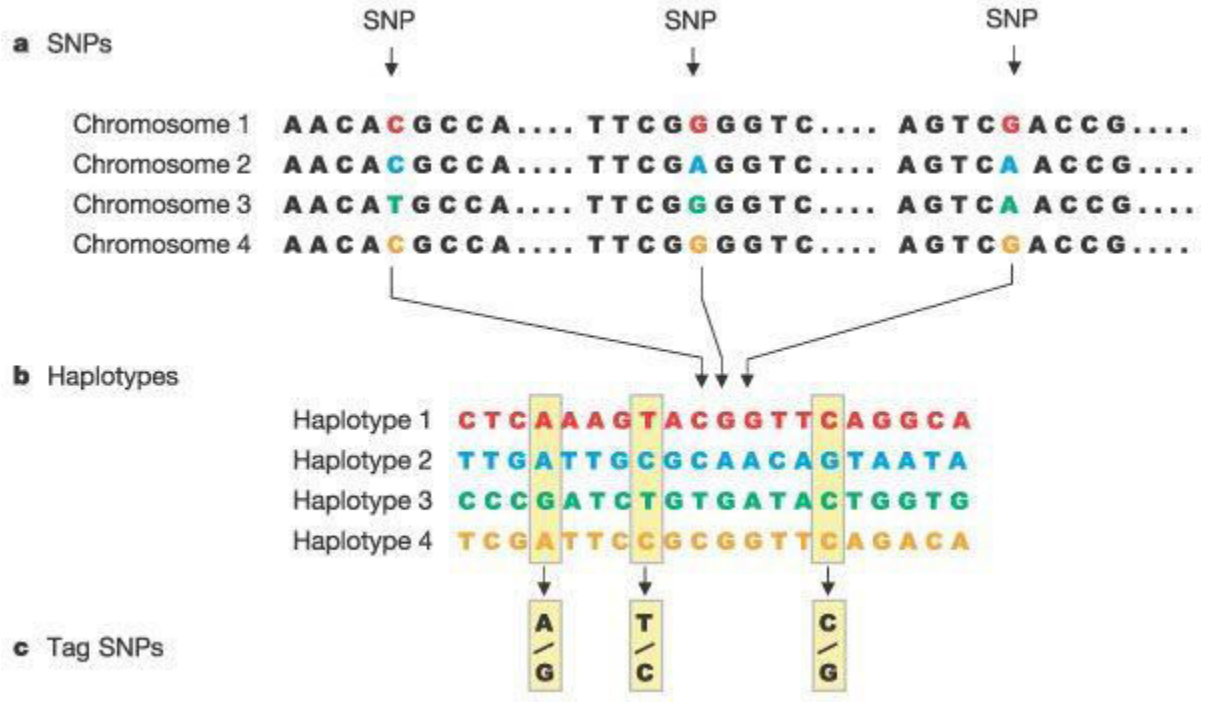

Figure 2. The selection of 'tag' SNPs. *a) In different populations, there are SNPs in the chromosome; b) a set of alleles form haplotypes; c) several SNPs can be selected to represent those SNPs. Adapted from (Consortium 2003).*

Taking advantage of the LD patterns, the HapMap project helps to identify a set of 'tag' SNPs to improve the efficiency by not genotyping all markers in the genome. The tag SNPs contain most of the information of the common variation within the region according to the LD (Johnson et al. 2001; Gabriel et al. 2002).

1.3.2.2 Key steps in GWAS -- Phasing and imputation

As expected, a set of genotyped polymorphisms provided by different GWA arrays achieves a modest power to detect the association. However, variants not being directly genotyped may play a role in the disease. Using imputation, which aims to predict the genotypes that are not directly assayed in the studied individuals and fill in missing genotype data, can lead to a boost of the association signaling and an improvement in the detection of associations (Marchini et al. 2007).S Haplotypes are related to each other because IBD exists among the unrelated individuals. By identifying the sharing between the haplotypes of the studied individuals and the haplotypes in the reference panel, imputation can infer the missing alleles that are not included in the commercial arrays (**Figure 3**).

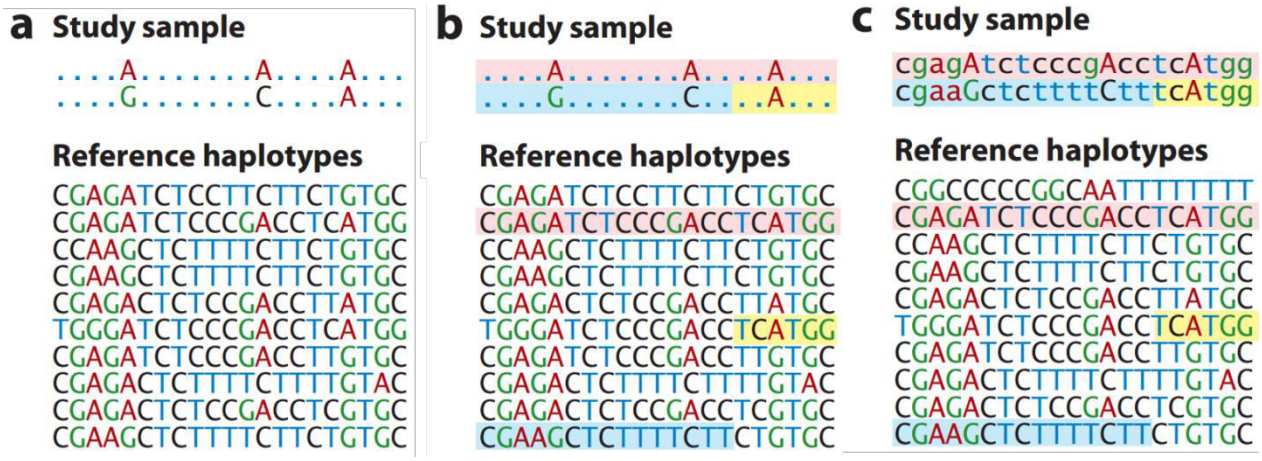

Figure 3. Genotype imputation in unrelated individuals*. a) The genotyped data still has missing genotypes. However, the reference panel with a large number of haplotypes is available. b) Identify the sharing region between study samples and the reference panel. c) Combine the study sample with reference haplotypes to filling the missing genotypes in the study sample. Adapted from Li et al. (Li et al. 2009).*

1.3.2.3 Phasing and pre-phasing

Phasing is the process that uses genotype data to estimate haplotypes for an individual from the reference panels (phasing) or GWAS sample (pre-phasing) (Figure I-3a,b), which can improve the imputation speed and accuracy (Delaneau et al. 2012; Howie et al. 2012). Phasing unrelated individuals considers the haplotype frequencies. The more individual samples there are, the better the phasing estimation is (Browning et al. 2011). Haplotype phase plays an important role in the imputation of untyped genetic variation. Generally, the more haplotypes the panel have, the higher the accuracy in the imputation is. However, imputing from a large reference panel is computationally challenging. Computational methods were developed to handle the large genotyped data. The most popular computational phasing method uses hidden Markov models (HMM) based on approximate coalescent models (McVean et al. 2005).

Under the HMM, approaches were designed to deal with population-based statistical phasing, like PHASE (Stephens et al. 2001; Stephens et al. 2005), SHAPE-IT (Delaneau et al. 2008), fastPHASE (Scheet et al. 2006), IMPUTE2 (Howie et al. 2009), and MACH (Li et al. 2010). PHASE provides accurate phasing in small numbers of markers with small sample sizes, but it is slow in large data size. fastPHASE is less computationally costly and can handle large unphased genotype data at the expenses of losing accuracy compared to PHASE. MACH and IMPUTE2 are imputation programs but can be used for phasing observed genotypes as well. Both methods have a high accuracy as the sample size increases. But IMPUTE2 is more time-consuming. SHAPE-IT was developed based on PHASE. This tool provides the same phasing accuracy but runs faster than PHASE. The updated SHAPEIT2 combined the advantages of SHAPEIT1 and IMPUTE2 and can deal with GWAS data. Moreover, it has better computational performance, accuracy, and less switch error rate compared to other methods for the sample sizes ~9,000 (Delaneau et al. 2013). Additionally, Eagle2 was reported to have faster speed and better accuracy compared to SHAPEIT2 with a lower switch error rate in the European-ancestry (Loh et al. 2016).

Due to large genotype databases, the phasing haplotypes of reference panels or GWAS samples need only to be done once. Matching a phased GWAS haplotype to one reference haplotype is

faster compared to matching the unphased GWAS genotypes (Howie et al. 2012). All these efforts are meant for a proper imputation.

1.3.2.4 Reference panels

Selecting a good reference panel is important to a successful imputation. Available reference panels include HapMap3 (Consortium et al. 2010), 1000 genomes phase 3 (Consortium et al. 2015), UK10K(Huang et al. 2015), The Haplotype Reference Consortium (HRC) (McCarthy et al. 2016), and The Trans-Omics for Precision Medicine (TOPMed) (Taliun et al. 2019). HAPMAP3 has 2,368 haplotypes from the genotyped 1.6 million common SNPs in 1,184 reference individuals from 11 global populations. Compared to HAPMAP3, 1000 genomes phase 3 provides a more comprehensive map of the genetic variations, which was used as the imputation reference panel in the current GWAS of MVP (Dina et al. 2015). This project combined data from deep exome sequencing, low-coverage whole-genome sequencing, and dense microarray genotyping. Finally, 5,008 haplotypes were predicted from 49 million SNPs in 2,504 individuals from 26 populations. Recently, HRC combined 20 low-coverage whole-genome sequencing data (including the UK10K) and created a large reference panel of 64,976 haplotypes that has over 39 million variants from European and 26 worldwide populations. TOPMed provides deep WGS data (39 X) and other omics data to get more than 400 million single nucleotide and short insertion/deletion variants (97% have frequency <1%) on ~53k samples.

Large reference panel contains more genetic variants, which increases the imputation accuracy and improve the imputation density. This leads to an improvement in the power of association analysis which increases the possibility to detect causal variants.

1.3.2.5 Imputation

Imputation is a cost-efficient way to map the variants that are not directly genotyped in the SNP genotyping (Figure I-3c). A high-resolution map provided by imputation increases the chance to detect causal SNPs, even non-SNPs (e.g. indels/deletions, CNV) (Marchini et al. 2010). With the emerging of large reference panels, tools with a low computational cost, but have high

imputation accuracy are indispensable. Several methods can achieve this goal, such as minimac3 (Das et al. 2016), IMPUTE2 (Howie et al. 2009), Beagle 4.1(Browning et al. 2016) and MaCH (Li et al. 2010). Using different reference panels, Das et al. (Das et al. 2016) reported that minimac3 has the best performance among the aforementioned tools in imputation speed, computational consuming and imputation accuracy (especially in rare variants (minor allele frequency (MAF) = 0.0001-0.5%)).

Based on the correlation between the true and the imputed dose of allele, r^2 (RSQ) measure is used to interpret the imputation accuracy (Howie et al. 2012). For Minimac and Beagle, the RSQ measure is based on the correlation between true and estimated haploid dosages across all haplotypes. Another measure, MaCH RSQ measure, is based on diploid dosages across all samples (Das et al. 2018).

Although using the aforementioned large reference panels is valuable in phasing and imputation, a few limitations need to be mentioned. A number of factors, e.g., sample size, marker density, genotype accuracy, relatedness in the sample, ethnicity and allele frequency (Browning et al. 2011) influence the computational phasing and imputation accuracy. Phasing on the different reference panels for imputation may cause different effects on association results. Meanwhile, phasing accuracy in the reference panel and the GWAS SNPs used for imputation can affect the imputation accuracy.

Imputation usually requires high computational cost and time consumption. In addition, direct access to large reference panels like HRC and TOPMed is limited. Instead, one needs to access a web-based imputation server to upload the genotyping data. Those servers perform the imputation with the predefined quality control process, such as checking and transforming the genome build between hg19 and hg38, flipping the strand, switching the allele, filtering the SNP with low call rate, removing multiallelic and duplicated sites. We also can select the interested reference panel. After the imputation, the imputed data and detailed information files can be downloaded. Thus, one can focus on analyzing the imputed data and interpreting the results, rather than on learning how to do the imputation. Currently, two imputation servers are broadly used, Michigan Imputation Server (Das et al. 2016) and Sanger Imputation Server (McCarthy et al. 2016).

1.3.3 Association analyses and applications

As long as well-mapped SNPs are genotyped or imputed, a genetic association that compares alleles or genotypes at the genetic markers becomes important to identify the genetic risks.

In a case-control study, the strength of an association can be measured by the odds ratio (OR), which is calculated from the allele or genotype frequencies. OR is the ratio of odds of the disease group to the odds of the control group. There are two alleles (A, a) and three possible genotypes (a/a, a/A, A/A). By comparing the odds of disease in an individual carrying allele A to the odds of disease in an individual carrying allele a, the allelic OR describes the association between the disease and allele. The genotypic ORs represent the association between the disease and genotype by comparing the odds of disease carrying different genotypes. Usually, there are two genotypic ORs, A/A vs a/a and a/A vs a/a (Clarke et al. 2011).

For disease penetrance, several models (additive, recessive, dominant and multiplicative models) (Lewis 2002) imply the relationship between the genotype and phenotype. Chi-squared (χ^2) test, Cochran-Armitage trend test, and logistic regression are the widely used association tests in the models. A single test can be used to calculate for each individual SNP. Usually, the null hypothesis is that there is no association between allele (or genotype) and the disease. The frequencies of allele or genotype in cases are supposed to equal to controls.

In a χ^2 test for association, the frequencies of allelic or the genotype can be collected into a two-dim or three-dim contingency table (Clarke et al. 2011). To test a dominant model of penetrance, only allele A increases the risk. To test the recessive model of penetrance, two copies of A are required for the increased risk. For a multiplicative model of penetrance, there is an r-fold risk for Aa, r2-fold risk for aa. It is necessary to analyze using alleles rather than using genotypes in the multiplicative model. However, genetic penetrance of many complex diseases is unknown. The additive model is likely to perform well and is commonly used (McCarthy et al. 2008; Clarke et al. 2011). For the additive model, there is an r-fold increased risk for Aa, 2r-fold increased risk for aa, which is always examined using the Cochran-Armitage trend test (Clarke et al. 2011).

Logistic regression models of association are used when additional covariates are needed to handle complex traits and environmental interactions or covariates, such as sex or age of onset (Balding 2006; Clarke et al. 2011). We can calculate the transformation logit (p) ~ β0 + β1X, where p is the expected value of phenotype given by the genotype X of the individual. When the potential confounders, such as ethnicity (E), genotyping batch (G), age (A), or sex (S) are needed to adjust the association, we can use the logistic transformation logit f(p) ~ β0 + β1X + β1E + β1G + β1A + β1S…. In the logistical model, the value of β1 is estimated. If β1 significantly differs from zero, it indicates the association between the SNP and disease (Clarke et al. 2011).

PLINK and SNPTEST can be used to perform association analyses. PLINK (http://zzz.bwh.harvard.edu/plink/) is a free, open source, genome association analysis toolset, which can rapidly manipulate and analyze large data sets comprising thousands of markers genotyped for thousands of individuals (Purcell et al. 2007). PLINK has five main functions: data management, summary statistics, assessment of population stratification, association analysis, and IBD estimation. All of these functions are applicable to GWAS. Different from PLINK, SNPTEST is a program for the analysis of SNP association in GWAS with frequentist association test and Bayesian association test (Marchini et al. 2007). SNPTEST can be carried out for both quantitative and binary traits and can work on user-specified covariates.

1.3.4 Multiple testing

Using GWAS, millions of SNPs have been tested. The proportion of false positive (type I error) results will be increased if P-value is still set to 0.05. To control the accuracy of multiple testing, the determination of a proper significance threshold is a very important aspect in GWA studies.

For association tests applied at each of n SNPs, the significance levels α* for a given family-wise error rate of α (for example, α = 0.05) can be simply approximated using Bonferroni adjustments (α* = α/n). Assuming independence between markers, Bonferroni adjustments provide a simple way to adjust for multiple testing. Permutation testing (Dudbridge et al. 2008) is an accurate correction, which applies only to the genotyped data set. Therefore, the permutation approaches cannot generate truly genome-wide significance thresholds unless the entire genome is

sequenced. The empirical genome-wide significance threshold as strong evidence for association accepted was 5×10^{-7} and recently, 5×10^{-8} (Consortium 2007; Pe'er et al. 2008). Further, a quantile-quantile plot is used to assess whether the observed P-values are consistent with or deviate from expected P-values (Weir et al. 2004). Nevertheless, a significant result in an association test does not imply that the SNP directly affects the disease risk. A direct, or causal association occurs when different alleles at the marker locus are directly involved in the etiology of the disease through one or more biological pathways.

1.3.5 GWAS based Meta-analyses

To guarantee the outcome of the association analysis, a multi-stage approach to minimize sample size and reduce the amount of required genotyping can be used. In stage one, the full set of SNPs are genotyped and the second stage is to type and analyze SNPs that seem promising from the first stage. The third stage can be used as well to re-test the selected SNPs and improve the stringency. This strategy would minimize the chance of false associations (Hirschhorn et al. 2005). However, the common alleles own small genetic effects and the detection of association signal needs a large sample size (Moonesinghe et al. 2008; Evangelou et al. 2013). Using a single GWAS is insufficient even if we use a multi-stage approach.

Meta-analysis, a statistical synthesis of information from multiple independent GWASs can increase the power of association signal and reduce false-positive findings. Large meta-analyses have dramatically increased the yield of discovered and validated genetic risk loci. Several approaches exist for GWAS meta-analysis and they are classified into P-value and Z-scores, fixed effects, random effect, optimal weights, and Bayesian-based meta-analysis (Evangelou et al. 2013). P-values based meta-analysis was used in different fields but hardly provides an overall estimation of effect size and finds heterogeneity issues. Z-scores based approaches, however, take the direction of the effect into account. Heterogeneity resulting from differences in study designs, differences due to interactions with other SNPs or differences due to environmental or lifestyle factors that influence the effects of genes is a potential issue in meta-analysis. (Schaid et al. 2018).

1.3.6 Important factors for a good genetic association study

A good GWAS is affected by several factors. The sample size is the key determinant. But a large sample size may lead to population structure, which is another risk factor in GWAS (Marchini et al. 2004). When genetic markers in cases and controls have different frequencies due to systematic ancestry differences, it can cause false-positive association, spurious associations or over-representation in association analysis (Hirschhorn et al. 2005). The principal-components analysis provides a fast and effective way to detect and correct the population structure (Patterson et al. 2006). Alternatively, mixed-model approaches can meet the requirements in mixed populations studies (Rosenberg et al. 2010). However, a good study design and the proper study sample size is not the only guarantee for a good GWAS. The quality of genotyping, imputation accuracy, the quality control process affect the association results also. Besides, a success GWAS also depends on the number of the segregated loci that affect the trait in the population, the effect size and allele frequency of these loci and the heterogeneous rate of the interested trait or disease (Visscher et al. 2017).

In summary, GWAS has achieved great success in many fields and has reported hundreds of complex traits (Visscher et al. 2017). According to the NHGRI-EBI GWAS Catalog, about 5687 studies and 71673 variant-trait associations from 3567 publications have been reported between genetic variants and one or more complex traits (Buniello et al. 2019). There is a boost of the discovery of new loci and the power of recent GWAS using larger sample sizes (Buniello et al. 2019). About 702 strong associations have been reported between genetic variants and cardiovascular disease, where “strong” is defined as statistically significant at the genome-wide P-value threshold of 5×10^{-8} (NHGRI-EBI GWAS Catalog).

Nevertheless, most genetic variants detected by GWAS are common in the population and have a MAF larger than 1%. The debate that whether a few variants with strong effects or a large number of variants with small effect lead to the inherited component has lasted for a long time. A large number of findings observed in the GWAS on the complex diseases, including most of the cardiovascular disease, are common variants with small effect (Schunkert et al. 2018). For complex traits, a possible combination of the sets of alleles that can increase or decrease the

disease risk plays a critical role. Moreover, many studied traits associated with variants at hundreds to thousands of loci in the genome strongly suggest that different traits may share the same underlying causal variants (Visscher et al. 2017). This so-called pleiotropy is a common characteristic and is identified as the phenomenon that one gene affects more than one phenotype (Erdmann et al. 2018).

1.4 Strategies to extract biological functions from GWAS findings

1.4.1 In silico annotations and enhancers annotations, eQTLs …etc

As GWAS has identified disease or trait associated loci, questions are arisen, 1) As most of the leading associations are located in the non-coding regions, what are their functions and how do they affect the disease? 2) Are those lead SNPs causal SNPs? If not, which one will be the potential causal SNPs? To answer the first two questions, functional annotations need to be carried out. As we know, only a small number of disease-associated SNPs lies in protein-coding regions and 88% of the associated SNPs are either intronic or intergenic (Hindorff et al. 2009). Depending on the location, the associated SNPs can influence gene activity through different mechanisms (**Figure 4**). Most SNPs identified by GWAS located in the non-coding region showed strong evidence of enrichment for regulatory elements, such as enhancers, promoters, insulators and silencers (Schaid et al. 2018). Those SNPs might act as "*cis*-regulatory elements" (CREs) that influence the gene expression by transcriptional, posttranscriptional, and posttranslational mechanisms (Edwards et al. 2013). It has been shown that disease-related variants are enriched in CREs, typically defined by transcription factor (TF) binding, open chromatin (DNase-seq, FAIRE-seq), and histone modifications known to be associated with transcriptional regulatory activity, such as H3K27ac, H3K4me1, and H3K4me3 (Edwards et al. 2013; Gallagher et al. 2018). Some regulatory elements are located far away from their target genes (>1000 kb), and the regulation of the transcription goes through the long-range interactions by the chromatin loops (Sexton et al. 2009). With the development of chromosome conformation capture methods such as Hi-C, the analysis of the role of genome-wide chromatin interaction progresses rapidly. The Encyclopedia of DNA Element (ENCODE) delineates all of the

functional elements encoded in the human genome sequence and develops advanced technologies for annotating the human genome.

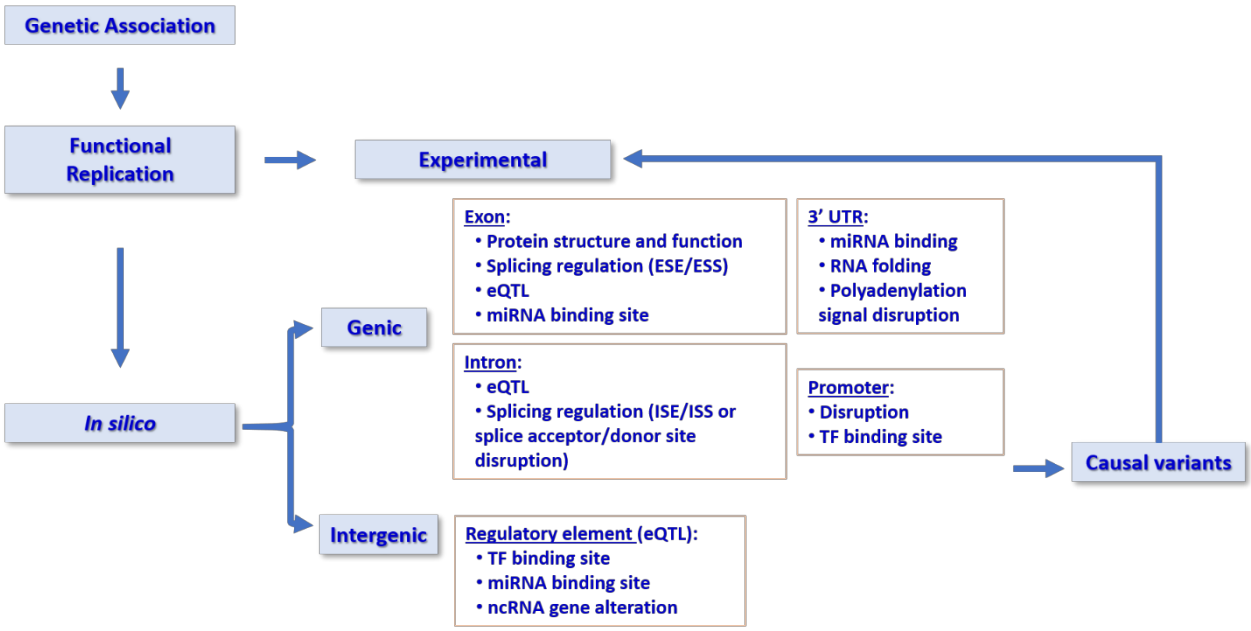

Figure 4. Functional annotation of the detected genetic associations for the diseases/trait. *Depending on the location of genetic associations, the genetic variants affect gene activities through different mechanisms. Abbreviations: eQTL: Expression quantitative loci, TF: transcription factor, miRNA: microRNA, ncRNA: non-coding RNA, UTR: untranslated region, ESE: exonic splicing enhancer, ESS: exonic splicing silencer, ISE: intronic splicing enhancer, ISS: intronic splicing silencer.*

Additionally, expression quantitative trait loci (eQTLs) are the variants that significantly affect gene expression (Pasaniuc et al. 2016). Gene expression can be differently regulated across tissues and the regulation of genes can be nearby (*cis* eQTLs, within 1 Mb) or distant (*trans* eQTLs) (Emilsson et al. 2008). GWAS association signals were found to be enriched with cis eQTLs (Edwards et al. 2013). To figure out whether the identified GWAS loci can affect the expression of their target genes, numerous public resources facilitate the exploration of eQTLs in different tissues, such as Genotype-Tissue Expression (GTEx). The project GTEx (https://gtexportal.org/home/) included genotypes, gene expression, and histological and clinical data from 714 human donors across 53 distinct tissues (Consortium 2013)

At the association locus, the lead SNP may not be the causal variant and haplotypes might also influence eQTL effects (Garnier et al. 2013). Annotating the co-inherited variants in strong LD with the most significant disease-associated variant is helpful to investigate SNPs that cause the association and genes affected by the causal SNP (Gallagher et al. 2018). The maps of regulatory annotations and connections in disease-relevant tissues generated by projects such as ENCODE (Consortium 2012), Epigenome RoadMap (Prakash et al. 2014), and GTEx have been crucial to interpreting variants that account for the majority of GWAS-identified risk alleles.

1.4.2 Pathway-based analysis using GWAS summary data

Despite the success of GWAS in detecting hundred thousands of SNP markers, it fails to properly reveal the relationship between the genetic findings and biological processes (Manolio 2013). The interpretation of polygenic characteristics of complex diseases from the biological point of view is a challenge. Besides, genes that may be genuinely associated with the trait may be excluded by the stringent genome-wide significant threshold in most GWAS (Wang et al. 2010). Moreover, genes do not work independently. The complex biological processes and networks are important in disease susceptibility and progression. Considering these limitations, pathway-based association approaches have been developed to bridge the gap between disease-related loci and the biological mechanisms behind those loci (**Figure 5**) (Wang et al. 2010).

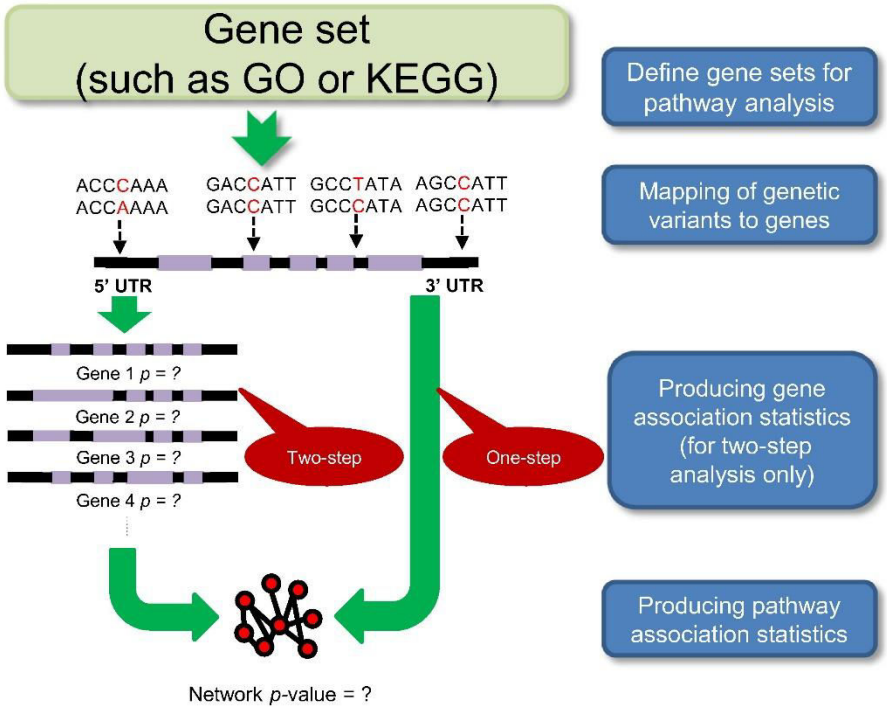

Figure 5.A general workflow for pathway analysis. *A set of pre-defined gene-set should be selected from diverse databases for the next pathway study. Then, SNPs are mapped to genes in the two-step approach or calculating pathway-based statistic directly using one-step approach. Finally, pathway analysis is performed. Picture from (Kao et al. 2017)*

1.4.2.1 Pathway databases

To perform a pathway analysis, a set of gene sets or pathways need to be predefined. Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al. 2000), Biocarta and Gene Ontology (GO) (Ashburner et al. 2000) are commonly used as the predefined gene sets. KEGG pathway and Biocarta are manually collected to represent the molecular interaction, reaction and relation networks (Kanehisa et al. 2000). A large number of electronic annotations were collected in GO for human genes where the gene sets are categorized into the biological process, cellular component and molecular function in a hierarchical structure (Ashburner et al. 2000). Additionally, databases providing gene expression data or protein-protein interaction data are widely used together with KEGG, Biocarta or GO, such as Molecular Signatures Database (MSigDB)(Subramanian et al. 2007), REACTOME (Vastrik et al. 2007), Mouse Genetics

Initiative (MGI) (Blake et al. 2014). MSigDB contains several types of gene set including genes coexpressed in response to genetic or chemical perturbations, genes in expression neighborhoods of cancer-related genes, and genes sharing conserved upstream regulatory motifs. REACTOME is another manually collected and peer-reviewed database. The Mouse Genome Database is a core component of MGI and contains comprehensive phenotype annotations including associations of mouse models with human diseases (Blake et al. 2014; Pers et al. 2015). With the increased identified pathways from different databases, more and more approaches are combined to make sure the results come from comprehensive coverage of the known pathways and hence to increase the precision of the findings (Wang et al. 2010).

1.4.2.2 Input data for pathway analysis

GWAS-based pathway analysis methods can be roughly classified into SNP P-value based and raw genotype based methods according to different types of inputs (Wang et al. 2010) (**Figure 6**).

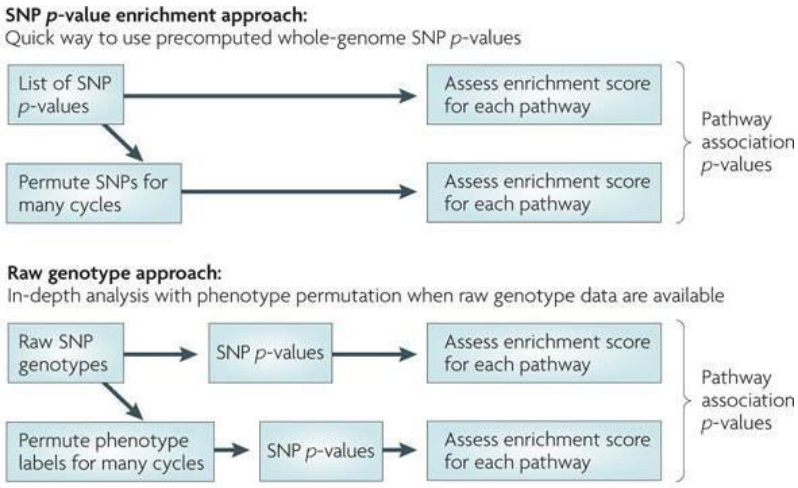

Figure 6. Input data for pathway analysis. *Pathways analysis approaches for GWAS data can be divided into two types based on the input data: SNP P-value enrichment approach and raw genotype approach. Picture adapted from (Wang et al. 2010)*

SNP P-value based study needs a set of P-values for SNPs but does not need individual-level genotypes. Many approaches require a P-value cut-off to define a set of significant SNPs in

GWAS for the following pathway analysis. Usually, a list of independent SNPs is needed. For example, in DEPICT, the independent 'lead SNPs'is defined by only keeping the most significant SNP in a SNP block that are in LD (pairwise r>0.1) and/or in a given physical distance (Pers et al. 2015). This strategy can help to narrow down the analyzing scope without missing important signals. Yet, the result of pathway analysis mostly depends on a user-defined cut-off setting. Another limitation of SNP P-value based approach is that bias can occur because of the gene size. The raw-genotype based approaches use SNP genotypes to perform pathway analysis, which usually requires phenotype permutations to adjust for statistical significance of pathway enrichment scores of the identified pathways. Different from the P-value based approaches, the raw-genotyped based approaches use phenotype permutation that can keep the LD patterns among nearby SNPs. These methods need large computational resources and the raw genotype data is not always available for every study (Wang et al. 2010). SNP ratio test (SRT) (Odushlaine2009) compares the ratio of the number of significant to non-significant SNPs in the original GWAS and the same ratio of simulated GWAS in a specific gene set. The simulated GWASs are created based on the original GWAS, which retain the original genotypes of the GWAS but generate random case/control status. To get a statistically significant P-value, more than 1000 times of simulation is necessary.

1.4.2.3 Gene mapping

The pathway approaches can also be classified into "one-step" and "two-step". For one-step approaches, SNPs are directly used to calculate the pathway-based statistics. For two-step approaches, pathway analysis is performed based on a gene-based score from the P-values of variants. Meanwhile, for two-step analysis, whether the approach is SNP P-value based or raw genotype based, mapping SNPs to genes is an inevitable step. Generally, two strategies are used to map SNP to genes, 1) physical position mapping -- setting gene boundaries, such as taking genes from 500kb, 200kb, 5kb in both upstream and downstream of the target SNP. 2) LD block mapping -- taking SNPs that are in LD with the genes (Wang et al. 2011). It is risky to adopt such strategies for gene mapping as it may include as many disease-associated genes as irrelevant genes and increase the computational burden. These strategies are still widely used in the pathway analysis, notwithstanding it may lead to bias. However, filtering SNPs by functional

annotations (eg. CADD and RegulomeDB) before the physical position or LD block mapping may be useful for mapping more disease-related genes (Watanabe et al. 2017). In addition, eQTL mapping and chromatin interaction mapping are available in FUMA. If the independent GWAS significant SNPs and SNPs in high LD of them are mapped to eQTLs in specific tissue types, those SNPs will be selected to map to genes up to 1 Mb apart, which refers to cis-eQTLs. For chromatin interaction mapping, interacting regions "A" in a specific tissue/cell types were overlapped with the independent significant GWAS SNPs and SNPs in LD between them. Then, these SNPs are mapped to genes whose promoter regions overlap with interactions region “B”, A and B are interacted. Chromatin interaction mapping is based on chromatin interactions such as Hi-C and ChIA-PET.

1.4.2.4 Calculations methods of statistics in pathway enrichment analyses

In the aspect of hypothesis tests, pathway analysis approaches can be classified by self-contained test and competitive test (**Figure 7**) (Goeman et al. 2007). The self-contained methods compare a defined pathway with a simulated dataset that has no significant phenotype association. Competitive enrichment methods compare the collective association within a pathway to the collective signal among genes not in the pathway (Goeman et al. 2007; Ramanan et al. 2012). The self-contained test can be used in candidate genes association study, which is impossible in the competitive test. Because for competitive approaches, genes that are not in the interested pathway are required, but self-contained approaches do not need such data. This requirement has limited the utilization of competitive approach in GWAS (Kao et al. 2017). However, the genomic inflation self-contained test facing the un-monitored genomic inflation which may cause type I error, in addition to some degrees of genomic inflation in GWAS. These may be the reasons why competitive test is widely used in the pathway analysis for GWAS data (Wang et al. 2010). For example, iGSEA4GWAS, a competitive approach, use an improved gene set enrichment analysis (GSEA) to test the enrichment of the associated genes within a pathway set (Zhang et al. 2010).

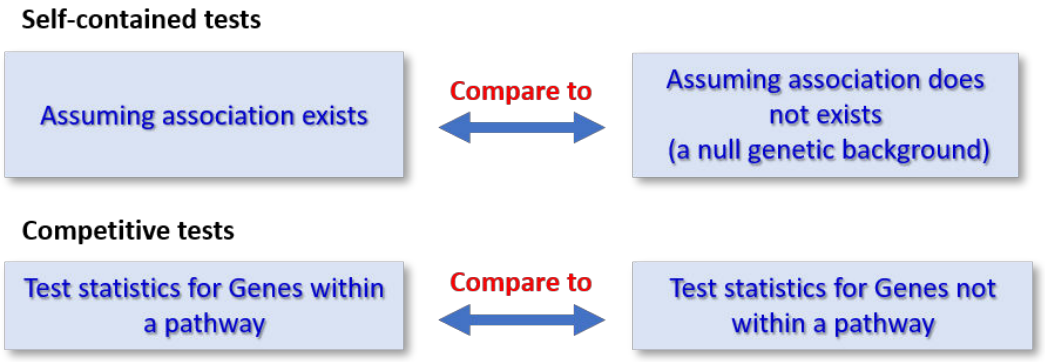

Figure 7. Competitive and self-contained approaches. *Self-contained tests calculate the pathway P-values by comparing the statistics assuming that the observed pathway is associated with the phenotype versus that there is no association (null hypothesis). For competitive approaches, statistics of genes within a pathway is compared with gene sets that contain genes not within the pathway.*

1.4.2.5 Multi-omics data based pathway analysis

Nowadays, the simple use of the database for pathway analysis cannot meet the needs to understand the complex biological system of human. Integrating data from multiple "-omics" platforms can provide a comprehensive understanding of how the risk variants affect the disease from a biology point of view (Kao et al. 2017). Tools are developed to integrate and analyze huge databases, such as iGWAS (Huang et al. 2015), SPATIAL (Bokanizad et al. 2016) and DEPICT. DEPICT uses gene expression data from 77,840 microarrays including two human, one rat and one mouse Affymetrix gene expression platforms in conjunction with pre-defined gene sets from diverse databases and data types (including protein-protein interactions database, Mouse Genetics Initiative database, Reactome, KEGG and GO terms), to build a large 'reconstituted' gene sets. This 'reconstituted' gene sets (contain a gene-gene set matrix of Z-scores comprising 19,997 gene rows and 14,461 gene sets columns.) were used for the gene-set enrichment analysis and tissues/cell types enrichment analysis.

1.4.3 Network analysis for GWAS data

Using pathway analysis, numbers of disease-related pathways are highlighted, which makes the interpretation difficult. Besides, these identified pathways are enriched based on existing database and literature which ignores the random selection of true biological relationships (Greene et al. 2012). Several significant genes with small effects may cause multiple distinct pathways which is difficult to explain. In addition, results from different methods may differ greatly. More importantly, identified pathways or genes from pathway analysis are independent and the connection among them is ignored. Hence, a network-based analysis to identify the interaction and to detect the biological modules underlying cellular architecture and functions becomes important.

Jia et al. (Jia et al. 2014) summarized several features that distinguish pathway-based and network-based analysis. Compared to pathway-based analysis, network-based analysis defines sub-networks and assesses the combined effects of multiple genes participated in the sub-networks. Generally, networks consist of nodes and edges. Nodes can be genes, proteins, pathways, or diseases. Edges represent a functional or physical connection between the nodes. The layout of nodes and edges in a network is network topology (Hu et al. 2016). Edges either come from existing database called “curated networks” or inferred based on available data called “inferred networks”. The curated networks provide a higher fidelity for the edges compared to inferred networks. But inferred networks have the possibility to identify the underlying relationships (Himmelstein et al. 2015).

Network analysis relies on network annotations that present functional interactions among genes and their products. Based on the hypothesis that genes or proteins grouped in a sub-network are more likely to share similar functions. Using guilt-by-association to search for sub-networks that are both enriched with association signals in GWAS data and contain functional relationships is helpful to detect disease susceptibility genes and their interactions (Jia et al. 2014).

In general, four core steps are included in the network-based analysis of GWAS data (**Figure 8**):

Step 1, prepare the network based on the known expression data. There are a variety of molecular networks that have been used to explore the biological mechanisms, such as protein-

protein network, regulatory network, gene co-expression network, metabolic network (Hu et al. 2016)). Interaction network can generally be classified into two groups if the interactions are direct and physical or indirect but functional. Physical interactions include protein-protein interaction (PPI), which come from large-scale high-throughput experiments in vitro (Barshir et al. 2013; Jia et al. 2014). Search Tool for the Retrieval of Interacting Genes (STRING) provides

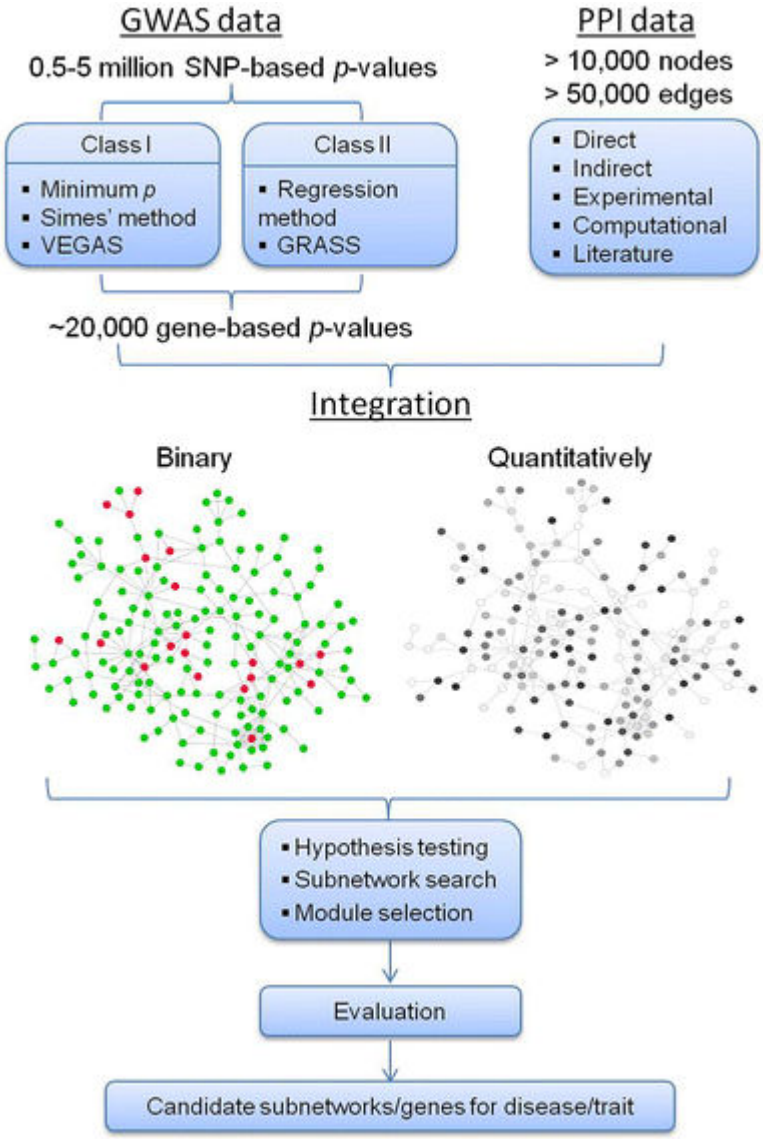

Figure 8. Overview of the network analysis. *GWAS data is used to generate gene-based P-values firstly. Then, protein-protein interaction data or other existing expression data is pre-defined. An integration analysis is performed to discover the interaction network and analyze the subnetwork/genes for the diseases/trait. Adapted from (Jia et al. 2014)*

access to both experimental as well as predicted PPI information (Szklarczyk et al. 2011). Functional interactions include tissue-specific gene co-expression, literature co-citation, and genetic interactions (Hu et al. 2011; Jia et al. 2014). Compared to physical interactions, functional interactions are more biologically associated with diseases/traits, which helped to discover relevant networks that are enriched with GWAS associations (Jia et al. 2014).

Step 2, similar to pathway-based analysis, mapping the association signals to genes and calculate the gene-based statistical value is needed. This is a critical step for the following analyses. See Section Pathway analysis – gene mapping section for details.

Step 3, perform an integrative analysis on the GWAS data and pre-defined interaction data to explore the interactions in the network, discover the relationship among genes or proteins of interest, and pick out the subnetworks. Diverse approaches were developed to detect biological interactions, such as Genome-wide Integrated Analysis of Gene Networks in Tissues (GIANT) (Greene et al. 2015), protein interaction network-based pathway analysis (PINBPA) (Wang et al. 2015) and DEPICT (Pers et al. 2015). GIANT provides a visualized interface to human tissue networks through multi-gene queries, analysis tools including NetWAS and downloadable networks. PINBPA is a Cytoscape app that can prioritize genes and detect enriched subnetworks. DEPICT is used for pathway analysis but also provides pathway related networks.

Step 4, evaluate the networks and perform follow-up functional analyses of genes in the sub-networks. Open source network exploration tools, such as Cytoscape (Shannon et al. 2003) and Gephi (Bastian et al. 2009) were developed to visualize the networks.

Although network approaches developed to date have brought promising results and help to explain GWAS data, a number of challenges remain in the network analysis of GWAS. To name a few, missing interactions still exist which needs further discovery (Leiserson et al. 2013). Incorporating different interaction databases from different platforms and diverse ethnic populations is a challenge as well (Jia et al. 2014). In conclusion, network analysis enables us to investigate the interaction among the associated genes and the biological mechanisms behind the GWAS.

2 CHAPTER II – The studies of heart valve diseases

With the development of sequencing and genotyping technologies and the growing annotation databases and sophisticated analyzing tools, the genetic study of complex human diseases is more comprehensive, interpretable and reliable. In this thesis, we study the genetics of mitral valve prolapse (MVP), which is transmitted in an autosomal dominant mode (Weiss et al. 1975). MVP could be acquired from rheumatic valve disease, endocarditis, and trauma. MVP was found in the clinical syndromes, such as Marfan syndrome. Nearly 70% of patients with Marfan syndrome were found to have mitral valve dysfunction (generally MVP) (Pyeritz et al. 1983). However, in most instances, MVP is a non-syndromic and isolated disease that can be familial or sporadic in the general population.

To understand the MVP better, the genetics of aortic valve disease (AVD), which is biologically close to MVP, is summarized first. Meanwhile, the aortic valve shares the same structure with the mitral valve, and AVD and mitral valve disease (MVD) are the most common valvular diseases. The understanding of the genetic mutations and mechanisms of AVD at the fundamental level may help to inspire our MVP genetic study and is essential to improving preventive and therapeutic strategies. Then, the structure of the mitral valve and the diagnosis are briefly recalled. A general overview of how mitral valves development, growth, and aging is presented. At last, the genetics of MVP from family studies and GWAS in a large cohort are also reviewed.

2.1 The structure of human heart valves

In the human heart, four valves, aortic valve (AV), mitral valve (MV), pulmonary valve (PV) and tricuspid valve (TV), are involved in the blood circulation (**Figure 9**). Problems arise when one or more of the valves are not opening or closing properly, which cause heart valve disease. Regurgitation and stenosis can happen when the heart valve is diseased or damaged. Regurgitation occurs when the valve does not close tightly so that the blood flows backward rather than being pumped out of the heart, due to valve prolapse or valve thicken and shorten.

For stenosis, valves turn thicker or become very stiff, resulting in the incomplete open of the valve and limited or blocked blood flow can be found. Although the screening of the general population indicates that many individuals with valvular diseases have few clinical symptoms, the valvular heart disease is associated with significant morbidity and mortality (Padang et al. 2012).

According to the position of the diseased valve, valvular diseases can be classified into the right-sided valvular heart diseases (RVHDs) and the left-sided valvular diseases (LVHDs). In the RVHDs, the prevalence of tricuspid regurgitation (TR) has not been well studied (Al-Mohaissen et al. 2012). Pulmonary valve diseases are mostly congenital diseases (Iung et al. 2011). In the Euro Heart Survey, isolated right-sided native valve diseases (severe pulmonary and tricuspid diseases) only accounted for 0.8% of all patients (Iung et al. 2007).

The LVHDs concern the aortic and the mitral valve. In Europe, aortic valve stenosis (AVS or AS) (43.1%) and mitral regurgitation (MR) (33.6%) are the most common types of valvular diseases and are mostly caused by degenerative diseases (Iung et al. 2007). AVS occurs when the heart's aortic valve narrows. AS prevents the valve from fully opening and reduces or blocks blood flow from the heart into the main artery to the rest of the body. Mitral valve prolapse (MVP) is the leading cause of mitral regurgitation (MR) (Hayek et al. 2005).

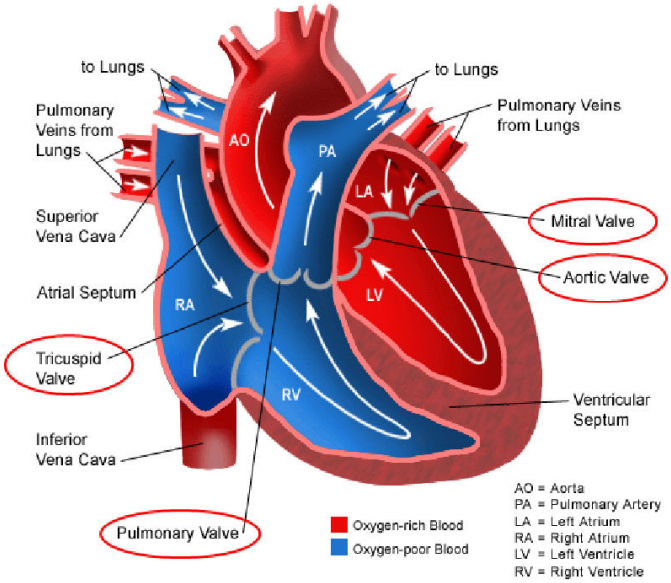

Figure 9. Drawing of a normal heart illustrating chambers and valves locations. *The heart has four chambers separated into left and right sides by the ventricular septum. The right and left sides of the heart are further divided into two top chambers called the atria which receive blood from the veins, and two bottom chambers called ventricles which pump blood into the arteries. The tricuspid and pulmonary valves belong to the right-sided valves, the mitral and aortic valve are left-sided valves.*

In this part, current genetic studies of the LVHDs, namely AS and MVP are summarized. In particular, sporadic MVP in the general population will be further highlighted in this thesis manuscript.

2.2 Aortic valve diseases

Normal aortic valves contain three tightly fitting, triangular-shaped flaps cusps, but some people are born with only two cups. This congenital valve disease is called bicuspid aortic valve (BAV). As aging, accumulated deposits of calcium in the aortic valve may result in calcific aortic valve diseases (CAVD), such as calcific aortic valve stenosis (CAVS). CAVD is the most common cause of AS among adults which include bicuspid and tricuspid valve (Roberts et al. 2005).

As the most common type of valvular heart diseases and the most frequent LVHDs, the prevalence of AS increases with age. Eveborn (Eveborn et al. 2013) reported that the prevalence

is only about 0.2% among adults aged 50–59 years, 1.3% in the 60–69 years, 3.9% in the 70–79 years, but increases to almost 10% in the adults over 80 years or older. Therefore, AS turns to be an important societal and economic burden.

2.2.1 Genetics of bicuspid aortic valve

BAV is a common congenital cardiac disease, affecting 0.5–2.0% of the general population (Prakash et al. 2014). BAV is entirely genetically determined and a highly heritable disease, which follows an autosomal dominant mode of transmission (Cripe et al. 2004; Laforest et al. 2012). Several molecular signaling pathways have been demonstrated to be involved in aortic valve formation, for example, migration of neural crest cells (Jain et al. 2011), extracellular matrix (ECM) remodeling (Fedak et al. 2003), pathways related to the formation of the outflow tract (OFT) and the endocardial-mesenchymal transition (Laforest et al. 2011; Laforest et al. 2012). Hence, BAV is a polygenic disease which influences the development of aortic valve by the combinational effects of multiple genes.

A number of studies have identified a large number of genes suggested having an impact on the pathogenesis of BAV. The associated genetic loci are summarized in **Table 1**. Mutations in *NOTCH1* were found in families with BAV and the hypoplastic left ventricle suggested that *NOTCH1* mutations may be the potential genetic basis for left heart syndrome (Garg et al. 2005).

Except for *NOTCH1*, genes in/related to the GATA family play a role in BAV as well. The deletion of *Gata5* in mice affected heart development and led to BVA by regulating several pathways involved in endocardial cell differentiation, including Bmp2, Tbx20, Nos3 and Notch pathways (Laforest et al. 2011; Padang et al. 2012). Besides, *GATA4* and *GATA6* were observed in human congenital heart defects. The missense mutation in *GATA4* resulted in the impaired epithelial-mesenchymal transition, which is a crucial step in valve formation (Michelena et al. 2011). *Gata6* heterozygote mice developed right and left leaflet fusion type of BAV. By downregulating *MMP9* and *BMP4*, *Gata6* disrupts valve remodeling and alters extracellular matrix composition (Gharibeh et al. 2018).

Table 1. Genes at associated loci in BAV.

Chr	Locus/ Genes	Phenotype	Study Population	Analytical methods	Ref
9	*NOTCH1*	BAV=9	European–American	Linkage analysis	(Garg et al. 2005; Kerstjens-Frederikse et al. 2016) (Mohamed et al. 2006; McKellar et al. 2007)
	18q22.1 5q21.2 13q34.3	BAV=74	European–American*	Linkage analyses	(Martin et al. 2007)
20	*GATA5*	BAV=100	Australia	Targeted sequencing	(Laforest et al. 2011; Padang et al. 2012; Shi et al. 2014)
15	*SMAD6*	BAV=2	British	Targeted sequencing	(Ankeny et al. 2011) (Tan et al. 2012)
10	*ACTA2*	BAV=4 in 13 families	American	Linkage analyses	(Guo et al. 2007)
5	*NKX2*.5	BAV=142	Chinese Han	Targeted sequencing	(Qu et al. 2014)
7	*EGFR*	BAV=48	French Canadians	Targeted sequencing	(Dargis et al. 2016)
16	*AXIN1-PDIA2* locus	BAV=68	European–American*	Gene Network Analysis	(Wooten et al. 2010)
16	*ENG*	BAV=68	European–American*	Gene Network Analysis	(Wooten et al. 2010)
8	*GATA4*	446 BAV cases 4660 controls	European	GWAS & Exome chip	(Yang et al. 2017)
18	*GATA6*	452 BAV cases 1849 controls	European	GWAS qRT-PCR immunohistochemistry	(Gharibeh et al. 2018)

Several other genes have been implicated in BAV. One of those is *SMAD6* expressed in cardiac valves and outflow tract and may reduce the efficiency of inhibiting the osteogenic potential of BMP signaling (Ankeny et al. 2011; Tan et al. 2012). The Endoglin gene (*ENG*) was found to be associated with BAV, which is expressed in heart valves and the aorta (Valeria et al. 2008) and is related to the endocardial cushion mesenchyme during valve formation (Qu et al. 1998) (Wooten et al. 2010). *AXIN1* is part of the Wnt pathway which is vital in heart valve formation (Armstrong et al. 2004).

2.2.2 Genetics of calcific aortic valve stenosis

CAVS is a common cardiac disease affecting 2% of the population ≥65 years old (Stewart et al. 1997; Probst et al. 2006). The prevalence of CAVS increases exponentially with age and was regarded as a degenerative and irreversible process (Bosse et al. 2008). CAVS can hide for years until the occurrence of symptoms, such as cardiac failure (Horstkotte et al. 1988; Probst et al. 2006). CAVS can be caused by several clinical risk factors, like smoking, male gender, age, diabetes, and hypertension (Stewart et al. 1997). Another major factor is the congenital BVA, which results in the hemodynamic alterations that may accelerate the calcifying process (Bosse et al. 2008). However, CAVS with the normal tricuspid aortic valve is suspected to have unique genetic mechanisms. (**Table 2**).

Mutations in vitamin D receptor (*VDR*) and apolipoprotein E (*APOE*) were found to increase the risk of CAVS suggesting a genetic background of the disease (Avakian et al. 2001; Ortlepp et al. 2001; Novaro et al. 2003). *LPA* was found as a potential therapeutic target for the prevention of AS (Cairns et al. 2017) (Kamstrup et al. 2014). In addition, *RUNX2* and genes from the calcium signaling pathway (*CACNA1C*) are the potential risks for CAVS (Guauque-Olarte et al. 2015). CACNA1C, a part of the protein complex that forms a voltage-dependent calcium channel was upregulated in calcified valves. *RUNX2* is a critical transcription factor for osteoblastic differentiation and skeletal morphogenesis (Guauque-Olarte et al. 2015). Although *NOTCH1* was identified as the causal gene for BAV, it is associated with CAVS as well and affects calcium deposition by differentiation of valve cells to the osteoblast-like cell (Rajamannan et al. 2003). In

addition, *Runx2* and the osteopontin expression was found to be induced by stimulation of human aortic valve interstitial cells with bone morphogenetic protein 2 (Yang et al. 2009).

Table 2. Genes at associated loci in CAVS.

Chr	Locus/ Genes	Phenotype	Study Population	Analytical methods	Ref
12	vitamin D receptor	100 CAVS cases 100 controls	European*	case-control study	(Ortlepp et al. 2001)
19	*APOE*	62 severe AS 62 control	Brazil*	case-control study	(Avakian et al. 2001)
19	*APOE*	802 patients (43 patients with AS)	European–American*	case-control study	(Novaro et al. 2003)
6	*LPA*	In total 6942	White European	GWAS	(Thanassoulis et al. 2013; Kamstrup et al. 2014; Cairns et al. 2017)
6 12	*RUNX2* *CACNA1C*	960 cases 4852 controls	Canada & European	Meta-GWAS	(Guauque-Olarte et al. 2015)
1	*PALMD*	1009 tricuspid CAVS cases and 1017 controls	Quebec	TWAS	(Helgadottir et al. 2018; Theriault et al. 2018)
2	*TEX41*	2,457 AS cases and 349,342 controls	Iceland	GWAS	(Helgadottir et al. 2018)

We do not know much about the mechanisms of the calcification in the tricuspid aortic valve, even though the variants were found in *LPA*, *RUNX2*, *CACNA1C*, *PALMD*. The processing of osteoblastic differentiation associated with *RUNX* and *CACNA1C* is related to the calcification of BAV. This suggests that the calcification of bicuspid and tricuspid aortic valve may share similar mechanisms.

In most of the family studies of BAV, the bicuspid valve is not the only phenotype in the family pedigree. Mitral valve abnormalities are frequently observed in the family individuals, such as stenosis, prolapse, thickening, myxomatous changes or regurgitation (Garg2005, Kerstjens-

Frederikse2016, Martin2007). BAV and CAVS related genes and their corresponding functions are important to MVP as well, such as *GATA4* and *GATA6*. Meanwhile, biological functions linked to those genes, such as endothelial to mesenchymal transition (EMT) of endocardial cells, ECM composition, are important to both aortic and mitral valve (Puceat2013). Genetic mutations in genes of the signaling pathways related to EMT, such as NOTCH, BMP and TGFβ signaling pathways may affect both the aortic and mitral valvulogenesis through different mechanisms.

2.3 Mitral valve Diseases

Mitral valve prolapse (MVP) is defined as an abnormal mitral leaflet displacement into the left atrium during systole (Guy et al. 2012) (**Figure 10**) It is firstly recognized by Barlow where the mitral origin of late-systolic murmurs often associated with clicks by using left ventricular cineangiography in the 1960s (Barlow et al. 1966). Criley subsequently termed this condition mitral valve prolapse (Criley et al. 1966). Although MVP usually begins with no serious symptoms and most patients with MVP may not need surgical intervention, it can lead to serious complications, such as progressive mitral regurgitation, heart failure, arrhythmia, and sudden cardiac death (Sriram et al. 2013; Basso et al. 2015; Nalliah et al. 2019). Prospective studies have shown that asymptomatic patients with low-risk presentation (e.g., moderate MR and ejection fraction ≥50%) can develop adverse MVP-related events, indicating wide heterogeneity in outcomes among individuals (Avierinos et al. 2002). MVP is the most common and an increasing indication for surgical repair of MR (Avierinos et al. 2002; Gammie et al. 2018). Based on the histopathological evidence, degenerative MVD can be classified into Barlow's Disease and Fibroelastic Deficiency (Anyanwu et al. 2007; Hjortnaes et al. 2017).

The prevalence of MVP is similar in South Asian (2.7%), European (3.1%) and Chinese (2.2%), but a familial clustering of MVP in the community found that parental MVP has a higher prevalence compared to non parental MVP (5.4% VS 1.1%) (Theal et al. 2004; Delling et al. 2014). Recently, a meta-analysis based on a literature survey indicated that the prevalence of MVP in the community population is 1.2% (Nalliah et al. 2019). A high level of heterogeneity (I^2=97.5, X^2 $p<0.001$) was shown in this study, which suggested that different populations have

variable genetic penetration (Nalliah et al. 2019). However, the prevalence of MVP was widely considered at 2.4% (Freed et al. 1999).

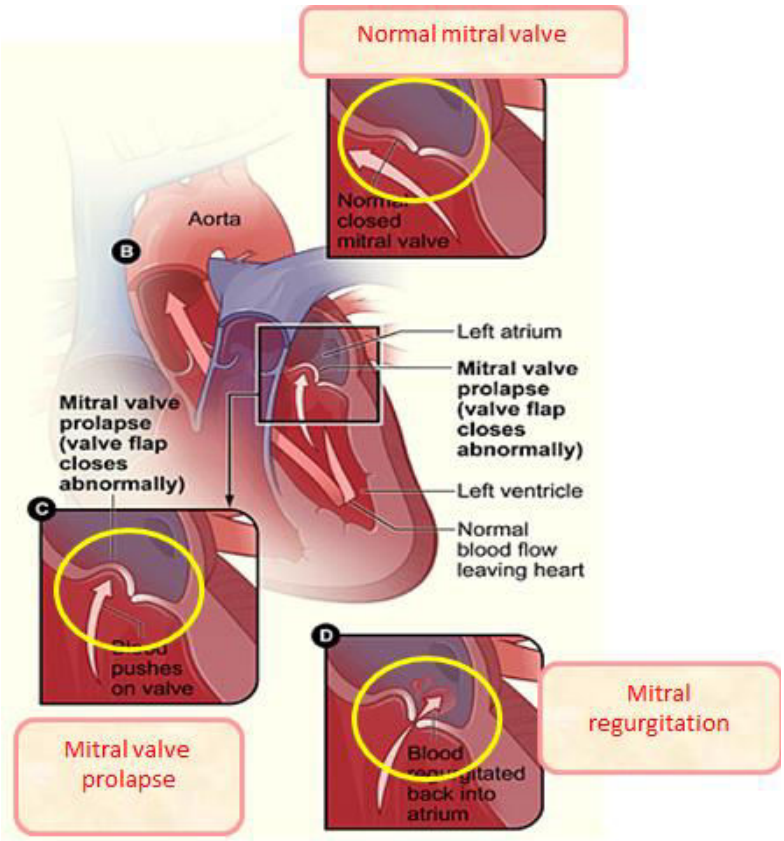

Figure 10. Mitral valve prolapse and Mitral Regurgitation. *MVP is an abnormal protrusion or billowing of one or both of the leaflets of the mitral valve into the left atrium during systole. MVP is a risk for mitral regurgitation, which leads to leakage of blood back into the left atrium through the mitral valve during systole.*

Mitral valve disease (MVD) was considered as an unavoidable process until the evidence indicated that the mitral valve is a dynamic and accessible structure for treatment (Judge et al. 2011; Levine et al. 2015). Nowadays, the treatment of MVP is limited to surgical valve repair or replacement, and the underlying mechanisms are not assessed (Levine et al. 2015). By integrating diverse aspects, such as valve imaging, biology, and genetics, to understand the mechanisms of mitral valve function, novel therapies and simplification of the clinical progression of MVD can be possible.

In this part, functions of mitral valves, the basic anatomy of the mitral valve, potential biological processes and genes related to the development, growth, and aging of the mitral valve are described. In addition, we summarize current genetics studies on syndromic and non-syndromic MVP.

2.3.1 Diagnosis approaches

Echocardiography (Echo) is an ideal clinical tool to diagnose and assess abnormal leaflet and measure the structural and functional consequence of MVDs (Dal-Bianco et al. 2013). Two-dimensional (2D) and 3-dimensional (3D) Echo provide a detailed morphological and functional assessment and are widely used tools in the diagnosis and valve repair surgery (Dal-Bianco et al. 2013). However, 2D Echo with old echocardiographic criteria makes the diagnosis less accurate, especially when the detected prolapsed valve comes from normal saddle shape without actual pathological leaflet displacement (Levine et al. 2015). The incidence of MVP was as high as 10-15% using 2D Echo (Procacci et al. 1976). 3D Echo improves the detecting accuracy of the valve pathology. It is also helpful to predict ease of repair, or suggest possible repair techniques (Guy et al. 2012). Using the criteria in 2D Echo (mitral leaflet displacement exceeds more than 2 mm above the normal range of mitral annulus during systole and mitral leaflet thickness more than 5 mm during diastole), 3D Echo has revised the prevalence of MVP to 2.4% without loss of sensitivity (Freed et al. 1999). The improved diagnostic specificity laid solid foundations for genetic studies of MVP. Alternative imaging techniques for visualizing the mitral valve are cardiac computed tomography (CT) and magnetic resonance imaging (MRI). CT performs well in detecting MVP but the required radiation dose limits its utilization (Ghosh et al. 2012; Smith et al. 2012). MRI has a very high precision in measuring mitral regurgitation but it is not widely used for mitral valve disease diagnosis yet (Chan et al. 2008; Chinitz et al. 2013).

2.3.2 Mitral valve structure

To improve the understanding of MVP, the reorganization of the mitral structure is essential. Mitral valve is one of the heart valve which has two leaflets, the anterior leaflet (contiguous with

the aorto-mitral curtain) and the posterior leaflet (composed of three scallops, P1–P3) (**Figure 1-11**). Both leaflets attach to the dynamic mitral annulus at their basal ends. Emerging from the papillary muscles (PM) heads or from the basal posterior myocardium, the multiple chordae tendineae insert into the ventricular aspects of the anterior, posterior and commissural leaflets (Dal-Bianco et al. 2013). The elastin and collagen network in chordae restrain the PM-valve force transmission (Millington-Sanders et al. 1998). This complex shape ensures the one-way flow of blood in the heart through the mitral valve.

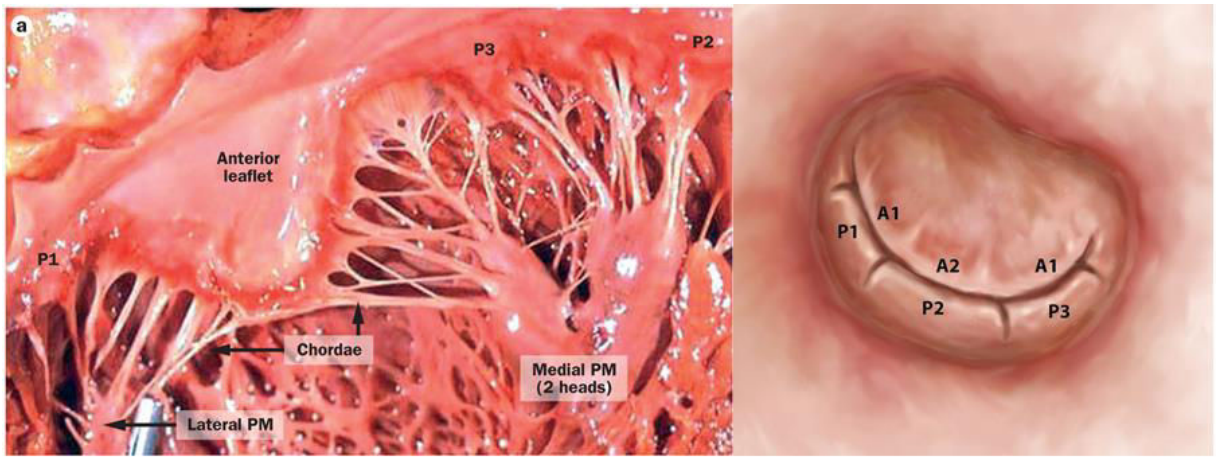

Figure 11. Anatomical structure and leaflet composition of the mitral valve. *Mitral valve has two leaflets, the anterior leaflet below the aortic valve and the posterior leaflet composed of three scallops, P1, P2, and P3. On the leaflet edges, the chordae tendineae link the leaflets to the PM anchors on the left ventricular wall. Picture adapted from (Guy et al. 2012; Levine et al. 2015).*

The region connecting the mitral annulus and leaflet edge contains atrial myocytes, nerve fibers, collagen, small vessel, and smooth muscle cells. The nerve fibers innervate the cardiac muscle of the atrioventricular valves and maintain the electrophysiological continuity with the rest of the heart (Cooper et al. 1966; Sonnenblick et al. 1967). Continuous endothelial cells cover the valve on both atrial and ventricular aspects. Mitral valve leaflets have three well-defined tissue layers, atrialis, spongiosa and fibrosa/ventricularis (Wit et al. 1979; Levine et al. 2015) (**Figure 1-12**). Unique extracellular matrix (ECM) characteristics of each layer ensure the proper function of the mitral valve.

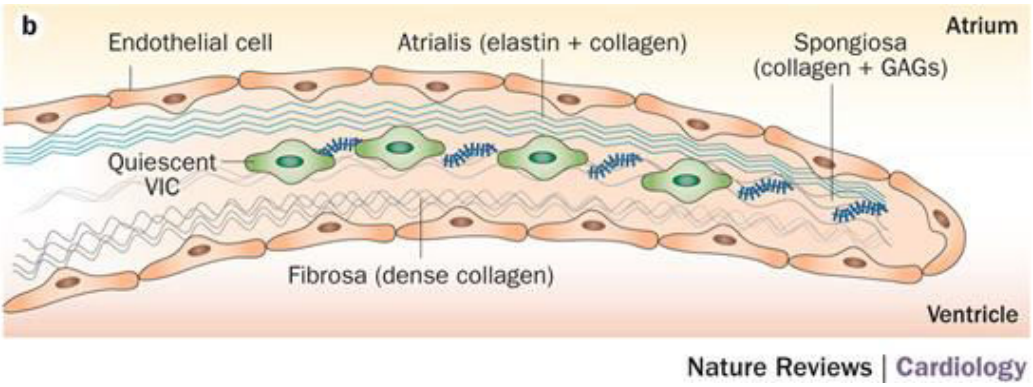

Figure 12. Normal mitral valve structure. *The leaflet structure is well organized by three layers, known as the atrialis, spongiosa, and fibrosa. Endothelial cells cover the valve. GAG, glycosaminoglycan; VIC, valvular interstitial cell. Picture from (Levine et al. 2015)*

Beneath the endothelial cells, the atrialis layer extends from the left atrial endocardium into the leaflet. The atrialis layer contains cardiac muscle cells, well-defined cell bundles, and several myofibrils thick, and arranged lamellar collagen and elastin sheets (Levine et al. 2015). The collagen and elastin are critical to leaflet adaptation, remodeling and enabling elastic recoil in diastole (Dal-Bianco et al. 2009). The myocytes, smooth muscle cells and cell bundles in this layer may be responsible for transiently modulated leaflet stiffness and deformation during the closure. Thus, the atrialis layer can sustain high frequency and tensile deformations in systole (Itoh et al. 2009). The spongiosa, beneath the atrialis, is thinner at the mitral annulus and becomes thicker towards the leaflet free edge (Wit et al. 1979; Levine et al. 2015). This layer contains loosely arranged and less organized collagen, but is enriched in glycosaminoglycans (GAGs). GAGs are the small proteoglycans decorin and biglycan related and enriched in the tensile regions (Grande-Allen et al. 2004). The GAGs are helpful in the retention of water, hyaluronan, and glucuronate in the spongiosa, which can accommodate the tensile stress and compressive loading (Grande-Allen et al. 2004). Along the high pressure left ventricular aspect, the fibrosa consisted of dense collagen is extended from the mitral annulus to the leaflet edge (Wit et al. 1979; Levine et al. 2015). The dense network of collagen fibers in this layer is helpful in resisting high tensile forces and maintaining stability (Sacks et al. 2007).

The deep subendothelial layers contain a large number of valvular interstitial cells (VICs), unique cells of cardiac valves that are originated from embryonic progenitor endothelial/mesenchymal cells (Liu et al. 2007). The "activated" VICs can alter the extracellular environment rapidly by various processes, such as the expression of proteolytic enzymes and smooth muscle associated contractile proteins (Levine et al. 2015; Spadaccio et al. 2016). In healthy MVs, interstitial cells should be mostly dormant, and the ECM turnover should be slow as well (Spadaccio et al. 2016).

A normal mitral valve structure is the basic guarantee for its dynamics and competence. Elongation of the anterior leaflet, loss of annular dynamics and restricted leaflet motion might change the vertical structures, cause adverse ventricular dilatation, increase diastolic filling pressures, and impair the efficient ventricular pumping (Eriksson et al. 2010; Ro et al. 2014). The mechanisms of MV adaptation are not well understood and the determinants of MV diseases needs further exploration.

2.3.3 Mitral valve development, growth, and aging

Mitral valve development at early stage is complex. As the embryo folds ventrally, the two heart fields coalesce into a linear heart tube composed of myocardial and endocardial cells which contribute to mitral valve formation in the embryo (Milgrom-Hoffman et al. 2011; Pucéat 2013). Vascular endothelial growth factor receptor 2 (VEGFR2) is highly expressed in the endocardial progenitor cells but less expressed in myocardial progenitor cells. The myocardial progenitor cells differentiate into the first and second heart lineage which have the potential to become endocardial progenitor cells with the increased expression of VEGFR2 (Pucéat 2013). The endocardial progenitor cells undergo EMT to generate different cell types are involved in the formation of valve leaflets (Levine et al. 2015). Heart endocardium is derived from vascular endothelial progenitors which has been demonstrated in both avian and mouse models (Milgrom-Hoffman et al. 2011). VEGFR2 expression is required for the endocardium only within an early developmental window and is essential for the formation, but not for the maintenance of endocardial cells (Milgrom-Hoffman et al. 2011).

Mitral valve development begins soon after looping the early embryonic heart tube (Levine et al. 2015) (**Figure 13**). In the early embryonic heart, to respond to the growth factors such as TGFβ1-3, BMP2-4, and SMAD, a subset of endothelial cells are activated and transformed into a mesenchymal phenotype, known as EMT (Armstrong et al. 2004; Inai et al. 2008; Levine et al. 2015). The vascular endothelial growth factor A (VEGF) expressed in endothelial cells promotes endothelial cell proliferation. The endothelial cells that do not undergo EMT express nuclear factor of activated T cells cytoplasmic 1 (NFATc1), which suppresses transcriptional factors Snail1 and Snail2. Snial1/2 are important for initiating EMT and needed for cells to proliferate and elongate (Wu et al. 2011).

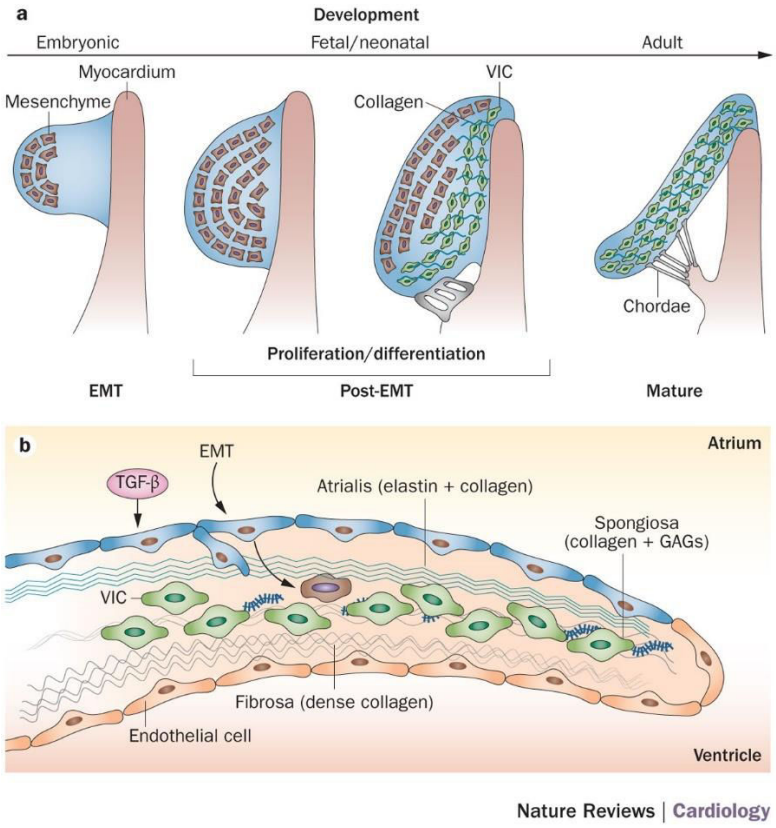

Figure 13. Mitral valve growth and development. *a) Mitral valve development from an endocardial cushion. The cells undergo the EMT process to generate mesenchyme (brown cells). The mesenchyme undergoes proliferative expansion and then differentiate into collagen-secreting VICs (green). Finally, those stratified and aligned cells and matrix form a mature valve b) Valve growth and response to stress. A growing or stressed valve can be stimulated by TGF-β, then endothelial cells undergo EMT to increase the number of matrix-producing interstitial cells. EMT, endothelial-to-mesenchymal transition; GAG, glycosaminoglycan; TGF-β, transforming growth factor β; VIC, valve interstitial cell. Picture came from (Levine et al. 2015).*

The VIC generated by the EMT process where the mesenchymal cells undergo migration and proliferation at the Post-EMT stage (de Vlaming et al. 2012; Pucéat 2013). Then, VICs are proliferated under the action of VEGF and then the valve elongates and thins to form the maturing leaflets. A few genes including Twist, Tbx20, and Sox9 and their downstream targets Sox5 and Sox6 are essential to regulating this proliferative stage (Thiery et al. 2009; Pucéat 2013).

During the aging process, the mitral valve contains less cellular, dense, thick and nonparallel collagen bundles, exhibiting an increased number of elastic fibers which decreases in elastin and increases in stiffness, excessive collagenolytic activity, lipid accumulation and calcifications, and reduces density of mucopolysaccharides (Pomerance 1967; Spadaccio et al. 2016). These phenomena are more serious in the individuals elder than 60 years (Pomerance 1967). Increasing evidence supports that ECM plays an important role in valve aging and remodeling. An excessive accumulation of GAGs, reduced and fragmented elastin fibers, and loss of collagen are demonstrated in aged valves. Besides, the activated state of VICs and valvular endothelial cells (VECs) are affected by the physiological states and tissue environment which is influenced by the ECM changes (Spadaccio et al. 2016). In addition, aging VECs showed lower cell proliferation rates, attenuated nitric oxide production, increased reactive oxygen species (ROS), disorganized mitochondria, reduced cell membrane self-repair in response to stretch-induced injury and decreased SMA expression in response to TGFβ1 treatment. The biological processes related to focal adhesion and cell cycle are negatively impacted by aging as well (Anstine et al. 2016). Whether the aging process of the mitral valve is triggered by age-related changes in VICs or VECs activity or by an intrinsic ECM age-related degradation, the genetic model may play a crucial role in the determinism of the disease and need more investigation.

2.3.4 Genetics of MVP

In this part, we present the 3 models that helped us to understand the genetic of MVP. Firstly, we summarize the genetics of syndromic MVP, in which several known MVP-related genes, such as fibrillin 1 (*FBN1*), TGF beta were firstly identified to be associated with MVP. Secondly, we will illustrate the family studies on MVP. Last and the most importantly, we put more stress on

the GWAS of non-syndromic MVP which is a common cardiac disease and the topic of interest of my thesis.

2.3.4.1 Genetics of syndromic MVP

MVP commonly occurs in patients with connective tissue disorders, including Marfan syndrome (MFS), Loeys-Dietz syndrome, Aneurysm-osteoarthritis syndrome, Ehlers–Danlos syndrome, osteogenesis imperfecta, and pseudo-xanthoma elasticum (PXE) (Levine et al. 2015; Le Tourneau et al. 2018). In addition, MVP can be caused by Williams-beuren syndrome (WBS), Larsen-like syndrome (Le Tourneau et al. 2018).

Marfan syndrome, an autosomal-dominant multisystemic inherited disorder, is known to be caused by a mutation on chromosome 15q21.1 in the gene *FBN1*, which encodes the glycoprotein fibrillin-1 and is the principal component of microfibrils in the ECM (Dietz et al. 1991; Collod-Béroud et al. 2002; Thacoor 2017). A recent study using 965 patients who carried pathogenic *FBN1* mutations reported an increased MV disease rate with age in the patients. The rate of MVP increased from 43% at 30 years old to 77% at 60 years old. Thirteen percent of Marfan patients were referred for MV surgery (Détaint et al. 2010). This work clearly showed that MVD is associated with subdiagnostic variants of Marfan syndrome.

As reported by Kaartinen et al. (Kaartinen et al. 2003), in Marfan syndrome, fibrillin-1 deficiency leads to excessive amounts of activated TGF beta to be liberated from the matrix in heart valves. In the development of heart valves, TGF beta signaling induces endocardial transformation that is essential for the proper formation of endocardial atrioventricular cushions and valves (Nakajima et al. 2000). Excessive signaling by members of the TGF beta superfamily led to many features of Marfan syndrome (Levine et al. 2015).

Antagonism of TGF beta signaling can prevent the thickening and prolongation of mitral valves (Ng et al. 2004). TGF beta family members can be inhibited by blocking type-1 angiotensin II receptor (Cohn et al. 2007; Brooke et al. 2008), a pathway that is implicated in nonsyndromic myxomatous MVP (Geirsson et al. 2012). These reports have demonstrated that the dysregulation of TGF beta plays a role in mitral valve diseases in MFS. However, the

abnormalities of the *FBN1* gene in patients or families with isolated MVP have not been revealed (Levine et al. 2015).

Loeys–Dietz syndrome shows autosomal dominant inheritance like MFS (Loeys et al. 2005). This disease is associated with heterozygous mutations in genes encoding the TGF beta receptors components, *TGFBR1* and *TGFBR2*, or the *TGF* beta *2* ligand, *TGFB2* (Loeys et al. 2005; Lindsay et al. 2012). Some individuals with *TGFBR1* or *TGFBR2* mutations have phenotypic overlap with Marfan syndrome but are not associated with MFS in multiple organ systems (Loeys et al. 2005). Four of 14 (29%) affected individuals with Loeys–Dietz syndrome had substantial MVP (Loeys et al. 2005). In Attias's study, the prevalence of MVP among individuals with *FBN1* mutations was compared with that in individuals with *TGFBR2* mutations. More people had MVP with *FBN1* mutation (45%) versus with *TGFBR2* mutations (21%) (Attias et al. 2009)). Individuals who carry *FBN1* mutations or *TGFBR2* mutations were significantly related to MVP compared to controls (Attias et al. 2009).

Like MFS and Loeys–Dietz syndrome which has the main feature of connective tissue disorders, aneurysm-osteoarthritis syndrome has been reported to be caused by mutations in *SMAD3* located in chromosome 15 and encoding a key regulator in the TGF beta pathway (van de Laar et al. 2011). MVP phenotype in patient with aneurysm-osteoarthritis syndrome has been reported in 45% (10/22) – 50% (18/36) patients (van de Laar et al. 2011; van de Laar et al. 2012), which is equally common in patients with MFS (54 %) (Faivre et al. 2007) and less common in LDS (21%) (Attias et al. 2009).

Except for syndromes related to TGF beta activation and signaling, Ehlers–Danlos syndrome(EDS) was found to be caused by mutations in various genes, especially in collagen-related genes, and other connective tissue proteins in the EMC, such as elastin, proteoglycans and macromolecular proteins (https://emedicine.medscape.com/article/943567-overview). But only 6% of EDS patients were reported to have MVP (Atzinger et al. 2011). Pseudoxanthoma elasticum (PXE) is a rare autosomal recessive disorder related to elastic fibers in connective tissue and caused by mutations in the *ABCC6* gene. MVP is an infrequent finding in approximately 4.5% in PEX patients (Prunier et al. 2013). Williams-beuren syndrome (WBS) is a rare developmental disorder related to the elastin gene on chromosome 7q and is highly

associated with supravalvar aortic stenosis (65%) and pulmonary stenosis (51%) (Del Pasqua et al. 2009). WBS was reported to have 6.2% MVP phenotype (Del Pasqua et al. 2009). MVP has been occasionally reported in osteogenesis imperfecta, but the link between them was not well studied (Le Tourneau et al. 2018). Furthermore, a mutation in *B3GAT3* encoding glucuronyltransferase-I (GlcAT-I) can cause impaired proteoglycan maturation which is termed as Larsen-like syndrome. In a family with five affected children, MVP was found in four patients (Baasanjav et al. 2011).

2.3.4.2 Family studies on non-syndromic MVP

MVP was suggested as a familial inheritance disease with a familial prevalence of 47% (Hancock et al. 1966; Weiss et al. 1975). Family genetic linkage studies identified several susceptibility loci for MVP. The first locus was mapped in a 54-cM region in chr16p11.2-12.1 with a significant maximum two-point LOD score (Disse et al. 1999). The second locus was mapped in a 4.3-cM region between the markers D11S1923 and D11S1331 on chromosome 11p15.4 (Freed et al. 2003). Thereafter, an 8.2-Mb candidate region on chr13q31.3-q32.1 with a peak nonparametric linkage was found (Nesta et al. 2005). Furthermore, X-linked myxomatous valvular dystrophy, a rare disorder with histopathological features similar to severe MVP has been mapped to chromosome Xq28 (Kyndt et al. 1998). These family studies have identified the related regions, but the identification of causal mutations and genes at these loci is challenging.

Until recently, filamin A (*FLNA)* on chromosome X was the only known causative gene for MVP. Kyndt (Kyndt et al. 2007) identified *FLNA* by the expanded linkage analysis of the previously mapped region of chromosome Xq28 in 4 families with XMVD of French, British, African, and Chinese origin (Kyndt et al. 2007). *FLNA* is important for the organization of the cytoskeleton and ECM. The role of *Flna* in the development of MVP was confirmed by using conditional gene knockout mouse model. Tissue-specific ablation of *Flna* altered the development of atrioventricular valves and resulted in abnormally enlarged mitral valve (Sauls et al. 2012). A recent study of 246 relatives from 4 families with MVD that harbor *FLNA* missense mutations reported that patients with *FLNA* mutations have severe MVP and require early diagnosis and close monitoring. As one of the sub-phenotypes of MVP, FLNA-MVP, an X

linked disease is characterized by a severe elongation of AL and PL in male patients and a mild to moderate expression in female patients. Besides, frequent polyvalvular involvement, mitral leaflet thickening, and elongation were found in FLNA related MVD. Myxomatous degeneration has also been reported in FLNA-MVP (Le Tourneau et al. 2018).

In a follow-up study of *MMVP2* locus, using targeted sequencing in four patients from three families with severe MVP, Durst et al. (Durst et al. 2015) have shown that the loss of function mutations in dachsous cadherin-related 1 (*DCHS1*) gene is associated with MVP. This gene is a member of the cadherin superfamily, and it encodes calcium-dependent cell-adhesion protein. *DCHS1* deficiency in mitral VICs derived from MVP patients resulted in altered cellular patterning and increased VICs migration (Durst et al. 2015). However, given that both missenses *DCHS1* mutations are rare in the population and MVP is a common disease, additional genetic causes of sporadic MVP still need to be determined.

2.3.4.3 Genetics of non-syndromic MVP -- GWAS

As MVP is identified as a common cardiac disease, genetic determinants of the isolated MVP worth being noticed. Differ from the family linkage studies, GWAS provides the opportunity to discover the causal genetic variants in a large population.

Collaborating with researchers from France, the United States, Canada, and Spain, a two-stage GWAS was conducted (Dina et al. 2015) (**Figure 14**). In the discovery stage, 1412 MVP patients and 2439 control individuals from two independent French cohorts were genotyped using Illumina/Affymetrix genotyping arrays. After quality control on the genotyped data, we estimated the haplotypes using SHAPE-IT (Delaneau et al. 2008) and then performed imputation on the 1000 Genomes Phase I reference panel using IMPUTE V2 (Howie et al. 2009). Finally, we did association tests on approximately 4.8 million common SNPs whose minor allele frequencies are greater than 0.05 in the Europeans. Meta-analyses of the 2 independent French GWAS was performed and 47 SNPs at 23 loci were selected for the second stage, including three genome-wide significant loci (P-value $< 5\times10^{-8}$). In the second stage, we performed an intermediate meta-analysis with two follow-up sets from Spain and United States and further identified 24 SNPs at 15 loci, which were replicated in another two follow-up sets from Canada

and France. The multiple stage approach used in GWAS is helpful and efficient when genotyping all SNPs for every individual is unachievable. Three addition loci reached the genome-wide significant level in the global meta-analysis containing 2,864 cases and 9,218 controls.

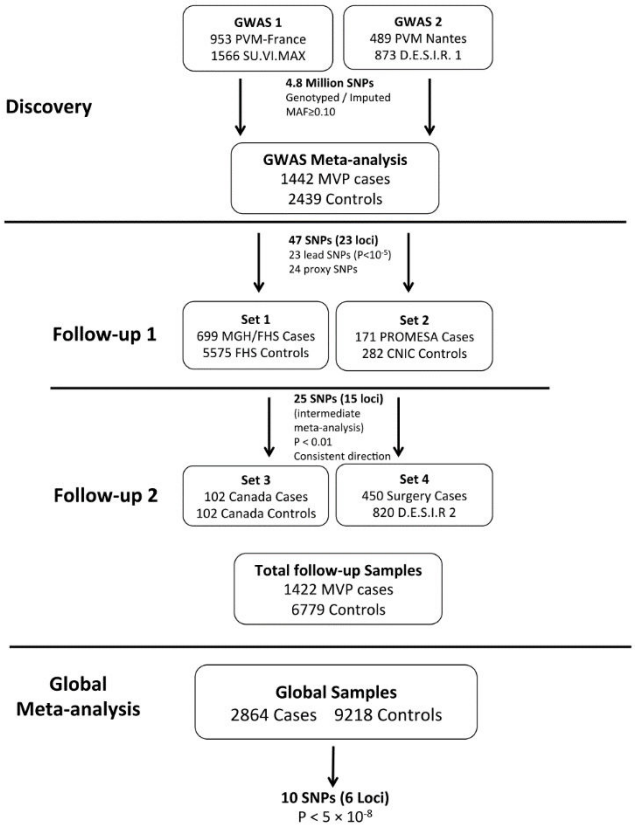

Figure 14. Study design of current GWAS and the filter strategy of SNPs. *Two discovery GWAS were followed up by another four replications. Then, we performed a global meta-analysis. Picture from (Dina et al. 2015).*

In summary, 6 risk loci associated with MVP validated the polygenic nature of MVP (**Figure 1-15**). Two of them were intronic (*LMCD1* and *SMG6*), and four were located in intergenic (*IGFBP5-TNS1*, *SETD4-CBR1*, *PITBNB-MN,* and *PCNX-SIPA1L1*) regions, which is not surprising, as it has been shown that ~95% of the disease-associated SNPs described by GWAS are located in non-coding regions of the human genome (Maurano et al. 2012).

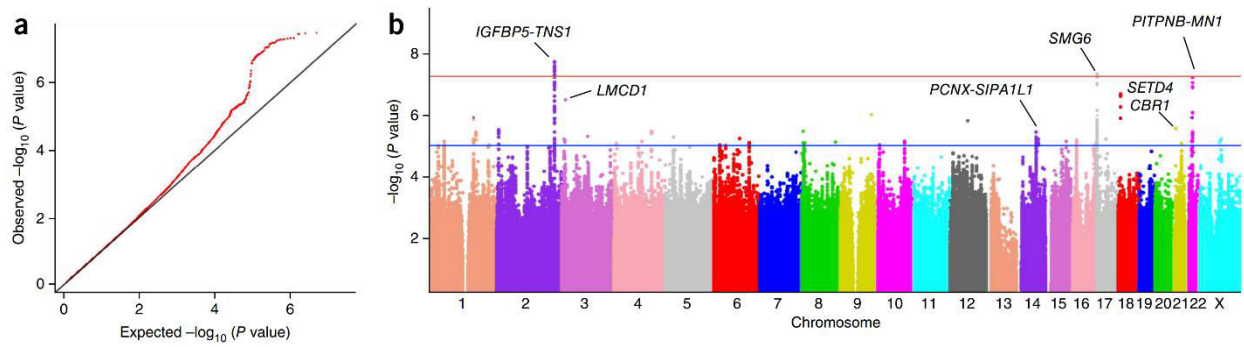

Figure 15. Quantile-Quantile (QQ-plot) *(a) and Manhattan plot (b) of the discovery genome-wide meta-analysis. Approximately 4.8 million SNPs in the GWAS meta-analysis are shown. The blue line indicates the P-value threshold used for follow-up (P-value < 1×10^{-5}), and the red line represents the genome-wide significance threshold (P-value < 5×10^{-8}). Picture from (Dina et al. 2015).*

At the *IGFBP5-TNS1* locus located on chromosome 2, *TNS1* was identified as a causal gene for MVP (**Figure 16a**). *TNS1* encodes a protein that localizes focal adhesions, interacts with actin, and may play a role in cell migration. The knockdown of *tns1* in zebrafish model resulted in a significant AV regurgitation and diminished the aggregation of endothelial cells at the developing valve (**Figure 16b,c**). The expression of tensin 1 was detected during valve morphogenesis and maintained during adulthood in mice and was localized to endothelial and valvular interstitial cells (**Figure 16d**). In 9-month-old *Tns1*$^{-/-}$ mice, enlarged posterior mitral leaflets were shown compared to the wild-type mice (**Figure 16e**). The valves *Tns1* knock out mice showed the evidence of myxomatous degeneration, loss of normal matrix stratification and expansion of proteoglycan expression in *Tns1*$^{-/-}$ mitral leaflets were observed (**Figure 16f**). These results support that *TNS1* is the best candidate contributing to MVP at this locus.

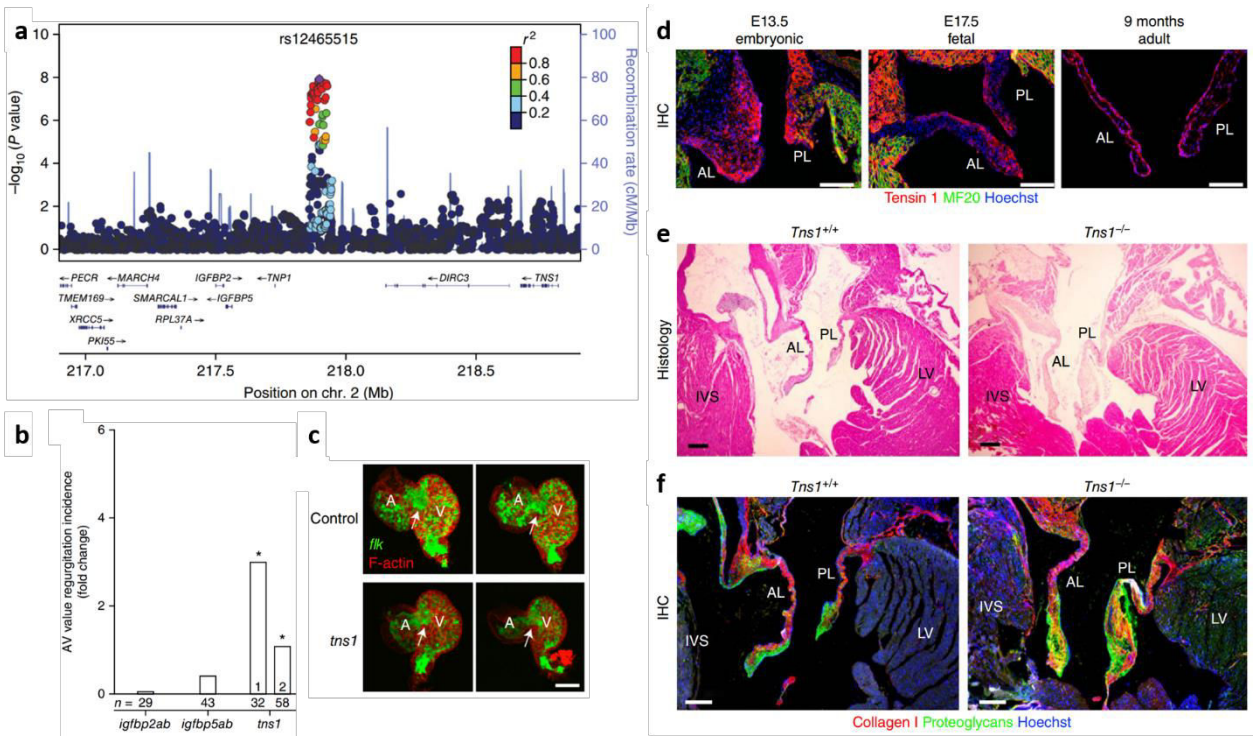

Figure 16. Morpholino-mediated knockdown of orthologs to tns1 in zebrafish embryos and Tensin 1 expression and knockout phenotype in mouse model. *(a) The association signals at IGFBP5-TNS1 locus using Locuszoom. Genes around ±500 bp of the top SNP are shown. (b) Fold change in observed atrioventricular regurgitation in 72-hpf zebrafish embryos after morpholino-mediated knockdown. All results are relative to clutchmate controls. n is the number of biological replicates per morpholino. *P < 0.05. (c) Two-dimensional projections of z-series image stacks taken on hearts excised from a control zebrafish (injected with nontargeting morpholino) and one with tns1 knockdown. This figure shows the distribution of F-actin, which is highly expressed in the functional myocardium (stained in red) and the EGFP expression, a marker of the endothelium (colored in green). A: atrium; V: ventricle. White arrows indicate the AV canal. Scale bar, 50 µ. (d) Tensin 1 expression in the developing mouse heart. IHC was performed for tensin 1 (red) at E13.5, E17.5 and 9 months of age. MF20 labels myocytes (green); Hoechst labels nuclei (blue). Scale bars, 200 µm. (e) Tns1 knockout mice show enlarged mitral leaflets. The wild-type ($Tns1^{+/+}$) vs knockout ($Tns1^{-/-}$) mice. Scale bars, 200 µm. (f) Myxomatous mitral leaflets were observed in Tns1 knockout mice. IHC for collagen (red); proteoglycans (green). Scale bars, 200 µm. AL, anterior leaflet; PL, posterior leaflet; LV, left ventricle; IVS, interventricular septum. Picture adapted from (Dina et al. 2015).*

Another highlighted locus is located on chromosome 3 in the intron of *LMCD1* (**Figure 17a**), which encodes a protein that may act as a co-regulator of transcription with other transcription factors. Knockdown of *lmcd1* in zebrafish model led to a significant AV regurgitation and cardiac looping defects (**Figure 17b**). It has been shown that *Lmcd1* is a transcriptional co-factor

that represses *Gata6*, an important regulator of cardiac development (Rath et al. 2005). Also, *LMCD1* plays an important role in the calcineurin/NFAT signaling, which is required for endocardial to mesenchymal transition during heart valve morphogenesis. These data provide a strong rationale for studying the role of this gene in MVP disease. One more gene that should be mentioned is *PITPNB* located on chromosome 22 at intergenic *PITBNB-MN* locus. The ortholog of Pitpnb was shown to be expressed in endothelial and interstitial mitral valve cells during development in adult mouse mitral leaflets. The role of *PITPNB* and other candidate genes from MVP-associated loci should be further investigated.

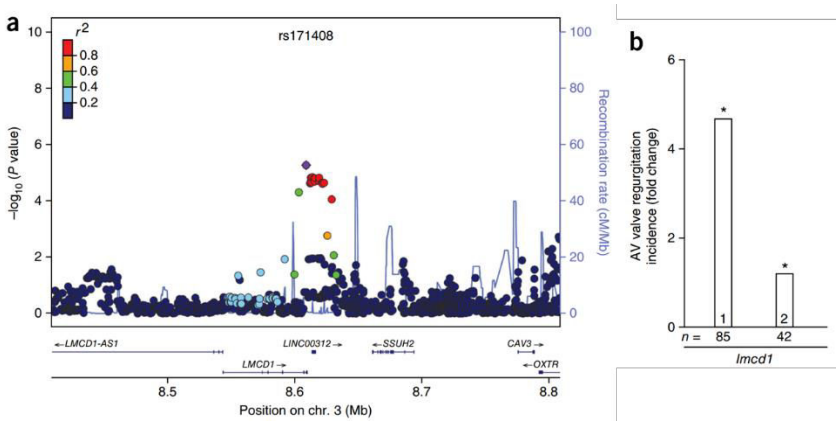

Figure 17. The association signals at LMCD1 locus and morpholino-mediated knockdown of orthologs to lmcd1 in zebrafish embryos. *(a) The association signals at LMCD1 locus using Locuszoom. Genes around ±500 bp of the top SNP are shown. (b) Fold change in observed atrioventricular regurgitation in 72-hpf zebrafish embryos after morpholino-mediated knockdown. All results are relative to clutchmate controls. n is the number of biological replicates per morpholino. *P < 0.05. Picture adapted from (Dina et al. 2015).*

Progresses in genetic studies of non-syndromic MVP helped to understand this disease. However, in those small number of loci, I only know the candidate function with certainty for 1 locus (TNS1). For the remaining 5 loci, I still do not have a clear idea of what are the genes involved in what are the functions. The genes that have been linked to non-syndromic MVP so far can only explain a proportion of MVP heritability. Thus, GWAS-identified MVP-associated loci still need to be explored and further investigation is necessary to identify additional MVP risk loci and to discover the biological mechanisms of MVP.

Objectives of the thesis

The objective of my thesis was to improve our knowledge about the genetic basis of mitral valve prolapse, the most common indication for valve repair or replacement due to mitral regurgitation, and understand the biological mechanisms behind the genetic association.

To achieve these goals, I performed GWAS-based studies in the largest existing case-control available so far for MVP; 2 French GWAS and a newly defined UK Biobank case-control GWAS.

In my first study, I re-analyzed the meta-analysis that was available at the beginning of my Ph.D. that included (1442 cases and 2439 controls) and applied three gene-sets and pathways enrichment analyses. I hypothesize that loci under the stringent statistical threshold (P-value$<5\times10^{-8}$) are likely to be located near biologically relevant genes with candidate and novel mechanisms to MVP. This study provided interesting and unexpected clues on the biology of MVP related to genetic susceptibility and allowed the prioritization and follow-up of the seventh locus of susceptibility to MVP on Chr 1 named *GLIS1* (Article 1). I was also involved in a large biological study that identified cilia genes as significantly enriched and associated with MVP (See my contribution in Appendix C) that I will not detail in this manuscript. In the second study, I attempted to overcome the limitations of the small sample size where GWAS data is available by, on the one hand, increasing genotyping information using large and recently available imputation panels (HRC and TOPMed) imputation panel and, on the other hand adding a new case-control dataset from the UKBiobank study (434 cases and 4527 controls). In this study, I was able to improve the association resolution at previously confirmed loci and describe 6 new suggestive loci. I also provided an extensive functional annotation and provide promising new genes near the new loci, mainly *PDGFD* on Chr 11 and *ACTN4* on Chr 19 (Article 2)

PART II

Leveraging genome-wide association studies data to inform the biology behind the genetic risk for mitral valve prolapse

1 CHAPTER 1 – GWAS-Driven enrichment analyses to understand the biology behind genetic association with MVP

1.1 Main results

I firstly applied the SNP ratio test (O'dushlaine et al. 2009) (SRT) method that uses simulated datasets to estimate the significance of a given pathway. I analyzed ~2 million SNPs that mapped within genes and found that among the top 10 enriched pathways ($P_{empirical}$<0.05), several pathways are consistent with cardiac function, including viral myocarditis, hypertrophic cardiomyopathy, dilated cardiomyopathy, and cardiac muscle contraction. I also noted enrichment in the phosphatidylinositol signaling system, MAPK and WNT signaling that are important for focal adhesion and the gap junction functions. Next, I employed i-GSEA4GWAS (Zhang et al. 2010; Zhang et al. 2015) which highlighted 244 enriched pathways (FDR < 0.05) for MVP. According to the semantic similarity between GO Biological Process, the enriched pathways have been classified into four groups. Except for two of the known mechanisms in MVP that have been enriched, such as actin filament organization, cytoskeleton-dependent intracellular transport and cell migration, other two pathways related to the regulation of transcription and heart development, vasculature development were highlighted as well.

The SRT method and i-GSEA4GWAS were applied in the study of primary cilia (additional paper, check appendix C for my contribution in details), which is supposed to be the cause of MVP. Instead of using existing gene sets/pathways, I added 278 known cilia ciliopathy genes as a pre-defined gene set ("Cilia Gene Set"). I tested whether the "Cilia Gene Set" is enriched by the associated SNPs with MVP. Among nominally associated SNPs in the GWAS (SNPs<0.05), I found that SNPs that map in the "Cilia Gene Set" are more likely (trend P=0.03) to be enriched for associated SNPs with MVP using SRT. The significant enrichment in "Cilia Gene Set" (p=0.009, FDR= 0.024) was found using i-GSEA4GWAS as well.

I also applied the integrative method DEPICT(Pers et al. 2015) and obtained 309 nominally enriched gene sets for MVP (P-value<0.05), which were clustered into 36 clusters. DEPICT enrichment results were consistent with the i-GSEA4GWAS analysis. DEPICT also enriched the

MVP-risk loci into 21 tissue/cell types including cardiovascular system, mesenchymal stem cells, and muscle. To simplify the reporting of our results in the article, following suggestions from journals editors during the reviewing process, we only provided the main results from iGSEA4GWAS-and DEPICT.

Among the loci that contribute to the enrichment analyses, I provided a focus on an association signal located on chromosome 1, rs1879734 (OR=1.23, 95%CI=1.14-1.33, P-value =1.29×10^{-7}), which is in the first intron of *GLIS1*. We replicated this locus in a UK Biobank MVP case-control GWAS (Lead SNP: rs18797434, $OR_{Overall}$=1.20, 95%CI=1.12-1.27, P =4.36×10^{-10}). I performed functional annotation for all 111 SNPs in moderate LD with rs18797434 (r2>0.5) that showed association with MVP (P-value < 0.05). I note that rs1879734 is located in a TFBS for endothelial transcription factor GATA2 and shows nominal eQTL association with GLIS1 expression in GTEx (p=0.048). I also found that Glis1 is expressed predominantly in nuclei of endothelial cells of the valves as well as the valvular interstitial cells during embryonic development. The atrioventricular regurgitation was found in Glis1 knockdown in zebrafish.

1.2 Article 1

Mengyao Yu, Adrien Georges, Nathan R Tucker, SergiyKyryachenko, Katelyn Toomer, Jean-Jacques Schott, Francesca N. Delling, Leticia Fernandez-Friera, Jorge Solis, Patrick T. Ellinor, Robert A. Levine, Susan A. Slaugenhaupt, Albert A. Hagège, Christian Dina, Xavier Jeunemaitre, David J. Milan, Russell A. Norris, Nabila Bouatia-Naji. **GWAS-driven Gene-set Analyses, Genetic and Functional Follow-up Suggest GLIS1 as a Susceptibility Gene for Mitral Valve Prolapse.** *Circulation: Genomic and Precision Medicine*; Accepted

Online link: https://www.ahajournals.org/doi/10.1161/CIRCGEN.119.002497

Other publication: bioRxiv 433268; doi: https://doi.org/10.1101/433268

GWAS-driven Gene-set Analyses, Genetic and Functional Follow-up Suggest *GLIS1* as a Susceptibility Gene for Mitral Valve Prolapse

Running title: *Yu et al., Biological mechanisms at MVP genetic risk loci*

Mengyao Yu, Msc[1,2], Adrien Georges, PhD[1,2], Nathan R. Tucker, PhD[3,4], Sergiy Kyryachenko, PhD[1,2], Katelyn Toomer, PhD[5], Jean-Jacques Schott, PhD[6,7,8], Francesca N. Delling, MD[9], Leticia Fernandez-Friera, MD[10,11] Jorge Solis, MD[10,11], Patrick T. Ellinor, MD[3,4], Robert A. Levine, MD[12], Susan A. Slaugenhaupt, PhD[13], Albert A. Hagège, MD, PhD[1,2,14], Christian Dina, PhD[6,7,8], Xavier Jeunemaitre, MD, PhD[1,2,15], David J. Milan, MD[3*], Russell A. Norris, PhD[5*], Nabila Bouatia-Naji, PhD[1,2*]

[1] INSERM, UMR970, Paris Cardiovascular Research Center, 75015 Paris, France

[2] University Paris Descartes, Sorbonne Paris Cité, Faculty of Medicine, 75006 Paris, France

[3] Cardiovascular Research Center, Cardiology Division, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Boston, Massachusetts 02114 USA,

[4] Precision Cardiology Laboratory, The Broad Institute, Cambridge, MA 02124 USA

[5] Cardiovascular Developmental Biology Center, Department of Regenerative Medicine and Cell Biology, College of Medicine, Children's Research Institute, Medical University of South Carolina, 171 Ashley Avenue, Charleston, SC 29425, USA

[6] Inserm U1087; institut du thorax; University Hospital Nantes, France

[7] CNRS, UMR 6291, Nantes, France

[8] Université de Nantes, Nantes, France.

[9] Department of Medicine, Division of Cardiology, University of California, San Francisco, 94143

[10] HM Hospitales-Centro Integral de Enfermedades Cardiovasculares HM-CIEC, Madrid, Spain

[11] Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, Spain

[12] Cardiac Ultrasound Laboratory, Cardiology Division, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Boston, Massachusetts 02114 USA

[13] Center for Human Genetic Research, Massachusetts General Hospital and Department of Neurology, Harvard Medical School, 185 Cambridge St., Boston, MA 02114 USA.

[14] Assistance Publique – Hôpitaux de Paris, Department of Cardiology, Hôpital Européen Georges Pompidou, 75015, Paris, France

[15] Assistance Publique – Hôpitaux de Paris, Department of Genetics, Hôpital Européen Georges Pompidou, 75015, Paris, France

* Joint last authors.

Corresponding author

Nabila BOUATIA-NAJI, PhD

Paris Cardiovascular Research Center, INSERM UMR970, 56 Rue Leblanc, F-75015, Paris FRANCE.

Tel. +33 1 53 98 79 95

Fax. +33 1 53 98 79 52

Email. nabila.bouatia-naji@inserm.fr

Manuscript word count: 5,262

Subject Terms: Valvular Heart Disease; Genetic, Association Studies; Developmental Biology; Computational Biology.

Abstract

Objective. Mitral valve prolapse (MVP) is a common heart valve disease, the most frequent indication for valve repair and/or replacement. MVP is characterized by excess extracellular matrix secretion and cellular disorganization which leads to bulky valves that are unable to co-apt correctly during ventricular systole resulting in mitral regurgitation, and it is associated with sudden cardiac death. Here we aim to characterize globally the biological mechanisms underlying genetic susceptibility to MVP to better characterize its triggering mechanisms.

Methods and Results. Using pathway enrichment tools (i-GSEA4GWAS, DEPICT) applied to GWAS data, we show that genes at risk loci are involved in biological functions relevant to actin filament organization, cytoskeleton biology and cardiac development. The enrichment for positive regulation of transcription, cell proliferation and migration motivated the follow-up of *GLIS1*, a transcription factor from the Krüppel-like zinc finger family. We followed-up the association with MVP in an independent dataset of cases and controls This research was conducted using the UK Biobank Resource. In combination with previously available data, we found a genome-wide significant association with MVP (OR=1.20, P=4.36 $\times 10^{-10}$). Immunohistochemistry experiments in mouse indicated that *Glis1* is expressed during embryonic development predominantly in nuclei of endothelial and interstitial cells of mitral valves. We also show that *Glis1* knockdown using morpholinos cause atrioventricular regurgitation in zebrafish.

Conclusions. Our findings define globally molecular and cellular mechanisms underlying common genetic susceptibility to MVP and implicate established and unprecedented mechanisms. Through the *GLIS1* association and function, we point at regulatory functions during cardiac development as common mechanisms to mitral valve degeneration.

Keywords: heart valve disease, mitral valve prolapse, GWAS based enrichment analysis, GLIS1

Abbreviations

MVP:	Mitral Valve Prolapse
EndoMT:	Endothelial To Mesenchymal Transition
ECM:	Extracellular Matrix
GLIS1:	GLIS family zinc finger 1
SNP:	Single Nucleotide Polymorphism
GWAS:	Genome-Wide Association Studies
DEPICT:	Data-driven Expression Prioritized Integration for Complex Traits
i-GSEA4GWAS:	Improved Gene Set Enrichment Analysis for GWAS
VAI:	Variant Annotation Integrator
ENCODE:	Encyclopedia of DNA Elements
IGB:	Integrated Genome Browser
eQTLs:	Expression Quantitative Trait Locus
KEGG:	Kyoto encyclopedia of genes and genomes
GO:	Gene Ontology
FDR:	False Discovery Rate

Introduction

Mitral valve prolapse (MVP) is a common heart valve disease with important health consequences and an estimated prevalence of 2.4% in general populations[1, 2]. It is defined as an abnormal mitral leaflet displacement into the left atrium during systole[3] and is the most common and an increasing indication for surgical repair of mitral regurgitation[4, 5]. MVP in associated with infective endocarditis in native valves[6] and is an established a risk factor for heart failure and sudden death in community-based studies[1, 4, 7]. Prospective studies have showed that asymptomatic patients with low-risk presentation (e.g., moderate mitral regurgitation and ejection fraction ≥50%) can develop adverse MVP-related events, indicating wide heterogeneity in outcomes among individuals[4, 8].

The process that leads to MVP and myxomatous degeneration is poorly understood. Previous reports have suggested that a potential mechanism is activation of quiescent valvular interstitial cells to myofibroblasts. These activated cells secrete matrix metalloproteinases that drive collagen and elastin fragmentation and release TGF-β that in turn promotes further cell proliferation and myofibroblast differentiation[8]. However, the triggering mechanisms of myxomatous degeneration are still to be identified, as are the etiological mechanisms of advert valvular complications in patients.

Previous familial and population genetic studies have contributed to the current understanding of MVP biology. The identification of mutations in the filamin A gene (*FLNA)*[9] followed by research in mice[10] confirmed the role of this actin-binding protein during fetal valve development, mainly by providing stability to F-actin networks and linking them to cell membranes, which protect cells against mechanical stress[11]. More recently, we found that loss of function mutations in *DCHS1*, coding a protein from the cadherin superfamily involved in cell adhesion, cause familial MVP[12]. *DCHS1* deficiency in valvular interstitial cells altered migration and cellular patterning from patients and provoked loss of cell polarity during valve development in mice and zebrafish models[12]Additional clues to the etiology of MVP came through genome-wide association studies (GWAS) where we identified six risk loci that are robustly

associated with genetic susceptibility to MVP[13]. Genetic and biological evidence supported the role of two genes; tensin 1 (*TNS1*), a focal adhesion protein involved in cytoskeleton organization, and a transcription factor called LIM and cysteine-rich domain 1 (*LMCD1*[13]. However, a broad assessment of the biological functions and potential mechanisms provided by the GWAS data is still missing.

In this study, we hypothesize that several loci, including those not prioritized according to the stringent GWAS statistical threshold (P-value$<5\times10^{-8}$), could contain biologically relevant genes and unknown mechanisms relevant to MVP onset. We applied computational-based analytical methods to the currently available GWAS data to 1) highlight globally enriched biological mechanisms at MVP loci, 2) characterize their expression pattern in tissues and 3) identify biologically pertinent genes for follow-up.

Methods

Details on the populations and methods are available in the Supplementary methods. Because of the sensitive nature of the data collected for this study, requests to access the human MVP datasets from qualified researchers trained in human subject confidentiality protocols may be sent to the study leaders of each cohort and/or to the corresponding author. The experimental data that support the findings of this study are available from the corresponding author upon reasonable request. Clinical and genetic data were obtained after full written consents in each cohort studied and after approval of studies protocols from local ethics committees (See and[14] for details). All animal experiments were performed under protocols approved by the Institutional Animal Care and Use Committee, Medical University of South Carolina. Prior to cardiac resection, mice were euthanized in accordance with the Guide for the Care and Use of Laboratory Animals (NIH Publication No. 85-23, revised 1996.). Zebrafish experiments were performed in accordance with approved Institutional Animal Care and Use Committee (IACUC) protocols.

Results

Gene-set enrichment analyses at MVP loci

We first employed i-GSEA4GWAS, a gene set enrichment analysis tool for meta-analysis of GWAS summary statistics that presents the advantage of calculating enrichment based on all association results generated by the meta-analysis of GWAS, without restricting the analysis to a significance threshold[15]. This method highlighted 244 pathways as significantly enriched (FDR < 0.05) for MVP (Supplementary Table 1). The most enriched pathway was Biocarta EDG-1 pathway with 26 out of 27 genes harbouring associated variants. EDG-1, also known as Sphingosine 1-phosphate receptor (S1PR1), is a G-protein receptor reported to interact with Filamin A[16], which is mutated in familial forms of MVP[17]. Globally, enriched gene-sets clustered into four main groups related to valve biology according to a semantic similarity between GO Biological Processes terms using REVIGO webserver[18]. Two groups supported known mechanisms in MVP and contained terms "actin filament organization" and "membrane organization", "cytoskeleton-dependent intracellular transport" and "cell migration" (Figure 1). Interestingly, one cluster contained several enriched terms related to regulation of transcription (e.g "positive regulation of transcription", DNA-template). Finally, this analysis highlights several enriched GO terms for developmental mechanisms including "heart development" and "vasculature development", in addition to related terms "regulation of angiogenesis" (Figure 1, Supplementary Table 1).

We then applied the integrative method *DEPICT* that allows to consider larger resources of gene-sets and takes into account tissue expression and mouse knockout phenotypes[19]. We used the recommended association cut-off (P_{GWAS}-value $<10^{-5}$) that identified 39 independent MVP loci harboring 50 genes. We obtained 309 nominally enriched gene sets for MVP (P-value<0.05) and the Pearson distance matrix provided by DEPICT defined 36 clusters from these enriched gene sets (Supplementary Table 2). Overall, these enrichment analyses and clustering were consistent with the i-GSEA4GWAS analyses, and supports

cell-cell and cell-ECM interactions, vascular development and regulatory genes to be highly relevant for MVP genetic susceptibility (Supplementary Table 2).

Enrichment for expression in tissues and cell types

Based on reconstituted gene-tissue matrix generated by DEPICT, we analyzed the expression profile from arrays-based human transcriptomic data and found that 21 tissues or cell types were enriched for the expression of genes at MVP risk loci. The most enriched physiological system is the cardiovascular system with five significantly enriched tissues. Connective tissue, epithelial and stem cells, specifically mesenchymal stem cells and muscle were the most enriched cell types and tissues for the expression of MVP genes (Supplementary Figure 1, Supplementary Table 3).

Prioritization of *GLIS1* as a susceptibility gene for MVP

Prioritization from contribution to enrichment analyses

Among the loci that contributed to the enrichment analyses, we provide a focus on an association signal located on chromosome 1 that deserved prioritization in the light of gene function candidacy and clues from the results obtained in the enrichment analyses. The association context at the *GLIS1* locus is presented in Figure 2a. The most associated SNP with MVP is rs1879734 and locates in the first intron, as reported previously (OR=1.23, 95%CI=1.14-1.33, P-value =1.29×10^{-7})[13]. *GLIS1* encodes a GLI-related *Kruppel*-like zinc finger protein that functions as an activator and repressor of transcription. We found that i-GSWA4GWAS revealed *GLIS1* as the best-ranked gene in six significantly (FDR<0.05) and two suggestively (FDR<0.25, P-value<0.05) enriched gene sets, all related to regulation of transcription, a biological mechanism that we found to be enriched for MVP loci (Figure 1, Supplementary Tables 1 and 4). We also found that *GLIS1* significantly contributes to the enrichment of several tissues and cell types namely the connective tissue (*GLIS1* Z-score=2.3) and cells (*GLIS1* Z-score=2.8), and mesenchymal stem

cells (Z-score=2.3) (Supplementary Table 3). We indicate in Supplementary Table 4 all 6 enriched gene sets at an FDR<0.05 where GLIS1 was found to be the best ranked gene by i-GSEA4GWAS analysis.

Follow-up of the association and functional annotation at the GLIS1 *locus*

We followed-up the association between the *GLIS1* locus in a UK Biobank MVP case control dataset and obtained an overall genome-wide significant association with MVP (Lead SNP: rs18797434 $OR_{Overall}$=1.20, 95%CI=1.12-1.27, P =4.36×10^{-10}, Figure 3). To further characterize this association signal, we performed functional annotation for all 111 SNPs in moderate LD with rs18797434 (r^2>0.5) that showed nominal association in the meta-analysis of GWAS with MVP (Pvalue<0.05). All SNPs mapped to introns 1 and 2 in *GLIS1* (Supplementary Table 5, Figure 2b). We found that 47 SNPs locate in DNase Hypersensitive region (DNASEV) in diverse normal tissues, 16 locate in transcription factor binding sites (TFBS) and 55 are nominal (P<0.05) eQTLs for *GLIS1* in heart atrial appendage tissue from GTEx (N=264). We note that rs1879734 is located in a TFBS for endothelial transcription factor GATA2 and shows nominal eQTL association with GLIS1 expression in GTEx (p=0.048). The examination of histone marks (Figure 2b) in mesenchymal stem cell originated from H1-hESC shows the presence of robust enhancer mark (H3K4me1) in the vicinity of rs1879734. In H1-hESC, strong signal of enhancer mark (H3K4me1) was presented at some SNPs in the locus. However, none of the 111 SNPs showed signals of histone marks in heart related tissue/cell types (pulmonary artery endothelial cell, cardiac mesoderm, heart left ventricle) (Figure 3b). *In silico* 4C experiment using four cell types including THP-1[20], HUVEC[21], H1-hESC[22] and NHEK[21] (Supplementary Figure 2) showed that the associated SNPs at this locus physically interact only with *GLIS1* regulatory sequences suggesting it is likely to be the potential target and causal gene at this locus.

Glis1 Expression during heart development in mouse

GLIS1 belongs to GLI-similar 1-3 (GLIS1-3) subfamily of Krüppel-like zinc finger proteins that act either as activators or repressors of gene transcription and are involved in multiple physiological processes and

diseases[23]. However, little is known about GLIS1 and its function in the cardiovascular system. In the absence of data about its role in valve biology, we performed immunohistochemistry experiments of the mouse ortholog, *Glis1* during embryonic, fetal and adult time points to study its pattern of expression during valve development. We found that Glis1 is expressed predominantly in nuclei of endothelial cells of the valves as well as the valvular interstitial cells during embryonic development (Figure 4). As the valves mature during fetal gestation, Glis1 is retained in a subset of endocardial and interstitial cells. By 6-months of age, Glis1 is much weaker in the valve leaflet with only scant cytoplasmic staining in endocardial cells. Weaker expression of Glis1 could be detected in the myocytes, epicardium and endocardium of the ventricular myocardium. These data show that Glis1 is embryonically expressed and that levels of this protein are rarely detected in the postnatal mouse, suggesting an important role for Glis1 in regulating valve morphogenesis during early development.

Knockdown of Glis-1 cause atrioventricular regurgitation in zebrafish

To analyze the potential effects of *GLIS-1* on valvular development and function, we chose to investigate its expression in the zebrafish. Due to genome duplication, zebrafish are predicted to have two orthologues of GLIS-1, *glis1a* and *glis1b*. We designed antisense morpholino oligonucleotides to target splice junctions in each with aims of rendering the transcript non-functional. Knockdown of *glis1a* was 65% efficient at 72 hours post fertilization, but had no discernible effect on atrioventricular valve function (Supplementary Figure 3). Knockdown of *glis1b* at 72 hours post fertilization was robust but slightly less efficient (37.3% and 36.5% reduction Figure 5a, 5b) and had minimal effects on the overt morphology of the developing embryo (Figure 5c). Compared to controls, this knockdown resulted in a significant increase in the incidence of severe atrioventricular regurgitation with a combined fold increase of 1.6 (P=0.01). (Figure 5d, Supplementary Video).

Discussion

Our study defines globally molecular and cellular mechanisms underlying common genetic susceptibility to MVP and implicate established and unprecedented mechanisms. Through the genetic association and the study of the expression during valve development, we point at a potential mechanism of mitral valve degeneration involving *GLIS1,* a regulatory gene from the Krüppel-like zinc finger family of proteins.

Our gene-set enrichment and integrative analyses applied to MVP currently available meta-analysis of GWAS confirmed that genes in associated loci are predominantly involved in the regulation of cell adherence and migration, cytoskeleton biology, focal adhesion and the interaction with the extracellular matrix. An example of this enrichment is the EDG-1/SIPR1 pathway that we describe as highly enriched for MVP associated SNPs. SIPR1 is well-known for its critical role in heart development and vascular maturation[24] as a key signaling pathway for actin assembly and lamellipodia formation. Interestingly, the reported interaction between SIPR1 and Filamin A[16] may encourage further functional investigation of the potential role of EDG1 in mitral valve degenerative process.

We also show that MVP genes are highly expressed in the cardiovascular system, connective and mesenchymal tissues. Several top contributing genes at MVP loci were transcription factors, especially to gene sets related to cardiovascular development. Consistently, we report enrichment for the protein localization to nuclear gene sets. Here, we followed-up specifically *GLIS1*, which was one of the top associated SNPs that we confirmed as a new genome-wide significant risk locus for MVP. In addition to the expression in developing heart valves, *GLIS1* was the most significantly contributing gene to the enrichment of several gene sets related to the regulation of transcription. Little is known about GLIS1 function in connection with valve biology. We found that GLIS1 is expressed during developing valves in mouse valve endocardial and interstitial cells, suggesting for the first time its potential regulatory role

during heart development. There is evidence for Glis1 to markedly increase the efficiency of generating induced pluripotent stem cells from mouse somatic cells in the presence of OCT3/4, SOX2 and KLF4[11]. This study demonstrated that GLIS1 directly interacts with KLF4 and induces the expression of Forkhead box genes, especially FOXA2 and several WNT genes to enhance mesenchymal to epithelial transition, a mechanism required for cell reprogramming[25]. Interestingly, a common variant in *WNT8A*, member of the WNT genes family, is associated with atrial fibrillation[26] and was previously shown to be regulated by GLIS1[11].

This work presents however several limitations. One critical step when performing enrichment analyses in GWAS loci is the attribution of variants to genes. The pathway analyses of i-GWSEA4GWAS and DEPICT rely on SNPs mapped to genes using physical distance and LD block information. This have limited the number of genes analyzed at risk loci and excluded more distant genes of interest. There is established evidence in favor of the functional role of distant long-range regulatory variants in predisposition to complex diseases[27]. In most cases, associated variants in GWAS loci are less likely to regulate the closest genes and be hundreds of kilobases away from culprit genes. This limitation explains the absence from the prioritization list generated by DEPICT of *TNS1*, a focal adhesion protein-coding gene that we have previously incriminated in MVP through genetic and functional investigation [13].

In conclusion, our pathway investigation supports that genes near MVP associated loci are involved in biological functions relevant to cell adhesion and migration during cardiac development and in response to shear stress, and highlight the importance of regulatory mechanisms. Our study also provides genetic *in silico* and *in vivo* functional follow-up of *GLIS1,* a transcription factor with potential role in mechanisms linked for EndoMT and cell migration during heart development and valve degenerative process.

Acknowledgments

We acknowledge the contribution of the Leducq Foundation, Paris for supporting a transatlantic consortium investigating the physiopathology of mitral valve disease, for which the genome-wide association study was a major project (coordinators: R.A.L. and A.A.H.). We acknowledge investigators who contributed access to validation in cohorts: Yohan Bossé and Philippe Pibarot from *Institut universitaire de cardiologie et de pneumologie de Québec-Université Laval*, Quebec, Canada, Ramachandran S. Vasan and Emilia J. Benjamin from the Framingham Heart Study, USA and Ronen Durst, Hassadah Hebrew University Medical Center, Jerusalem, IL.

Sources of funding

This study was supported by a PhD scholarship from the Chinese Scientific Council to MY, French Agency of Research (ANR-16-CE17-0015-02) to NB-N. AG and NB-N are supported by a European Research Council grant (ERC-Stg-ROSALIND-716628). The work at MUSC was performed in a facility constructed with support from the National Institutes of Health, Grant Number C06 RR018823 from the Extramural Research Facilities Program of the National Center for Research Resources. Other funding sources: National Heart Lung and Blood Institute: HL131546 (RAN), COBRE GM103342 (RAN), GM103444 (RAN), HL127692 (DJM, SAS, RAN, RAL), American Heart Association: 17CSA33590067 (RAN) and HL140187 (NRT). NIH K23 HL116652 (FND), contracts HHSN268201500001I, contract No1-HL25195, R01-HL-080124 and R01HL126136-01A1 (VSR), NIH R01 HL128099 (RAL), NIH HL141917 (RAL), Fondation Leducq Grant number 07CVD04 (RAL, AAH, SAS, XJ, JJS) and the Ellison Foundation, Boston, MA (RAL). The recruitment of the MVP France cohort was supported by the French Society of Cardiology (SFC).

Disclosures

None

REFERENCES

1. Freed LA, Levy D, Levine RA, Larson MG, Evans JC, Fuller DL, et al. Prevalence and clinical outcome of mitral-valve prolapse. *New England Journal of Medicine*. 1999;341:1-7.
2. Delling FN, Rong J, Larson MG, Lehman B, Osypiuk E, Stantchev P, et al. Familial clustering of mitral valve prolapse in the community. *Circulation*. 2014:CIRCULATIONAHA-114.
3. Guy TS and Hill AC. Mitral valve prolapse. *Annual review of medicine*. 2012;63:277-292.
4. Avierinos J, Gersh BJ, Melton LJ, Bailey KR, Shub C, Nishimura RA, et al. Natural history of asymptomatic mitral valve prolapse in the community. *Circulation*. 2002;106:1355-1361.
5. Gammie JS, Chikwe J, Badhwar V, Thibault DP, Vemulapalli S, Thourani VH, et al. Isolated Mitral Valve Surgery: The Society of Thoracic Surgeons Adult Cardiac Surgery Database Analysis. *Ann Thorac Surg*. 2018.
6. Katan O, Michelena HI, Avierinos JF, Mahoney DW, DeSimone DC, Baddour LM, et al. Incidence and Predictors of Infective Endocarditis in Mitral Valve Prolapse: A Population-Based Study. *Mayo Clin Proc*. 2016;91:336-42.
7. Freed LA, Benjamin EJ, Levy D, Larson MG, Evans JC, Fuller DL, et al. Mitral valve prolapse in the general population: the benign nature of echocardiographic features in the Framingham Heart Study. *J Am Coll Cardiol*. 2002;40:1298-1304.
8. Levine RA, Hagege AA, Judge DP, Padala M, Dal-Bianco JP, Aikawa E, et al. Mitral valve disease--morphology and mechanisms. *Nat Rev Cardiol*. 2015;12:689-710.
9. Kyndt F, Gueffet J-P, Probst V, Jaafar P, Legendre A, Le Bouffant F, et al. Mutations in the gene encoding filamin A as a cause for familial cardiac valvular dystrophy. *Circulation*. 2007;115:40-49.
10. Sauls K, De Vlaming A, Harris BS, Williams K, Wessels A, Levine RA, et al. Developmental basis for filamin-A-associated myxomatous mitral valve disease. *Cardiovascular research*. 2012;96:109-119.
11. Maekawa M, Yamaguchi K, Nakamura T, Shibukawa R, Kodanaka I, Ichisaka T, et al. Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. *Nature*. 2011;474:225-9.
12. Durst R, Sauls K, Peal DS, deVlaming A, Toomer K, Leyne M, et al. Mutations in DCHS1 cause mitral valve prolapse. *Nature*. 2015;525:109-113.
13. Dina C, Bouatia-Naji N, Tucker N, Delling FN, Toomer K, Durst R, et al. Genetic association analyses highlight biological pathways underlying mitral valve prolapse. *Nat Genet*. 2015;47:1206-11.
14. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:e1001779.
15. Zhang K, Cui S, Chang S, Zhang L and Wang J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res*. 2010;38:W90-W95.
16. Maceyka M, Alvarez SE, Milstien S and Spiegel S. Filamin A links sphingosine kinase 1 and sphingosine-1-phosphate receptor 1 at lamellipodia to orchestrate cell migration. *Molecular and cellular biology*. 2008;28:5687-97.
17. Le Tourneau T, Le Scouarnec S, Cueff C, Bernstein D, Aalberts JJJ, Lecointe S, et al. New insights into mitral valve dystrophy: a Filamin-A genotype-phenotype and outcome study. *European heart journal*. 2018;39:1269-1277.
18. Supek F, Bosnjak M, Skunca N and Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011;6:e21800.

19. Pers TH, Karjalainen JM, Chan Y, Westra H-J, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun*. 2015;6:5890.
20. Phanstiel DH, Van Bortle K, Spacek D, Hess GT, Shamim MS, Machol I, et al.. Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. *Molecular cell*. 2017;67:1037-1048 e6.
21. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159:1665-1680.
22. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485:376-80.
23. Jetten AM. GLIS1-3 transcription factors: critical roles in the regulation of multiple physiological processes and diseases. *Cell Mol Life Sci*. 2018;75:3473-3494.
24. Cannavo A, Liccardo D, Komici K, Corbi G, de Lucia C, Femminella GD, et al. Sphingosine Kinases and Sphingosine 1-Phosphate Receptors: Signaling and Actions in the Cardiovascular System. *Front Pharmacol*. 2017;8:556.
25. Scoville DW, Kang HS and Jetten AM. GLIS1-3: emerging roles in reprogramming, stem and progenitor cell differentiation and maintenance. *Stem cell investigation*. 2017;4:80.
26. Ellinor PT, Lunetta KL, Albert CM, Glazer NL, Ritchie MD, Smith AV, et al. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat Genet*. 2012;44:670-5.
27. Kleinjan DA and van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet*. 2005;76:8-32.

Figures

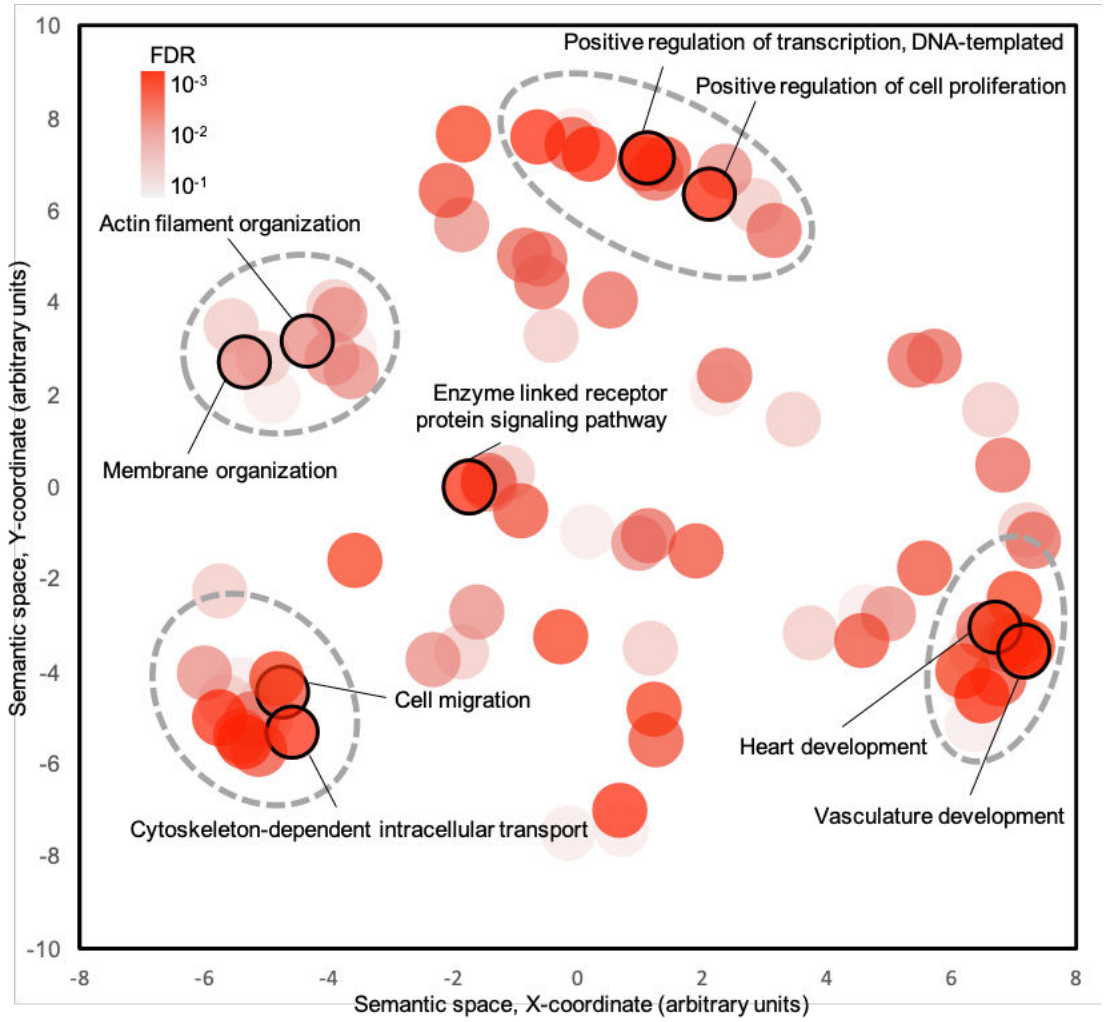

Figure 1: Semantic proximity between GO terms enriched for MVP associated risk loci. Each GO biological process term identified as enriched in MVP-associated variant with FDR>0.05 was compared to other associated terms to measure their semantic proximity. The matrix of semantic proximity was collapsed into two main component axis (X and Y) represented here. Each bubble represents one GOBP term, and the diameter of the bubble is proportional to the inverse logarithm of FDR. Main groups were delimited, and representative GOBP terms are indicated for each group. The full list of enriched terms is provided as Supplementary table

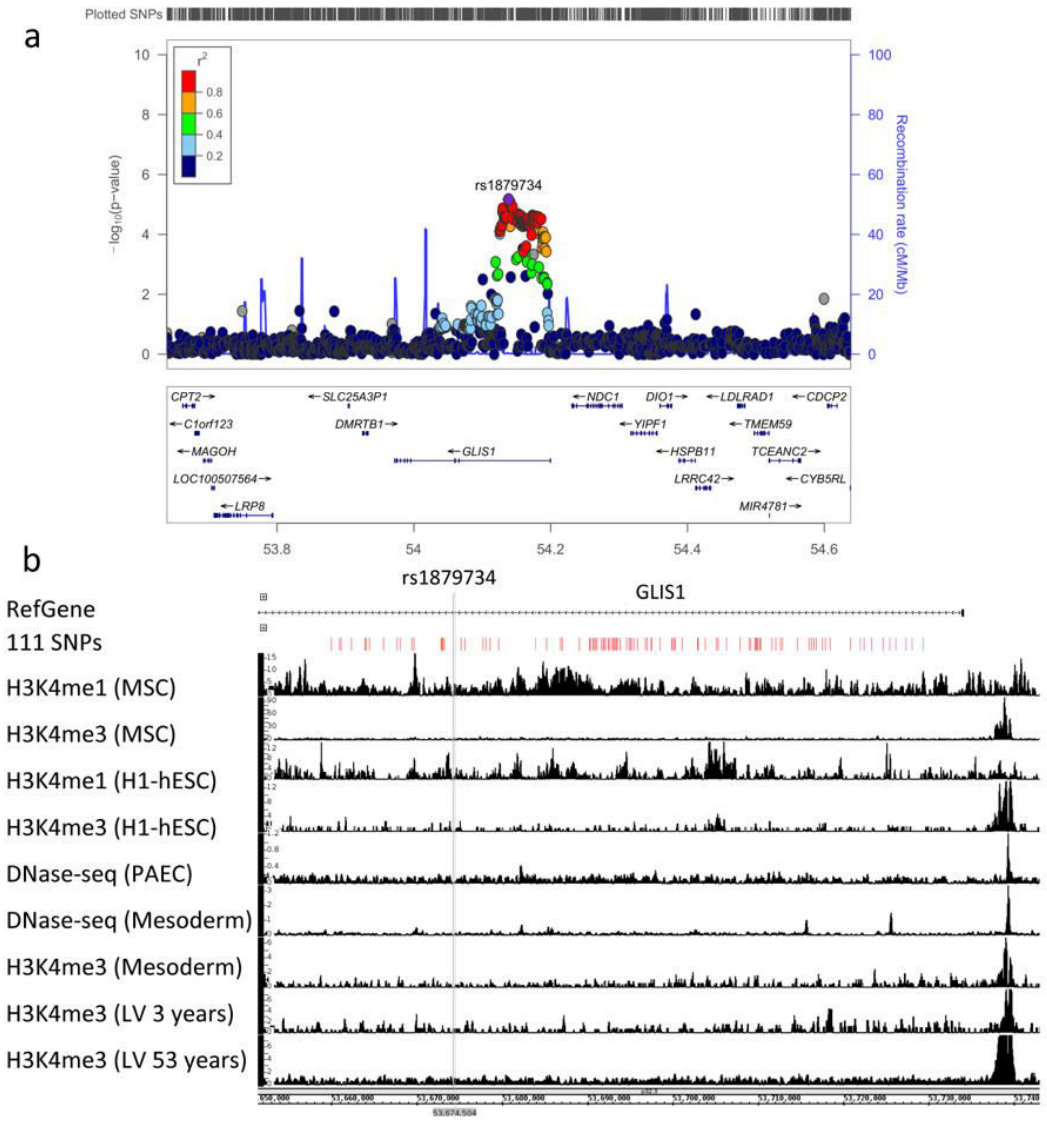

Figure 2. Genomic context and functional annotation of the association signal observed in the GWAS meta-analysis. **(a)** The regional association plot was generated using LocusZoom and displays surrounding genes (±500 Kb). The association signal is intronic to *GLIS1*. Round points represent SNPs in this region and purple point represent SNP rs1879734, the top associated SNP. **(b)** Visualization of histone marks and DNAse-seq density profiles in several tissues/cells based on ENCODE data. From top to bottom: reference gene; the selected 111 SNPs that is in high LD with rs1879734; H3K4me1 and H3K4me3 were from mesenchymal stem cell originated from H1-hESC; H3K4me1 and H3K4me3 from H1-hESC; DNase-seq from pulmonary artery endothelial cell; DNase-seq and H3K4me3 from cardiac mesoderm; H3K4me1 and H3K4me3 from heart left ventricle.

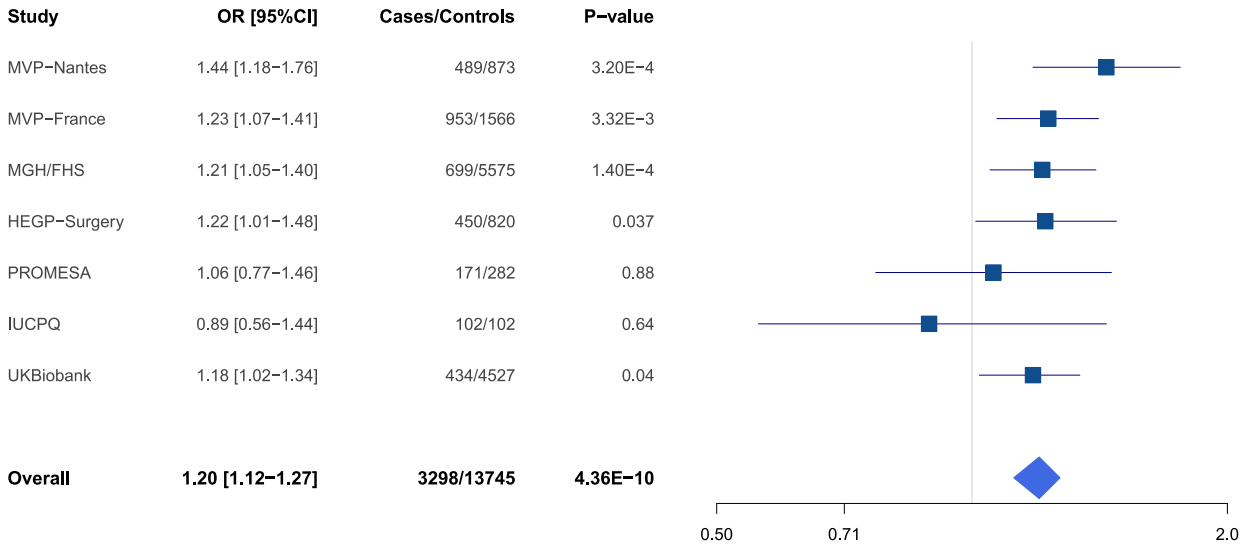

Figure 3. Forest plot shows associations in individual studies and the overall association between rs1879734 and mitral valve prolapse. OR: odds ratio, CI: confidence interval. All cases control cohorts were previously described, except UK Biobank, which is novel and reported in details in methods.

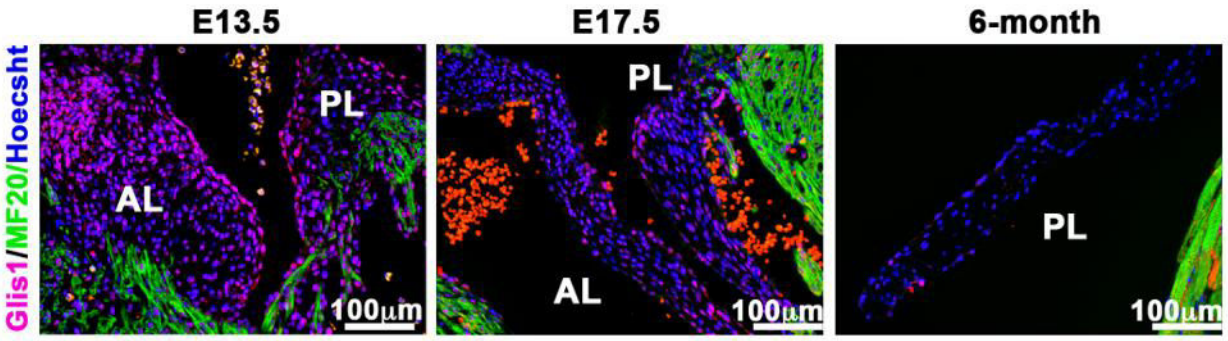

Figure 4. Glis1 expression in mouse developing and adult heart. Cardiac expression of Glis1 (red) was analysed at embryonic (E13.5), foetal (E17.5) and adult time points. Glis1 was detected during valve morphogenesis in mice (E.13 and E.17 stages), specifically during the completion of endothelial to mesenchymal transformation and valve sculpting and elongation and undetected in the adult valve (6 months). Glis1 is detected in nuclei from endothelial and valvular interstitial cells. Green tags are for MF20 marking sarcomeric myosin-myocytes, Blue is Hoescht coloration that indicates nuclei.

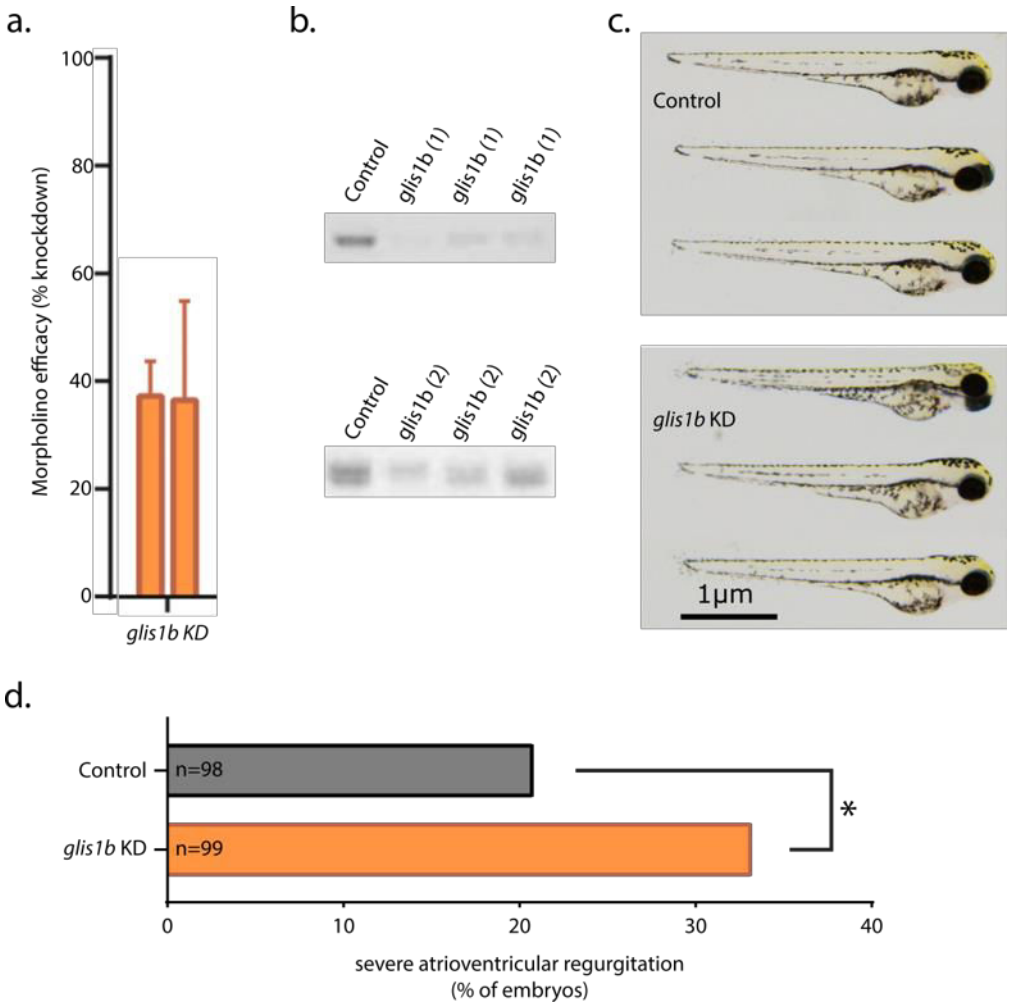

Figure 5. Assessment of cardiac regurgitation in zebrafish after morpholino knockdown of *glis1b*. **(a)** Morpholino-mediated knockdown efficacy. Efficacy for glis1b in embryonic zebrafish was measured by RT-PCR. **(b)** Representative gel images from analysis of morpholino efficacy. Control indicates samples amplified from control-injected embryos. All samples were obtained from 72-hpf embryos. **(c)**. Brightfield micrographs displaying gross morphology of 72-hpf embryos following glis1 knockdown. Scale bar represents 1 µm. **(d)** Fold change in observed atrioventricular regurgitation in 72-hpf zebrafish embryos after morpholino-mediated knockdown. All results are relative to clutchmate controls.

Supplementary material

GWAS-driven Gene-set Analyses, Genetic and Functional Follow-up Suggest *GLIS1* as a Susceptibility Gene for Mitral Valve Prolapse

Mengyao Yu, Msc Adrien Georges, PhD, Nathan R. Tucker, PhD, Sergiy Kyryachenko, PhD, Katelyn Toomer, PhD, Jean-Jacques Schott, PhD, Francesca N. Delling, MD, Leticia Fernandez-Friera, MD Jorge Solis, MD, Patrick T. Ellinor, MD, Robert A. Levine, MD, Susan A. Slaugenhaupt, PhD, Albert A. Hagège, MD, PhD, Christian Dina, PhD, Xavier Jeunemaitre, MD, PhD, David J. Milan, MD, Russell A. Norris, PhD, Nabila Bouatia-Naji, PhD

Table of Content

Supplementary Methods

Study populations

The clinical characteristics of MVP patients, controls and GWAS methods have been previously described[1]. Briefly, we applied pathway-based methods to a meta-analysis of two GWAS conducted in 1,442 French patients and 2,439 population-based controls, all of European ancestry[1]. In addition to the follow-up of rs1879734, the top associated variant at *GLIS1* locus[1], in 1,422 European MVP patients and 6,779 controls, We provide an additional case control dataset of 434 cases and 4,527 controls from the UK Biobank data[2]. All participants were British, Irish and any other white background. Inclusion criteria for the cases were main or secondary diagnosis of MVP alone or combined with mitral regurgitation using electronic health records (code= ICD10). We excluded patients with rare syndromes where MVP usually occurs (e.g., Ehlers-Danlos syndrome and Marfan's syndrome), and those with ischemic and rheumatic heart disease. Controls were participants without any kind of heart disease matched for age and sex in a ratio of 1 case to 10 controls. We performed the association in the UK Biobank MVP case control dataset using PLINK (v1.07, --logistic)[3] , under the additive model adjusted for the five first principal components as covariates. Regional association plot at *GLIS1* locus was created using LocusZoom[4]. We used METAL to compute the global effect on MVP risk of SNP rs1879734. As recommended[5] p-values, sample size and direction of effect were taken into account for each study as for the effective sample size per case control set ($W= 4/(1/N_{cases}+1/N_{controls})$)[5].

Gene-set enrichment analyses

Gene set enrichment analyses using i-GSEA4GWAS

The input data in the pathway analyses were 6.6 million (M) genotyped or imputed SNPs (MAF > 0.01) as previously described[1]. i-GSEA4GWAS v2 uses SNPs and their corresponding P-values from GWAS results as input and employs SNP label to correct for bias due to differences among genes with different numbers of mapped SNPs. This correction favors the enrichment of gene sets with non-randomly associated genes

with biological plausibility instead of randomly associated genes with large numbers of mapped SNPs[6, 7]. All SNPs from the GWAS meta-analysis were used and were mapped to genes if they are exonic/intronic or to the closest genes if they are within 20 Kb upstream/downstream a gene. Of the 6.6M tested, 4.3M variants were mapped to 21,167 different genes and then genes were attributed to pathways and/or gene sets using BioCarta, KEGG, and GO terms from MSigDB v4.0. The number of genes per gene sets was limited between 20 to 200 by default. Gene sets with lower or higher number of genes out of this range were not analyzed. We restricted GO terms to experimental evidence (codes IDA IPI, IMP IGI, IEP), computational analysis evidence code (ISS) and author evidence statement (TAS). In total, 936 gene sets contributed to calculate the significant proportion based enrichment score (SPES) that considered the contribution. We used REVIGO webserver[8], a method that uses semantical similarity between gene ontology biological processes (GOBP) Terms to generate clusters of gene-sets to get a global view on i-GSEA4GWAS enrichment analysis.

Integrative analyses using DEPICT

Data-driven expression prioritized integration for complex traits (DEPICT) (V1. Beta version) is an integrative tool that utilizes diverse sources to predict 14,461 'reconstituted' gene sets, it contains a membership probability for each gene in the genome compared to the traditional binary gene sets (genes are included or not). This is a key advantage over existing tools that enables prioritization of previously poorly annotated genes, and provides more specific and powerful gene set enrichment analysis[9]. DEPICT uses predefined gene sets including protein-protein interactions database, Mouse Genetics Initiative database, Reactome, KEGG and GO terms. It also uses a co-regulation data frame downloaded from the Gene Expression Omnibus database including human and rodent expression microarrays (total arrays 77,840, including 37,427 generated in human tissues). This information was used to perform gene prioritization at loci, gene set enrichment analyses, and provide genes expression profiles in 209 tissues or cell types defined by Medical Subject Heading (MeSH). Input of DEPICT is a list of the most significantly

associated SNPs ($P_{GWAS} < 1 \times 10^{-5}$). The identification of the independent loci in DEPICT was obtained using PLINK (v1.90b4.7, --clump-kb 500 --clump-p1 1e-05 --clump-r2 0.1). Loci were mapped to genes using LD ($r^2 > 0.5$). Genes where several SNPs map are counted once. To group overlapping gene sets, pairwise Pearson's correlation coefficient of all enriched gene sets was calculated and the Affinity Propagation (AP) method[9, 10] was used to cluster gene sets that are highly correlated to automatically define independent clusters based on the Pearson distance matrix. We set significance for enrichment to P-value < 0.05, given that no gene sets reached an FDR<0.05.

Variants Annotation of the *GLIS1* locus

We used the UCSC genome browser tool Variant Annotation Integrator (VAI) to annotate SNPs at the *GLIS1* locus, and indicate if SNPs were located on DNase Hypersensitive region or transcription factor binding site generated by ENCODE[11]. Tissue specific expression quantitative trait loci (eQTL) annotation was extracted from GETx portal (https://www.gtexportal.org/home/, Release V7, dbGaP Accession phs000424.v7.p2). Annotation of the 111 selected SNPs by Human ChIP-Seq (markers: H3k4ME1 and H3k4ME3) and DNase-Seq were downloaded from ENCODE (https://www.encodeproject.org/) and visualized by Integrated Genome Browser (IGB)[12]. In the absence of valvular cells in this database, selected tissues/cells used in this analyses are mesenchymal stem cell originated from H1-hESC[13], H1-hESC[13], pulmonary artery endothelial cell[14], cardiac mesoderm[14] and heart left ventricle[13]. Accession numbers of those files are (respectively): ENCFF152YQG (H3K4me1), ENCFF712CJP (H3K4me3), ENCFF623ZAW (H3K4me3), ENCFF593OAZ (H3K4me1), ENCFF719ZEX (DNase-seq), ENCFF591TLE (H3K4me3), ENCFF213FJN (DNase-seq), ENCFF094USN (DNase-seq), ENCFF254JZR (DNase-seq). The annotation

of the *GLIS1* region used the Hi-C data from four cell types including THP-1[15], HUVEC[16], H1-hESC[17] and NHEK[16] from ENCODE, all visualized by IGB.

Protein detection in mouse embryos and adult hearts

In all immunohistochemistry experiments, 5-min antigen retrieval was performed with VectaStain in a pressure cooker (Cuisinart). The antibodies used for immunohistochemistry were: Glis1 (Novus, NB100-41087) and myosin heavy chain (Developmental Hybridoma Banks, MF20). Primary antibodies were used at a 1:100 dilution; Hoechst 33342 (nuclear stain) was used at a 1:10,000 dilution. Appropriate secondary antibodies were used for detection. Three time points were used in immunohistochemistry: (i) completion of the endothelial-to-mesenchymal transition (EndoMT) (embryonic day (E) 13.5), (ii) valve sculpting and elongation (E17.5) and (iii) achievement of the mature adult form (at 9 months).

Zebrafish experiments

Zebrafish of the Tu-AB strain were reared according to standard techniques. Morpholinos were designed against the two zebrafish orthologues of GLIS1: *glis1a* (5'-GAGAATGGTCGTACATACCGTGTCC) and *glis1b* (1: 5'-AAGTGCACTGAGGTCTCACCCTGTG 2: 5'-TTAAGGTCAGGTACTCACAGTGTCC). Empirically established effective doses of antisense morpholino oligonucleotides were injected at the single-cell stage, and morpholino-injected animals were compared to controls injected with a non-targeting morpholino. Analysis of atrioventricular regurgitation was performed at 72 hours post fertilization as described[1]

REFERENCES

1. Dina C, Bouatia-Naji Ni, Tucker N, Delling FN, Toomer K, Durst R, Perrocheau M, Fernandez-Friera L, Solis J, PROMESAi, Le Tourneau T, Chen M-H, Probst V, Bosse Y, Pibarot P, Zelenika D, Lathrop M, Hercberg S, Roussel R, Benjamin EJ, Bonnet F, Lo SH, Dolmatova E, Simonet F, Lecointe S, Kyndt F, Redon R, Le Marec H, Froguel P, Ellinor PT, Vasan RS, Bruneval P, Markwald RR, Norris RA, Milan DJ, Slaugenhaupt SA, Levine RA, Schott J-J, Hagege AA, MVP-F, Jeunemaitre X and LTMITRALN. Genetic association analyses highlight biological pathways underlying mitral valve prolapse. *Nat Genet*. 2015;47:1206-1211.

2. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T and Collins R. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:e1001779.

3. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ and Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559-575.

4. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis G, ,alo R. and Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26:2336-2337.

5. Willer CJ, Li Y and Abecasis Gc, ,alo R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190-2191.

6. Zhang K, Cui S, Chang S, Zhang L and Wang J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res*. 2010;38:W90-W95.

7. Zhang K, Chang S, Guo L and Wang J. I-GSEA4GWAS v2: a web server for functional analysis of SNPs in trait-associated pathways identified from genome-wide association study. *Protein cell*. 2015;6:221.

8. Supek F, Bosnjak M, Skunca N and Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011;6:e21800.

9. Pers TH, Karjalainen JM, Chan Y, Westra H-J, Wood AR, Yang J, Lui JC, Vedantam S, Gustafsson S, Esko T, Frayling T, Speliotes EK, GIoATC, Boehnke M, Raychaudhuri S, Fehrmann RSN, Hirschhorn JN and Franke L. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun*. 2015;6:5890.

10. Frey BJ and Dueck D. Clustering by passing messages between data points. *science*. 2007;315:972-976.

11. Consortium ENCODEP. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS biology*. 2011;9:e1001046.

12. Freese NH, Norris DC and Loraine AE. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics*. 2016;32:2089-95.

13. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, Antosiewicz-Bourget J, Ye Z, Espinoza C, Agarwahl S, Shen L, Ruotti V, Wang W, Stewart R, Thomson JA, Ecker JR and Ren B. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell stem cell*. 2010;6:479-91.

14. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutyavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE and Stamatoyannopoulos JA. The accessible chromatin landscape of the human genome. *Nature*. 2012;489:75-82.

15. Phanstiel DH, Van Bortle K, Spacek D, Hess GT, Shamim MS, Machol I, Love MI, Aiden EL, Bassik MC and Snyder MP. Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. *Molecular cell*. 2017;67:1037-1048 e6.

16. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES and others. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159:1665-1680.

17. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS and Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485:376-80.

2 CHAPTER 2 – Using imputation and UKBiobank to identify new MVP risk loci

2.1 Main results

Previous GWAS have identified six genome-wide significant risk loci for MVP, such as *TNS1* and *LMCD1* (Dina et al. 2015). However, limited by the small sample size, the number of genotyped variants and the imputation panel, additional MVP determinants deserve further studies.

I performed an updated GWAS using HRC and TOPMed as the re-imputation panel. Instead of focusing on the association results, we firstly compared the imputation performance of HRC and TOPMed with our MVP cohorts. Both panels showed increasing trends of imputation accuracy when the allele frequency increases, but low performance in rare allele (MAF ≤ 0.01) imputation. After the quality control filtering alleles with RSQ < 0.3, the mean imputation accuracy of the rare allele region was dramatically increased in both panels. By comparing the imputation performance in the different cohorts, we found that cohort in bigger sample size has higher imputation accuracy in all MAF region.

I then conducted two meta-analyses on 1) MVP-Paris and MVP-Nantes using HRC as the reference panel, which contains 1,486 MVP cases and 2,331 controls, ~7.1 million common genotyped or imputed SNPs (MAF>0.01); 2) MVP-Paris, MVP-Nantes and MVP-UKB using TOPMed as the reference panel, including 1920 MVP cases and 6858 controls for ~8 million genotyped or imputed common SNPs (MAF>0.01).We replicated the associated in chr2 (lead SNP: rs7595393; P-value = 1.71×10^{-10}; OR = 1.31[1.20-1.42]), which further supported that *TNS1* plays a key role in MVP. Importantly, I identified 6 interesting loci.

To prioritize the causal variants and candidate genes within all the identified associated loci, I selected 259 SNPs in high LD (r2> 0.6) with the 6 lead SNPs that showed association with MVP (P-value < 0.05). We found that several SNPs may act as causal variants, which were eQTLs and/or showed biological evidence to be a regulatory element and/or showed to be deleterious.

At last, the gene-based association analysis highlighted 12 genes highly associated with MVP (P-value <2.709e-6). The gene-property analysis showed that our input SNPs are highly expressed in muscle skeletal, artery related tissue, heart left ventricle and heart atrial appendage.

*This is an unpublished data. The content has been removed from this version

2.2 Article 2

Mengyao Yu, SergiyKyryachenko, Stephanie Debette, Philippe Amouyel, Christophe Tzourio, Christian Dina, Jean-Jacques Schott, Albert A. Hagège, Xavier Jeunemaitre, Nabila Bouatia-Naji. **Up-dated Genome-Wide Association Study and Functional Annotation Reveal New Risk Loci for Mitral Valve Prolapse.**

Manuscript in preparation

*This is an unpublished data. The content has been removed from this version

PART III -- General discussion and Perspectives

1 CHAPTER 1 – Discussion

Based on previous GWAS study (Dina et al. 2015), employed pathway-based analysis and combined with the functional analysis, we have given a global view on the molecular and cellular mechanisms underlying common genetic susceptibility to MVP in my first study (Part II chapter 1). We identified that *GLIS1*, an MVP associated gene, is participated in the regulation process during cardiac development. The pathway-based analysis approach also showed that the cilia gene set was enriched suggesting the association between cilia genes and MVP. With the rapid development of sequencing technology, the imputation panel has made tremendous progress. In my second study (Part II chapter 2), taking advantages of HRC and TOPMed, the re-imputation of the previous GWAS provided a high-resolution of MVP-associated region and increased the chance to detect MVP-related SNPs and genes. I found 6 suggestive interesting loci and provided several candidate causal genes for MVP.

In this chapter, I discuss the process that links GWAS to the pathways related to MVP, and how important the re-imputation in the existing genotyped data is and why the functional annotation is a key to pick out causal SNPs and candidate genes for MVP. Finally, we will discuss the role of GWAS in detecting genetic risk loci for MVP.

1.1 From GWAS to diseases related pathways (pathway-driven results)

GWAS has successfully identified thousands of genetic variants underlying susceptibility to complex diseases. The challenge now is to link those genetic variants to the biological mechanisms that involved to the pathogenesis of diseases, which cannot be explained by a single molecular analysis (van der Sijde et al. 2014). Exploring the biological functions and mechanisms underlying the genetic susceptibility from a systemic point of view is necessary to characterize its triggering mechanisms. Pathway analyses combine signals of a set of selected genes represented by associated variants to infer the functions and biological meanings (Kao et al. 2017).

The pathway results highly depend on the parameters set and the statistical methods used. From our gene-set enrichment analysis at MVP risk loci, we first considered that analyzing only a small

amount of most significant SNPs may only reveal a small fraction of association functions. Thus, we try to use as many associated variants as input to increase the chance to explore the most relevant pathways. As a competitive approach, iGSEA4GWAS (Zhang et al. 2010; Zhang et al. 2015) was selected to test the relative enrichment of the associated genes per pathway. The algorithms of this tool improve the sensitivity that can yield a large number of "hits" relative to other algorithms (Kwon et al. 2012). Hence, we used a stringent threshold (FDR = 0.05) to avoid overestimation of enriched pathways, even though the suggested cut-off is 0.25. Yet this setting might exclude possibly related and redundant pathways/gene sets.

However, many factors can cause bias in the pathway analysis, such as the P-value cut-off for the input, different algorithms, and background of gene sets. It was suggested that using multiple methods to evaluate the reliability of the results is necessary (Jia et al. 2011). Especially integrating multiple "-omics" platforms into the pathway analysis turns important to explain the trait-specific mechanisms (Califano et al. 2012). Hence, we selected an additional tool – DEPICT, which is a non-tissue specific pathway analysis tool, with the recommended input (SNPs with P-value < 10^{-5}) (Pers et al. 2015).

Both tools use the existing database as the pre-defined gene sets (e.g. GO, KEGG) and calculate the gene P-value based on the SNP P-value. Except for the mentioned differences in the input, there are some obvious differences between iGSEA4GWAS and DEPICT. 1) From a method perspective, iGSEA4GWAS employs SNP label permutation to analyze P-value and correct for the bias due to differences among genes with different numbers of mapped SNPs. DEPICT performs gene set enrichment based on a "reconstituted", a consequence of the gene function prediction from the gene expression data. In iGSEA4GWAS, a given set of SNPs is mapped to ±20 Kb of the upstream/downstream genes. DEPICT sets the mapping boundaries to LD (r2>0.5) to the given lead SNPs. In addition, DEPICT is a Linux and command-based tool, but iGSEA4GWAS is a web-based tool, which is more convenient for researchers who are not familiar with Linux and programming. 2) From the point of enriched results, 244 and 309 gene sets were enriched using iGSEA4GWAS and DEPICT, respectively. A small part of different enriched pathways can be found in the results from the two approaches.

Nevertheless, in our first study, no matter how different the two approaches are, we were able to identify the overlapping pathways generated by the iGSEA4GWAS and DEPICT. Genes at associated loci were enriched in pathways that are known to be related to MVP, such as "actin filament organization", "cytoskeleton-dependent intracellular transport" and "cell migration". Furthermore, pathways enriched to "heart development" and "vasculature development", in addition to related terms "regulation of angiogenesis" were enriched as well.

In the second study, we did not perform pathway analysis. Nevertheless, many highlighted genes in this study are part of pathways that are functionally close to the enriched pathways in the first study. For example, we checked the highlighted genes in the GO Biological Process pathways and found that *GRAF3* is part of "negative regulation of vascular smooth muscle contraction"; *PDGFD* belongs to "positive regulation of cell proliferation and "cell migration" (Gaudet et al. 2011); *LEMD3* is participated in "angiogenesis", "negative regulation of TGF beta receptor signaling pathway", and "negative regulation of BMP signaling pathway" (Pan et al. 2005); *ACTN4* is part of "positive regulation of cell migration" and "actin and integrin binding" (Pavalko et al. 1993); *DMPK* is part of "regulation of heart contraction" (Kaliman et al. 2005). These highlighted pathways allow us to understand better how the associated SNPs and disease-related genes are involved in the process of disease formation.

Some issues and limitations in pathway analysis need to be confronted. No matter what approaches are used to map SNPs to genes, e.g., by boundary, LD or other approaches, unrelated genes are possibly included, the size of gene sets may vary considerably. Normally, gene sets size from 20 to 200 is acceptable and widely used as a criterion in the pathway analyses several genes with big effect size may enrich the pathways that are not related to the trait. Sometimes, removing susceptibility genes or using SNP-based test may be helpful. Finally, discordant results may occur in different approaches using data from the same GWAS. This may come from different statistical methods or different thresholds. Making conclusions based on different pathway analyses may help to have a global view of the analyses (Mooney et al. 2015).

1.2 Improve the resolution of GWAS by fine mapping using imputation

Genotype imputation is a critical step in GWAS because it enables researchers to inexpensively access the whole genome sequence data from genome-wide SNP array data. The imputation accuracy of imputed variants largely depends on the reference panel. The development of NGS techniques increases the sample size of the imputation panel, from 269 individuals of HapMap Project panel (Consortium 2003) to more than 100,000 deeply sequenced samples of TOPMed (Taliun et al. 2019). Taking advantages of the large reference panel, re-imputation of the published MVP sample is helpful to improve the power of downstream association analyses (Li et al. 2009). In the updated GWAS, we got almost 1.5-fold genotyped or imputed SNPs compared to the first published MVP GWAS. The increase of identified variants has led to a higher resolution of GWAS, which boosted the association signal and increased the chance to detect causal variants.

As the sample size of large reference panel increases, the heavier imputation burden makes the local imputation impossible. Compared to Sanger Imputation Server, Michigan imputation server can upload the local genotype data easily. After choosing the imputation server, we found that imputation accuracy is proportional to the size of MAF in the results from using both HRC and TOPMed panel, but TOPMed performed slightly better, especially for rare variants. The main reason is the larger haplotypes in TOPMed, but this may also be attributed to slightly better accuracy in Eagle compared to SHAPEIT2 (Loh et al. 2016).

We finally used TOPMed as the reference panel which resulted in a higher imputation accuracy. In addition, TOPMed can impute indels. It is expected to be widely used in the GWAS analysis in the future (Taliun et al. 2019).

1.3 Gene-based association analysis

GWAS focuses on the SNP-level association test and we usually infer the candidate genes for a disease by mapping the most associated SNP to genes according to the physical distance and then investigate the gene function from the known literature and databases (Dina et al. 2015).

Alternative approaches to highlight the disease-related gene is based on the pathway analysis. For example, my first study employed pathway analysis to highlight *GLIS1* as a causal gene for MVP. However, employing all these strategies brings a lot of work to select candidate genes. Instead, gene-based association analysis using GWAS summary statistics such as effect size and SNP P-value can be a complementary method to SNP-level association test (Svishcheva et al. 2019). In our second study, the increased resolution of GWAS may make the gene-based analysis more reliable, because this approach considers all the variants within a gene. Novel associated genes were highlighted by this analysis and conversely the most associated SNP in the gene can be used for the following replication (Liu et al. 2010).

1.4 Functional annotations

GWAS has identified several MVP-associated variants, which are in non-coding regions of the genome. It is suggested that those non-coding variants and variants in strong LD with them act as "*cis*-regulatory elements (CREs)" (Maurano et al. 2012; Schaub et al. 2012). They might change the genetic regulation of one or more target genes to affect disease risks. To determine the causal variants for MVP, different experimental approaches and genomic datasets were used to help annotate those putative associations. Checking if variants are eQTILs, which are more likely to alter the expression levels of target genes to affect disease risk, is a key step in the functional annotation.

In addition, the MVP-related variants might be enriched in CREs that map to the DNase hypersensitivity region (Degner et al. 2012), histone markers known to be associated with transcriptional regulatory activity, such as H3K4me1, H3K4me3, and CTCF,or transcription factor (TF) binding (Maurano et al. 2012; Schaub et al. 2012). Considering these factors together, RegulomeDB score is widely used to measure the function of the lead SNP at each locus and SNPs in strong LD with the lead SNP. Numbers of SNPs showed small scores and/or were found nearby the H3K4me1, H3K4me3, and/or CTCF peak, suggesting that those SNP may act as CREs to affect the target genes. Moreover, some of the studied SNPs are more likely to be deleterious and influence the function of target genes.

In the first study of my thesis, we did functional annotation of the SNPs using the UCSC genome browser tool Variant Annotation Integrator (VAI), which offered the cell types name for DNASEV and TFBS for genes. However, in the second study, a powerful tool FUMA (Watanabe et al. 2017) provides a more comprehensive way to annotate SNPs. It annotates SNPs with the information from RegulomeDB score (Boyle et al. 2012) and CADD score (Wang et al. 2010), as well as eQTL for genes at different tissue/cell types. Based on these information, I was able to provide a deeper understanding of the importance of the SNPs and how likely it is a causal variant for the newly identified suggestive loci for MVP.

Using diverse approaches to functionally annotate the top SNPs and SNPs in LD with them, we proposed several potential causal variants for MVP loci. The chromatin interaction analyses also provide many genes that interact with the detected MVP risk locus. However, all those functional annotations were tissue/cell types specific and we can only use heart-related tissue due to a lack of information on mitral valve related tissue/cell. The lack of accessibility of valve tissue is an important limitation for annotation of genetic loci for MVP. In both of my studies, we investigated if the most associated SNP and SNPs of strong LD are eQTLs in heart atrial appendage and heart left ventricle tissues as tissue proxies for mitral valves, or if SNPs is a potential regulatory element in mesenchymal stem cell, cardiac artery cell cardiac mesoderm and heart left ventricle.

1.5 Limitations of GWAS analysis

Despite the tremendous achievements of GWAS in identifying common variants associated with complex diseases and in pointing out the most potential disease-related genes, converting these findings into clinical applications remains limited. This is because 1) it is difficult to identify the causal variants or genes for diseases and traits. The 7 confirmed MVP-associated SNPs and 6 suggestive new loci are located in the intron or intergenic of the genes and they are annotated out of the protein coding region, indicating that they may play a role as regulatory elements (van der Sijde et al. 2014). Moreover, hundreds of other SNPs in strong LD with those lead SNPs are potentially causal variants as well. More work needs to be done to test the hypotheses generated by the functional annotation we provided for MVP loci. Many genes were mapped close to those

associated SNPs. Some of them were not obviously linked with MVP according to the existing literature on their biological function. These uncertain reasons make it more challenging to the identification of causal variants and genes and to clarify the direct functional consequences of the variants and genes. 2) The lead variants mapped to the MVP genes have a moderate effect size, and the odds ratios were between 1.2 and 1.4. This can be explained by the combined impact of the common variants. However, the identified SNPs cannot explain all the genetic risk for MVP. A large proportion of the heritability of MVP calls for more efforts to explore (Hofker et al. 2014). 3) GWAS mainly studies the common variants, but rare and structural variants with a big effect size might have an important impact on the diseases (Visscher et al. 2017). In our previous MVP GWAS (Dina et al. 2015), the MAF was set to 0.05 which excluded all the rare variants. In this updated GWAS, we used the TOPMed as the imputation panel, which can impute the indels. However, we only analyzed the common variants as well (MAF > 0.01) because many rare variants had low imputation accuracy and were excluded. Actually, whether for common, rare or structure variants, the sample size, genotyping array, and imputation panel can affect the association results. The sample size was limited by the difficulties in recruiting enough MVP patients and also because many cardiologists tend to under-estimate the health burden associated with MVP. But the statistical power to identify associations between variants and diseases greatly depends on the sample size (Visscher et al. 2017). Larger sample size may lead to new discoveries, especially the rare variants (Visscher et al. 2017). Fortunately, after we added MVP-UKB, we can use meta-analysis on several cohorts to improve the power to identify novel associations. Another basic factor to affect the success of GWAS is the density of genotype methods. GWASs rely on the LD and haplotype block, but not all the SNPs fall into a specific haplotype block with the tagSNPs. Potential causal SNPs might have low linkage with the tagSNPs and cannot be imputed or have a very low imputation accuracy, which cannot be included in the association analysis (Hofker et al. 2014). This problem might be solved using a large imputation panel that contains enough haplotypes (Loh et al. 2016). Moreover, in our case-control study, using genotyped data from different SNP chip resulted in the removal of many variants. This has led to low imputation accuracy and is the reason why we suggested to use the same SNP chip in the case-control study to improve the resolution of GWAS.

2 CHAPTER 2 -- Perspectives

To further extend the genetic study of MVP, the improvement can be made on the top of the list is to increase the sample size. Fortunately, we will have a new GWAS dataset from our collaborators in the US (David Milan, Patrick Ellinor) that is expected to include at least 2000 new cases. This upcoming MVP cohort was imputed with TOPMed as well and we will apply the same quality control criteria to the imputed data. A global meta-analysis on the independent GWAS will be performed so that we can identify additional risk loci, including rare variants and indels. Functional annotation using multi-omics summary data will be performed for the updated results. Of course, we will run an updated pathway analysis with -omics data as well.

We still need to further investigate 1) which variant or variants are the true causal variants and what are the molecular functions behind the causal variants; 2) which identified candidate genes are truly affected by causal variants and turned to be the real causal genes; 3) how do the causal genes affect MVP risk and which pathways and networks are involved in influencing the development of MVP; 4) how do trans-eQTLs affect the expression of the target genes. Though searching for more GWAS loci is critical (Gallagher et al. 2018), deep functional dissection of those identified GWAS loci is equally important. Besides the "polygenic" theory, an ''omnigenic'' model of complex disease was proposed (Boyle et al. 2017), which suggested that the number of "core genes" can directly affect the trait and play a role in disease etiology. It is interesting to investigate in the future whether a group of "core genes" with regulatory variants in the specific tissue contributing to the pathogenesis of MVP exists and how do they affect the risk for MVP. Furthermore, building a regulation network using gene expression data to identify MVP-relevant genes may be accomplished in future work.

Our work has provided a solid foundation based on a wealth of genetic data for assessing the functions of the identified SNP and genes *in vitro and vivo*. In the following experimental studies, animal models such as zebrafish and cell types such as valve interstitial cells can be used to test the relationships between those candidate SNP/genes and their functions with MVP. Investigating the chromatin integration between those target SNPs/genes with another part of the chromosome using mitral valve specific tissue/cell type is critical to reveal the biological mechanisms of MVP.

The biological studies will be provided by current (Dr. Kyryachenko) and future colleagues of the lab.

In conclusion, my thesis work has identified several MVP risk loci, proposed the potential causal SNPs and genes for MVP, and indicated important MVP-related biological mechanisms. More insight will be provided from future biological work to disentangle the potential applications of these findings to pave the way toward novel and personalized treatments as alternatives to the costly options of valve repair or replacement through surgery.

REFERENCES

Abecasis, G. R., S. S. Cherny, et al. (2002). "Merlin--rapid analysis of dense genetic maps using sparse gene flow trees." Nature genetics **30**: 97-101.

Al-Mohaissen, M. A. and K. L. Chan (2012). "Prevalence and mechanism of tricuspid regurgitation following implantation of endocardial leads for pacemaker or cardioverter-defibrillator." Journal of the American Society of Echocardiography : official publication of the American Society of Echocardiography **25**: 245-252.

Ankeny, R. F., V. H. Thourani, et al. (2011). "Preferential activation of SMAD1/5/8 on the fibrosa endothelium in calcified human aortic valves-association with low BMP antagonists and SMAD6." PloS one **6**(6): e20969.

Anstine, L. J., C. Bobba, et al. (2016). "Growth and maturation of heart valves leads to changes in endothelial cell distribution, impaired function, decreased metabolism and reduced cell proliferation." Journal of molecular and cellular cardiology **100**: 72-82.

Anyanwu, A. C. and D. H. Adams (2007). Etiologic classification of degenerative mitral valve disease Barlow's disease and fibroelastic deficiency. Seminars in thoracic and cardiovascular surgery.

Armstrong, E. J. and J. Bischoff (2004). "Heart valve development: endothelial cell signaling and differentiation." Circ Res **95**(5): 459-470.

Ashburner, M., C. A. Ball, et al. (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nature genetics **25**: 25-29.

Attias, D., C. Stheneur, et al. (2009). "Comparison of clinical presentations and outcomes between patients with TGFBR2 and FBN1 mutations in Marfan syndrome and related disorders." Circulation **120**: 2541-2549.

Atzinger, C. L., R. A. Meyer, et al. (2011). "Cross-sectional and longitudinal assessment of aortic root dilation and valvular anomalies in hypermobile and classic Ehlers-Danlos syndrome." J Pediatr **158**(5): 826-830 e821.

Avakian, S. D., J. M. Annicchino-Bizzacchi, et al. (2001). "Apolipoproteins AI, B, and E polymorphisms in severe aortic valve stenosis." Clinical genetics **60**(5): 381-384.

Avierinos, J., B. J. Gersh, et al. (2002). "Natural history of asymptomatic mitral valve prolapse in the community." Circulation **106**(11): 1355-1361.

Baasanjav, S., L. Al-Gazali, et al. (2011). "Faulty initiation of proteoglycan synthesis causes cardiac and joint defects." American journal of human genetics **89**: 15-27.

Balding, D. J. (2006). "A tutorial on statistical methods for population association studies." Nature Reviews Genetics **7**(10): 781-791.

Barlow, J. B. and C. K. Bosman (1966). "Aneurysmal protrusion of the posterior leaflet of the mitral valve. An auscultatory-electrocardiographic syndrome." American heart journal **71**: 166-178.

Barshir, R., O. Basha, et al. (2013). "The TissueNet database of human tissue protein-protein interactions." Nucleic acids research **41**: D841-D844.

Basso, C., M. P. Marra, et al. (2015). "Arrhythmic mitral valve prolapse and sudden cardiac death." Circulation: CIRCULATIONAHA-115.

Bastian, M., S. Heymann, et al. (2009). Gephi: an open source software for exploring and manipulating networks. Third international AAAI conference on weblogs and social media.

Blake, J. A., C. J. Bult, et al. (2014). "The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse." Nucleic acids research **42**: D810-D817.

Bokanizad, B., R. Tagett, et al. (2016). "SPATIAL: A System-level PAThway Impact AnaLysis approach." Nucleic Acids Res **44**(11): 5034-5044.

Bosse, Y., P. Mathieu, et al. (2008). "Genomics: the next step to elucidate the etiology of calcific aortic valve stenosis." Journal of the American College of Cardiology **51**(14): 1327-1336.

Boyle, A. P., E. L. Hong, et al. (2012). "Annotation of functional variation in personal genomes using RegulomeDB." Genome Res **22**(9): 1790-1797.

Boyle, E. A., Y. I. Li, et al. (2017). "An Expanded View of Complex Traits_From Polygenic to Omnigenic." Cell **169**: 1177-1186.

Brooke, B. S., J. P. Habashi, et al. (2008). "Angiotensin II blockade and aortic-root dilation in Marfan's syndrome." The New England journal of medicine **358**: 2787-2795.

Browning, B. L. and S. R. Browning (2016). "Genotype Imputation with Millions of Reference Samples." American journal of human genetics **98**: 116-126.

Browning, S. R. and B. L. Browning (2011). "Haplotype phasing: existing methods and new developments." Nature Reviews Genetics **12**(10): 703.

Brunner, A. L., D. S. Johnson, et al. (2009). "Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver." Genome research **19**: 1044-1056.

Buenrostro, J. D., P. G. Giresi, et al. (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." Nature methods **10**: 1213-1218.

Bumgarner, R. (2013). "Overview of DNA microarrays: types, applications, and their future." Current protocols in molecular biology **Chapter 22**: Unit 22.21.

Buniello, A., J. A. L. MacArthur, et al. (2019). "The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019." Nucleic acids research **47**: D1005-D1012.

Burton, P. R., M. D. Tobin, et al. (2005). "Key concepts in genetic epidemiology." Lancet (London, England) **366**: 941-951.

Bush, W. S. and J. Haines (2010). "Overview of linkage analysis in complex traits." Current protocols in human genetics **Chapter 1**: Unit 1.9.1-Unit 1.918.

Cairns, B., S. Coffey, et al. (2017). "A replicated, genome-wide significant association of aortic stenosis with a genetic variant for Lp (a): meta-analysis of published and novel data."

Califano, A., A. J. Butte, et al. (2012). "Leveraging models of cell regulation and GWAS data in integrative network-based association studies." Nat Genet **44**(8): 841-847.

Carlson, C. S., M. A. Eberle, et al. (2004). "Mapping complex disease loci in whole-genome association studies." Nature **429**: 446-452.

Carneiro, M. O., C. Russ, et al. (2012). "Pacific biosciences sequencing technology for genotyping and variation discovery in human data." BMC genomics **13**: 375.

Chan, K. M. J., R. Wage, et al. (2008). "Towards comprehensive assessment of mitral regurgitation using cardiovascular magnetic resonance." Journal of cardiovascular magnetic resonance : official journal of the Society for Cardiovascular Magnetic Resonance **10**: 61.

Chinitz, J. S., D. Chen, et al. (2013). "Mitral apparatus assessment by delayed enhancement CMR: relative impact of infarct distribution on mitral regurgitation." JACC. Cardiovascular imaging **6**: 220-234.

Clarke, G. M., C. A. Anderson, et al. (2011). "Basic statistical analysis in genetic case-control studies." Nat Protoc **6**(2): 121-133.

Clement, C. A., K. D. Ajbro, et al. (2013). "TGF-beta signaling is associated with endocytosis at the pocket region of the primary cilium." Cell Rep **3**(6): 1806-1814.

Cohn, R. D., C. van Erp, et al. (2007). "Angiotensin II type 1 receptor blockade attenuates TGF-beta-induced failure of muscle regeneration in multiple myopathic states." Nature medicine **13**: 204-210.

Collod-Béroud, G. and C. Boileau (2002). "Marfan syndrome in the third Millennium." European journal of human genetics : EJHG **10**: 673-681.

Consortium, E. N. C. O. D. E. P. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**: 57-74.

Consortium, E. P. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.

Consortium, G. (2013). "The Genotype-Tissue Expression (GTEx) project." Nature genetics **45**: 580-585.

Consortium, G. P., A. Auton, et al. (2015). "A global reference for human genetic variation." Nature **526**: 68-74.

Consortium, I. H. (2003). "The International HapMap Project." Nature **426**: 789-796.

Consortium, I. H., D. M. Altshuler, et al. (2010). "Integrating common and rare genetic variation in diverse human populations." Nature **467**: 52-58.

Consortium, W. T. C. C. (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." Nature **447**: 661-678.

Cooper, T., L. M. Napolitano, et al. (1966). "Structural basis of cardiac valvar function." Archives of surgery (Chicago, Ill. : 1960) **93**: 767-771.

Corbit, K. C., A. E. Shyer, et al. (2008). "Kif3a constrains beta-catenin-dependent Wnt signalling through dual ciliary and non-ciliary mechanisms." Nat Cell Biol **10**(1): 70-76.

Cordell, H. J. and D. G. Clayton (2005). "Genetic association studies." Lancet (London, England) **366**: 1121-1131.

Criley, J. M., K. B. Lewis, et al. (1966). "Prolapse of the mitral valve: clinical and cine-angiocardiographic findings." British heart journal **28**: 488-496.

Cripe, L., G. Andelfinger, et al. (2004). "Bicuspid aortic valve is heritable." Journal of the American College of Cardiology **44**(1): 138-143.

Dal-Bianco, J. P., E. Aikawa, et al. (2009). "Active adaptation of the tethered mitral valve: insights into a compensatory mechanism for functional mitral regurgitation." Circulation **120**(4): 334-342.

Dal-Bianco, J. P. and R. A. Levine (2013). "Anatomy of the Mitral Valve Apparatus - Role of 2D and 3D Echocardiography." Cardiology clinics **31**(PMC3856635): 10.1016/j.ccl.2013.1003.1001.

Daly, M. J., J. D. Rioux, et al. (2001). "High-resolution haplotype structure in the human genome." Nature genetics **29**: 229-232.

Dargis, N., M. Lamontagne, et al. (2016). "Identification of gender-specific genetic variants in patients with bicuspid aortic valve." The American journal of cardiology **117**(3): 420-426.

Das, S., G. R. Abecasis, et al. (2018). "Genotype Imputation from Large Reference Panels." Annual review of genomics and human genetics **19**: 73-96.

Das, S., L. Forer, et al. (2016). "Next-generation genotype imputation service and methods." Nature Genetics **48**(10): 1284-1287.

Das, S., L. Forer, et al. (2016). "Next-generation genotype imputation service and methods." Nat Genet **48**(10): 1284-1287.

Davey, J. W. and M. L. Blaxter (2010). "RADSeq: next-generation population genetics." Brief Funct Genomics **9**(5-6): 416-423.

Dawn Teare, M. and J. H. Barrett (2005). "Genetic linkage studies." Lancet (London, England) **366**: 1036-1044.

de Vlaming, A., K. Sauls, et al. (2012). "Atrioventricular valve development: new perspectives on an old theme." Differentiation; research in biological diversity **84**: 103-116.

Degner, J. F., A. A. Pai, et al. (2012). "DNase I sensitivity QTLs are a major determinant of human expression variation." Nature **482**(7385): 390-394.

Del Pasqua, A., G. Rinelli, et al. (2009). "New findings concerning cardiovascular manifestations emerging from long-term follow-up of 150 patients with the Williams-Beuren-Beuren syndrome." Cardiology in the young **19**: 563-567.

Delaneau, O., C. Coulonges, et al. (2008). "Shape-IT: new rapid and accurate algorithm for haplotype inference." BMC bioinformatics **9**: 540.

Delaneau, O., J. Marchini, et al. (2012). "A linear complexity phasing method for thousands of genomes." Nat Methods **9**(2): 179-181.

Delaneau, O., J.-F. Zagury, et al. (2013). "Improved whole-chromosome phasing for disease and population genetic studies." Nature methods **10**: 5-6.

Delling, F. N., J. Rong, et al. (2014). "Familial clustering of mitral valve prolapse in the community." Circulation: CIRCULATIONAHA-114.

Détaint, D., L. Faivre, et al. (2010). "Cardiovascular manifestations in men and women carrying a FBN1 mutation." European heart journal **31**: 2223-2229.

Dietz, H. C., G. R. Cutting, et al. (1991). "Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene." Nature **352**: 337-339.

Dina, C., N. i. Bouatia-Naji, et al. (2015). "Genetic association analyses highlight biological pathways underlying mitral valve prolapse." Nat Genet **47**(10): 1206-1211.

Disse, S., E. Abergel, et al. (1999). "Mapping of a first locus for autosomal dominant myxomatous mitral-valve prolapse to chromosome 16p11. 2-p12. 1." The American Journal of Human Genetics **65**(5): 1242-1251.

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

Durst, R., K. Sauls, et al. (2015). "Mutations in DCHS1 cause mitral valve prolapse." Nature **525**(7567): 109-113.

Edwards, S. L., J. Beesley, et al. (2013). "Beyond GWASs: illuminating the dark road from association to function." Am J Hum Genet **93**(5): 779-797.

Eid, J., A. Fehr, et al. (2009). "Real-time DNA sequencing from single polymerase molecules." Science (New York, N.Y.) **323**: 133-138.

Emilsson, V., G. Thorleifsson, et al. (2008). "Genetics of gene expression and its effect on disease." Nature **452**: 423-428.

Erdmann, J., T. Kessler, et al. (2018). "A decade of genome-wide association studies for coronary artery disease: the challenges ahead." Cardiovascular research **114**: 1241-1257.

Eriksson, J., C. J. Carlhäll, et al. (2010). "Semi-automatic quantification of 4D left ventricular blood flow." Journal of cardiovascular magnetic resonance : official journal of the Society for Cardiovascular Magnetic Resonance **12**: 9.

Evangelou, E. and J. P. A. Ioannidis (2013). "Meta-analysis methods for genome-wide association studies and beyond." Nature reviews. Genetics **14**: 379-389.

Eveborn, G. W., H. Schirmer, et al. (2013). "The evolving epidemiology of valvular aortic stenosis. the Tromsø study." Heart (British Cardiac Society) **99**: 396-400.

Faivre, L., G. Collod-Beroud, et al. (2007). "Effect of mutation type and location on clinical outcome in 1,013 probands with Marfan syndrome or related phenotypes and FBN1 mutations: an international study." American journal of human genetics **81**: 454-466.

Fedak, P. W., M. P. de Sa, et al. (2003). "Vascular matrix remodeling in patients with bicuspid aortic valve malformations: implications for aortic dilatation." J Thorac Cardiovasc Surg **126**(3): 797-806.

Flint, J. (2013). "GWAS GUIDE." Current Biology.

Freed, L. A., J. S. Acierno, et al. (2003). "A locus for autosomal dominant mitral valve prolapse on chromosome 11p15. 4." The American Journal of Human Genetics **72**(6): 1551-1559.

Freed, L. A., D. Levy, et al. (1999). "Prevalence and clinical outcome of mitral-valve prolapse." New England Journal of Medicine **341**(1): 1-7.

Gabriel, S. B., S. F. Schaffner, et al. (2002). "The structure of haplotype blocks in the human genome." Science (New York, N.Y.) **296**: 2225-2229.

Gallagher, M. D. and A. S. Chen-Plotkin (2018). "The Post-GWAS Era: From Association to Function." Am J Hum Genet **102**(5): 717-730.

Gammie, J. S., J. Chikwe, et al. (2018). "Isolated Mitral Valve Surgery: The Society of Thoracic Surgeons Adult Cardiac Surgery Database Analysis." Ann Thorac Surg **106**(3): 716-727.

Garg, V., A. N. Muth, et al. (2005). "Mutations in NOTCH1 cause aortic valve disease." Nature **437**(7056): 270.

Garnier, S., V. Truong, et al. (2013). "Genome-wide haplotype analysis of cis expression quantitative trait loci in monocytes." PLoS Genet **9**(1): e1003240.

Gaudet, P., M. S. Livstone, et al. (2011). "Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium." Brief Bioinform **12**(5): 449-462.

Geirsson, A., M. Singh, et al. (2012). "Modulation of transforming growth factor-$$ signaling and extracellular matrix production in myxomatous mitral valves by angiotensin II receptor blockers." Circulation **126**(11 suppl 1): S189-S197.

Gharibeh, L., H. Komati, et al. (2018). "GATA6 Regulates Aortic Valve Remodeling and its Haploinsufficiency Leads to RL-Type Bicuspid Aortic Valve." Circulation.

Ghosh, N., H. Al-Shehri, et al. (2012). "Characterization of mitral valve prolapse with cardiac computed tomography: comparison to echocardiographic and intraoperative findings." The international journal of cardiovascular imaging **28**: 855-863.

Goeman, J. J. and P. Bühlmann (2007). "Analyzing gene expression data in terms of gene sets: methodological issues." Bioinformatics (Oxford, England) **23**: 980-987.

Goodwin, S., J. D. McPherson, et al. (2016). "Coming of age ten years of next-generation sequencing technologies." Nature Reviews Genetics **17**(6): 333-351.

Grande-Allen, K. J., A. Calabro, et al. (2004). "Glycosaminoglycans and proteoglycans in normal mitral valve leaflets and chordae: association with regions of tensile and compressive loading." Glycobiology **14**: 621-633.

Greene, C. S., A. Krishnan, et al. (2015). "Understanding multicellular function and disease with human tissue-specific networks." Nat Genet **47**(6): 569-576.

Greene, C. S. and O. G. Troyanskaya (2012). "Accurate evaluation and analysis of functional genomics data and methods." Annals of the New York Academy of Sciences **1260**: 95-100.

Guauque-Olarte, S., D. Messika-Zeitoun, et al. (2015). "Calcium signaling pathway genes RUNX2 and CACNA1C are associated with calcific aortic valve disease." Circulation: Cardiovascular Genetics: CIRCGENETICS-115.

Guo, D.-C., H. Pannu, et al. (2007). "Mutations in smooth muscle $$-actin (ACTA2) lead to thoracic aortic aneurysms and dissections." Nature genetics **39**(12): 1488-1493.

Guy, T. S. and A. C. Hill (2012). "Mitral valve prolapse." Annual review of medicine **63**: 277-292.

Halldórsson, B. V., V. Bafna, et al. (2004). "Optimal haplotype block-free selection of tagging SNPs for genome-wide association studies." Genome research **14**: 1633-1640.

Hancock, E. W. and K. Cohn (1966). "The syndrome associated with midsystolic click and late systolic murmur." The American journal of medicine **41**: 183-196.

Hayek, E., C. N. Gring, et al. (2005). "Mitral valve prolapse." The Lancet **365**(9458): 507-518.

Hill, W. G. and A. Robertson (1968). "Linkage disequilibrium in finite populations." TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **38**: 226-231.

Himmelstein, D. S. and S. E. Baranzini (2015). "Heterogeneous Network Edge Prediction: A Data Integration Approach to Prioritize Disease-Associated Genes." PLoS computational biology **11**: e1004259.

Hindorff, L. A., P. Sethupathy, et al. (2009). "Potential etiologic and functional implications of genome-wide association loci for human diseases and traits." Proceedings of the National Academy of Sciences of the United States of America **106**: 9362-9367.

Hirschhorn, J. N. and M. J. Daly (2005). "Genome-wide association studies for common diseases and complex traits." Nature Reviews Genetics **6**(2): 95-108.

Hjortnaes, J., J. Keegan, et al. (2017). Comparative Histopathological Analysis of Mitral Valves in Barlow Disease and Fibroelastic Deficiency. Seminars in Thoracic and Cardiovascular Surgery.

Hodges, E., Z. Xuan, et al. (2007). "Genome-wide in situ exon capture for selective resequencing." Nature genetics **39**: 1522-1527.

Hofker, M. H., J. Fu, et al. (2014). "The genome revolution and its role in understanding complex diseases." Biochimica et biophysica acta **1842**: 1889-1895.

Hofker, M. H., J. Fu, et al. (2014). "The genome revolution and its role in understanding complex diseases." Biochim Biophys Acta **1842**(10): 1889-1895.

Horstkotte, D. and F. Loogen (1988). "The natural history of aortic valve stenosis." Eur Heart J **9 Suppl E**: 57-64.

Howie, B., C. Fuchsberger, et al. (2012). "Fast and accurate genotype imputation in genome-wide association studies through pre-phasing." Nat Genet **44**(8): 955-959.

Howie, B., C. Fuchsberger, et al. (2012). "Fast and accurate genotype imputation in genome-wide association studies through pre-phasing." Nat Genet **44**(8): 955-959.

Howie, B. N., P. Donnelly, et al. (2009). "A flexible and accurate genotype imputation method for the next generation of genome-wide association studies." PLoS genetics **5**: e1000529.

Hu, J. X., C. E. Thomas, et al. (2016). "Network biology concepts in complex disease comorbidities." Nature Reviews Genetics.

Hu, T., N. A. Sinnott-Armstrong, et al. (2011). "Characterizing genetic interactions in human disease association studies using statistical epistasis networks." BMC bioinformatics **12**: 364.

Huang, J., B. Howie, et al. (2015). "Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel." Nat Commun **6**: 8111.

Huang, Y. T., L. Liang, et al. (2015). "iGWAS: Integrative Genome-Wide Association Studies of Genetic and Genomic Data for Disease Susceptibility Using Mediation Analysis." Genet Epidemiol **39**(5): 347-356.

Hulin, A., V. Moore, et al. (2017). "Loss of Axin2 results in impaired heart valve maturation and subsequent myxomatous valve disease." Cardiovasc Res **113**(1): 40-51.

Inai, K., R. A. Norris, et al. (2008). "BMP-2 induces cell migration and periostin expression during atrioventricular valvulogenesis." Developmental biology **315**: 383-396.

Itoh, A., G. Krishnamurthy, et al. (2009). "Active stiffening of mitral valve leaflets in the beating heart." American journal of physiology. Heart and circulatory physiology **296**: H1766-H1773.

Iung, B., G. Baron, et al. (2007). "Valvular heart disease in the community: a European experience." Current problems in cardiology **32**: 609-661.

Iung, B. and A. Vahanian (2011). "Epidemiology of valvular heart disease in the adult." Nature reviews. Cardiology **8**: 162-172.

Jain, M., I. T. Fiddes, et al. (2015). "Improved data analysis for the MinION nanopore sequencer." Nature methods **12**: 351-356.

Jain, R., K. A. Engleka, et al. (2011). "Cardiac neural crest orchestrates remodeling and functional maturation of mouse semilunar valves." J Clin Invest **121**(1): 422-430.

Jia, P., L. Wang, et al. (2011). "Pathway-based analysis of GWAS datasets: effective but caution required." Int J Neuropsychopharmacol **14**(4): 567-572.

Jia, P. and Z. Zhao (2014). "Network.assisted analysis to prioritize GWAS results: principles, methods and perspectives." Human genetics **133**: 125-138.

John, S., N. Shephard, et al. (2004). "Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites." American journal of human genetics **75**: 54-64.

Johnson, G. C., L. Esposito, et al. (2001). "Haplotype tagging for the identification of common disease genes." Nature genetics **29**: 233-237.

Judge, D. P., R. R. Markwald, et al. (2011). "Translational research on the mitral valve: from developmental mechanisms to new therapies." Journal of cardiovascular translational research **4**(6): 699-701.

Kaartinen, V. and D. Warburton (2003). "Fibrillin controls TGF-beta activation." Nature genetics **33**: 331-332.

Kaliman, P., D. Catalucci, et al. (2005). "Myotonic dystrophy protein kinase phosphorylates phospholamban and regulates calcium uptake in cardiomyocyte sarcoplasmic reticulum." J Biol Chem **280**(9): 8016-8021.

Kamstrup, P. R., r.-H. Tybj\ae, Anne, et al. (2014). "Elevated lipoprotein (a) and risk of aortic valve stenosis in the general population." Journal of the American College of Cardiology **63**(5): 470-477.

Kanehisa, M. and S. Goto (2000). "KEGG: kyoto encyclopedia of genes and genomes." Nucleic acids research **28**: 27-30.

Kao, P. Y., K. H. Leung, et al. (2017). "Pathway analysis of complex diseases for GWAS, extending to consider rare variants, multi-omics and interactions." Biochim Biophys Acta Gen Subj **1861**(2): 335-353.

Kao, P. Y. P., K. H. Leung, et al. (2017). "Pathway analysis of complex diseases for GWAS, extending to consider rare variants, multi-omics and interactions." Biochimica et biophysica acta. General subjects **1861**: 335-353.

Kent, W. J., C. W. Sugnet, et al. (2002). "The human genome browser at UCSC." Genome Res **12**(6): 996-1006.

Kerstjens-Frederikse, W. S., I. M. B. H. van de Laar, et al. (2016). "Cardiovascular malformations caused by NOTCH1 mutations do not keep left data on 428 probands with left-sided CHD and their families." Genetics in Medicine **18**(9): 914-923.

Koren, S., M. C. Schatz, et al. (2012). "Hybrid error correction and de novo assembly of single-molecule sequencing reads." Nature biotechnology **30**: 693-700.

Kwon, J. S., J. Kim, et al. (2012). "Performance Comparison of Two Gene Set Analysis Methods for Genome-wide Association Study Results: GSA-SNP vs i-GSEA4GWAS." Genomics Inform **10**(2): 123-127.

Kyndt, F., J.-P. Gueffet, et al. (2007). "Mutations in the gene encoding filamin A as a cause for familial cardiac valvular dystrophy." Circulation **115**(1): 40-49.

Kyndt, F., J.-P. Gueffet, et al. (2007). "Mutations in the gene encoding filamin A as a cause for familial cardiac valvular dystrophy." Circulation **115**(1): 40-49.

Kyndt, F., J.-J. Schott, et al. (1998). "Mapping of X-linked myxomatous valvular dystrophy to chromosome Xq28." The American Journal of Human Genetics **62**(3): 627-632.

Laforest, B., G. Andelfinger, et al. (2011). "Loss of Gata5 in mice leads to bicuspid aortic valve." The Journal of clinical investigation **121**(7): 2876.

Laforest, B. and M. Nemer (2012). "Genetic insights into bicuspid aortic valve formation." Cardiology research and practice **2012**.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**: 860-921.

Le Tourneau, T., S. Le Scouarnec, et al. (2018). "New insights into mitral valve dystrophy: a Filamin-A genotype-phenotype and outcome study." European heart journal **39**: 1269-1277.

Le Tourneau, T., J. Mérot, et al. (2018). "Genetics of syndromic and non-syndromic mitral valve prolapse." Heart (British Cardiac Society) **104**: 978-984.

Leiserson, M. D. M., J. V. Eldridge, et al. (2013). "Network analysis of GWAS data." Current opinion in genetics \& development **23**: 602-610.

Levine, R. A., A. A. Hagége, et al. (2015). "Mitral valve disease--morphology and mechanisms." Nature reviews. Cardiology **12**: 689-710.

Lewis, C. M. (2002). "Genetic association studies: design, analysis and interpretation." Briefings in bioinformatics **3**(2): 146-153.

Lewontin, R. C. (1964). "The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models." Genetics **49**: 49-67.

Li, Y., C. Willer, et al. (2009). "Genotype imputation." Annual review of genomics and human genetics **10**: 387-406.

Li, Y., C. J. Willer, et al. (2010). "MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes." Genetic epidemiology **34**: 816-834.

Lindsay, M. E., D. Schepers, et al. (2012). "Loss-of-function mutations in TGFB2 cause a syndromic presentation of thoracic aortic aneurysm." Nature genetics **44**(8): 922.

Liu, A., B. Wang, et al. (2005). "Mouse intraflagellar transport proteins regulate both the activator and repressor functions of Gli transcription factors." Development **132**(13): 3103-3111.

Liu, A. C., V. R. Joag, et al. (2007). "The emerging role of valve interstitial cell phenotypes in regulating heart valve pathobiology." The American journal of pathology **171**: 1407-1418.

Liu, J. Z., A. F. McRae, et al. (2010). "A versatile gene-based test for genome-wide association studies." Am J Hum Genet **87**(1): 139-145.

Liu, L., Y. Li, et al. (2012). "Comparison of next-generation sequencing systems." Journal of biomedicine \& biotechnology **2012**: 251364.

Loeys, B. L., J. Chen, et al. (2005). "A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2." Nat Genet **37**(3): 275-281.

Loh, P.-R., P. Danecek, et al. (2016). "Reference-based phasing using the Haplotype Reference Consortium panel." Nature genetics **48**: 1443-1448.

Loh, P. R., P. F. Palamara, et al. (2016). "Fast and accurate long-range phasing in a UK Biobank cohort." Nat Genet **48**(7): 811-816.

Lumiaho, A., R. Ikaheimo, et al. (2001). "Mitral valve prolapse and mitral regurgitation are common in patients with polycystic kidney disease type 1." Am J Kidney Dis **38**(6): 1208-1216.

Manolio, T. A. (2013). "Bringing genome-wide association findings into clinical use." Nature reviews. Genetics **14**: 549-558.

Marchini, J., L. R. Cardon, et al. (2004). "The effects of human population structure on large genetic association studies." Nature genetics **36**: 512-517.

Marchini, J. and B. Howie (2010). "Genotype imputation for genome-wide association studies." Nat Rev Genet **11**(7): 499-511.

Marchini, J., B. Howie, et al. (2007). "A new multipoint method for genome-wide association studies by imputation of genotypes." Nat Genet **39**(7): 906-913.

Martin, L. J., V. Ramachandran, et al. (2007). "Evidence in favor of linkage to human chromosomal regions 18q, 5q and 13q for bicuspid aortic valve and associated cardiovascular malformations." Human genetics **121**(2): 275-284.

Maurano, M. T., R. Humbert, et al. (2012). "Systematic localization of common disease-associated variation in regulatory DNA." Science **337**(6099): 1190-1195.

McCarthy, M. I., G. c. Abecasis, ,alo R., et al. (2008). "Genome-wide association studies for complex traits: consensus, uncertainty and challenges." Nat Rev Genet **9**(5): 356-369.

McCarthy, S., S. Das, et al. (2016). "A reference panel of 64,976 haplotypes for genotype imputation." Nature genetics **48**: 1279-1283.

McVean, G. A. T. and N. J. Cardin (2005). "Approximating the coalescent with recombination." Philosophical transactions of the Royal Society of London. Series B, Biological sciences **360**: 1387-1393.

Michelena, H. I., A. D. Khanna, et al. (2011). "Incidence of aortic complications in patients with bicuspid aortic valves." JAMA **306**(10): 1104-1112.

Milgrom-Hoffman, M., Z. Harrelson, et al. (2011). "The heart endocardium is derived from vascular endothelial progenitors." Development (Cambridge, England) **138**: 4777-4787.

Millington-Sanders, C., A. Meir, et al. (1998). "Structure of chordae tendineae in the left ventricle of the human heart." Journal of anatomy **192 (Pt 4)**: 573-581.

Moonesinghe, R., M. J. Khoury, et al. (2008). "Required sample size and nonreplicability thresholds for heterogeneous genetic associations." Proceedings of the National Academy of Sciences of the United States of America **105**: 617-622.

Mooney, M. A. and B. Wilmot (2015). "Gene set analysis: A step-by-step guide." American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics **168**: 517-527.

Morton, N. E. (1955). "Sequential tests for the detection of linkage." Am J Hum Genet **7**(3): 277-318.

Nakajima, Y., T. Yamagishi, et al. (2000). "Mechanisms involved in valvuloseptal endocardial cushion formation in early cardiogenesis: roles of transforming growth factor (TGF)-beta and bone morphogenetic protein (BMP)." The Anatomical record **258**: 119-127.

Nalliah, C. J., R. Mahajan, et al. (2019). "Mitral valve prolapse and sudden cardiac death: a systematic review and meta-analysis." Heart (British Cardiac Society) **105**: 144-151.

Nesta, F., M. Leyne, et al. (2005). "New locus for autosomal dominant mitral valve prolapse on chromosome 13." Circulation **112**(13): 2022-2030.

Ng, C. M., A. Cheng, et al. (2004). "TGF-beta-dependent pathogenesis of mitral valve prolapse in a mouse model of Marfan syndrome." The Journal of clinical investigation **114**: 1586-1592.

Novaro, G. M., R. Sachar, et al. (2003). "Association between apolipoprotein E alleles and calcific valvular heart disease." Circulation **108**(15): 1804-1808.

O'dushlaine, C., E. Kenny, et al. (2009). "The SNP ratio test: pathway analysis of genome-wide association datasets." Bioinformatics **25**(20): 2762-2763.

Ortlepp, J. R., R. Hoffmann, et al. (2001). "The vitamin D receptor genotype predisposes to the development of calcific aortic valve stenosis." Heart **85**(6): 635-638.

Ott, J., J. Wang, et al. (2015). "Genetic linkage analysis in the age of whole-genome sequencing." Nat Rev Genet **16**(5): 275-284.

Padang, R., R. D. Bagnall, et al. (2012). "Rare non-synonymous variations in the transcriptional activation domains of GATA5 in bicuspid aortic valve disease." Journal of molecular and cellular cardiology **53**(2): 277-281.

Padang, R., R. D. Bagnall, et al. (2012). "Genetic basis of familial valvular heart disease." Circulation: Cardiovascular Genetics **5**(5): 569-580.

Pan, D., L. D. Estevez-Salmeron, et al. (2005). "The integral inner nuclear membrane protein MAN1 physically interacts with the R-Smad proteins to repress signaling by the transforming growth factor-{beta} superfamily of cytokines." J Biol Chem **280**(16): 15992-16001.

Park, P. J. (2009). "ChIP-seq: advantages and challenges of a maturing technology." Nature reviews. Genetics **10**: 669-680.

Pasaniuc, B. and A. L. Price (2016). "Dissecting the genetics of complex traits using summary association statistics." Nature Reviews Genetics.

Patterson, N., A. L. Price, et al. (2006). "Population structure and eigenanalysis." PLoS genet **2**(12): e190.

Pavalko, F. M. and S. M. LaRoche (1993). "Activation of human neutrophils induces an interaction between the integrin beta 2-subunit (CD18) and the actin binding protein alpha-actinin." J Immunol **151**(7): 3795-3807.

Pe'er, I., R. Yelensky, et al. (2008). "Estimation of the multiple testing burden for genomewide association studies of nearly all common variants." Genetic epidemiology **32**: 381-385.

Pers, T. H., J. M. Karjalainen, et al. (2015). "Biological interpretation of genome-wide association studies using predicted gene functions." Nat Commun **6**: 5890.

Pomerance, A. (1967). "Ageing changes in human heart valves." British heart journal **29**: 222-231.

Pop, M. and S. L. Salzberg (2008). "Bioinformatics challenges of new sequencing technology." Trends in genetics : TIG **24**: 142-149.

Prakash, S. K., Boss\'e, Yohan, et al. (2014). "A roadmap to investigate the genetic basis of bicuspid aortic valve and its complications: insights from the International BAVCon (Bicuspid Aortic Valve Consortium)." Journal of the American College of Cardiology **64**(8): 832-839.

Probst, V., S. Le Scouarnec, et al. (2006). "Familial aggregation of calcific aortic valve stenosis in the western part of France." Circulation **113**(6): 856-860.

Prunier, F., G. Terrien, et al. (2013). "Pseudoxanthoma elasticum: cardiac findings in patients and Abcc6-deficient mouse model." PloS one **8**: e68700.

Pucéat, M. (2013). "Embryological origin of the endocardium and derived valve progenitor cells: from developmental biology to stem cell-based valve repair." Biochimica et biophysica acta **1833**: 917-922.

Purcell, S., B. Neale, et al. (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." Am J Hum Genet **81**(3): 559-575.

Pyeritz, R. E. and M. A. Wappel (1983). "Mitral valve dysfunction in the Marfan syndrome. Clinical and echocardiographic study of prevalence and natural history." Am J Med **74**(5): 797-807.

Qu, R., M. M. Silver, et al. (1998). "Distribution of endoglin in early human development reveals high levels on endocardial cushion tissue mesenchyme during valve formation." Cell Tissue Res **292**(2): 333-343.

Qu, X.-K., X.-B. Qiu, et al. (2014). "A novel NKX2. 5 loss-of-function mutation associated with congenital bicuspid aortic valve." The American journal of cardiology **114**(12): 1891-1895.

Rajamannan, N. M., B. Gersh, et al. (2003). "Calcific aortic stenosis: from bench to the bedside--emerging clinical and cellular concepts." Heart **89**(7): 801-805.

Ramanan, V. K., L. Shen, et al. (2012). "Pathway analysis of genomic data: concepts, methods, and prospects for future development." Trends Genet **28**(7): 323-332.

Rath, N., Z. Wang, et al. (2005). "LMCD1/Dyxin is a novel transcriptional cofactor that restricts GATA6 function by inhibiting DNA binding." Molecular and cellular biology **25**(20): 8864-8873.

Rizzo, J. M. and M. J. Buck (2012). "Key principles and clinical applications of "next-generation" DNA sequencing." Cancer prevention research (Philadelphia, Pa.) **5**: 887-900.

Ro, R., D. Halpern, et al. (2014). "Vector flow mapping in obstructive hypertrophic cardiomyopathy to assess the relationship of early systolic left ventricular flow and the mitral valve." Journal of the American College of Cardiology **64**: 1984-1995.

Roberts, W. C. and J. M. Ko (2005). "Frequency by decades of unicuspid, bicuspid, and tricuspid aortic valves in adults having isolated aortic valve replacement for aortic stenosis, with or without associated aortic regurgitation." Circulation **111**: 920-925.

Rosenberg, N. A., L. Huang, et al. (2010). "Genome-wide association studies in diverse populations." Nature reviews. Genetics **11**: 356-366.

Sacks, M. S. and A. P. Yoganathan (2007). "Heart valve function: a biomechanical perspective." Philosophical transactions of the Royal Society of London. Series B, Biological sciences **362**: 1369-1391.

Sauls, K., A. De Vlaming, et al. (2012). "Developmental basis for filamin-A-associated myxomatous mitral valve disease." Cardiovascular research **96**(1): 109-119.

Schaid, D. J., W. Chen, et al. (2018). "From genome-wide associations to candidate causal variants by statistical fine-mapping." Nature reviews. Genetics.

Schaub, M. A., A. P. Boyle, et al. (2012). "Linking disease associations with regulatory information in the human genome." Genome Res **22**(9): 1748-1759.

Scheet, P. and M. Stephens (2006). "A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase." American journal of human genetics **78**: 629-644.

Schunkert, H. and N. J. Samani (2018). "Into the great wide open: 10 years of genome-wide association studies." Cardiovascular research **114**: 1189-1191.

Sexton, T., F. Bantignies, et al. (2009). "Genomic interactions: chromatin loops and gene meeting points in transcriptional regulation." Semin Cell Dev Biol **20**(7): 849-855.

Shannon, P., A. Markiel, et al. (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome research **13**: 2498-2504.

Shi, L. M., J. W. Tao, et al. (2014). "GATA5 loss-of-function mutations associated with congenital bicuspid aortic valve." Int J Mol Med **33**(5): 1219-1226.

Slatkin, M. (2008). "Linkage disequilibrium--understanding the evolutionary past and mapping the medical future." Nature reviews. Genetics **9**: 477-485.

Smith, T., S. Gurudevan, et al. (2012). "Assessment of the morphological features of degenerative mitral valve disease using 64-slice multi detector computed tomography." Journal of cardiovascular computed tomography **6**: 415-421.

Sonnenblick, E. H., L. M. Napolitano, et al. (1967). "An intrinsic neuromuscular basis for mitral valve motion in the dog." Circulation research **21**(1): 9-15.

Spadaccio, C., P. Mozetic, et al. (2016). "Cells and extracellular matrix interplay in cardiac valve disease: because age matters." Basic research in cardiology **111**: 16.

Sriram, C. S., F. F. Syed, et al. (2013). "Malignant bileaflet mitral valve prolapse syndrome in patients with otherwise idiopathic out-of-hospital cardiac arrest." Journal of the American College of Cardiology **62**: 222-230.

Stephens, M. and P. Scheet (2005). "Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation." American journal of human genetics **76**: 449-462.

Stephens, M., N. J. Smith, et al. (2001). "A new statistical method for haplotype reconstruction from population data." American journal of human genetics **68**: 978-989.

Stewart, B. F., D. Siscovick, et al. (1997). "Clinical factors associated with calcific aortic valve disease." Journal of the American College of Cardiology **29**(3): 630-634.

Subramanian, A., H. Kuehn, et al. (2007). "GSEA-P: a desktop application for Gene Set Enrichment Analysis." Bioinformatics (Oxford, England) **23**: 3251-3253.

Svishcheva, G. R., N. M. Belonogova, et al. (2019). "Gene-based association tests using GWAS summary statistics." Bioinformatics.

Szklarczyk, D., A. Franceschini, et al. (2011). "The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored." Nucleic acids research **39**: D561-D568.

Taliun, D., D. N. Harris, et al. (2019). "Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program." BioRxiv: 563866.

Tan, H. L., E. Glen, et al. (2012). "Nonsynonymous variants in the SMAD6 gene predispose to congenital cardiovascular malformation." Human mutation **33**(4): 720-727.

Teare, M. D. and M. F. Santibañez Koref (2014). "Linkage analysis and the study of Mendelian disease in the era of whole exome and genome sequencing." Briefings in functional genomics **13**: 378-383.

Thacoor, A. (2017). "Mitral valve prolapse and Marfan syndrome." Congenital Heart Disease.

Theal, M., K. Sleik, et al. (2004). "Prevalence of mitral valve prolapse in ethnic groups." The Canadian journal of cardiology **20**: 511-515.

Thiery, J. P., H. Acloque, et al. (2009). "Epithelial-mesenchymal transitions in development and disease." Cell **139**: 871-890.

Tomkinson, A. E., S. Vijayakumar, et al. (2006). "DNA ligases: structure, reaction mechanism, and function." Chemical reviews **106**: 687-699.

Valeria, B., G. Maddalena, et al. (2008). "Endoglin (CD105) expression in the human heart throughout gestation: an immunohistochemical study." Reprod Sci **15**(10): 1018-1026.

van de Laar, I. M. B. H., R. A. Oldenburg, et al. (2011). "Mutations in SMAD3 cause a syndromic form of aortic aneurysms and dissections with early-onset osteoarthritis." Nature genetics **43**: 121-126.

van de Laar, I. M. B. H., D. van der Linde, et al. (2012). "Phenotypic spectrum of the SMAD3-related aneurysms-osteoarthritis syndrome." Journal of medical genetics **49**: 47-57.

van der Sijde, M. R., A. Ng, et al. (2014). "Systems genetics: From GWAS to disease pathways." Biochim Biophys Acta **1842**(10): 1903-1909.

Vastrik, I., P. D'Eustachio, et al. (2007). "Reactome: a knowledge base of biologic pathways and processes." Genome biology **8**: R39.

Visscher, P. M., N. R. Wray, et al. (2017). "10 Years of GWAS Discovery: Biology, Function, and Translation." American journal of human genetics **101**: 5-22.

Wang, K., M. Li, et al. (2010). "Analysing biological pathways in genome-wide association studies." Nat Rev Genet **11**(12): 843-854.

Wang, K., M. Li, et al. (2010). "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data." Nucleic Acids Res **38**(16): e164.

Wang, L., P. Jia, et al. (2011). "Gene set analysis of genome-wide association studies: methodological issues and perspectives." Genomics **98**(1): 1-8.

Wang, L., T. Matsushita, et al. (2015). "PINBPA: cytoscape app for network analysis of GWAS data." Bioinformatics **31**(2): 262-264.

Wang, Z., M. Gerstein, et al. (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nature reviews. Genetics **10**: 57-63.

Watanabe, K., E. Taskesen, et al. (2017). "Functional mapping and annotation of genetic associations with FUMA." Nature communications **8**(1): 1826.

Weir, B. S., W. G. Hill, et al. (2004). "Allelic association patterns for a dense SNP map." Genetic epidemiology **27**: 442-450.

Weiss, A. N., J. W. Mimbs, et al. (1975). "Echocardiographic detection of mitral valve prolapse. Exclusion of false positive diagnosis and determination of inheritance." Circulation **52**: 1091-1096.

Wigginton, J. E., D. J. Cutler, et al. (2005). "A note on exact tests of Hardy-Weinberg equilibrium." The American Journal of Human Genetics **76**(5): 887-893.

Wit, A. L., J. J. Fenoglio, et al. (1979). "Ultrastructure and transmembrane potentials of cardiac muscle in the human anterior mitral valve leaflet." Circulation **59**: 1284-1292.

Wooten, E. C., L. K. Iyer, et al. (2010). "Application of gene network analysis techniques identifies AXIN1/PDIA2 and endoglin haplotypes associated with bicuspid aortic valve." PLoS One **5**(1): e8830.

Wu, B., Y. Wang, et al. (2011). "Nfatc1 coordinates valve endocardial cell lineage development required for heart valve formation." Circulation research **109**: 183-192.

Yang, B., W. Zhou, et al. (2017). "Protein-altering and regulatory genetic variants near GATA4 implicated in bicuspid aortic valve." Nature communications **8**: 15481.

Yang, X., X. Meng, et al. (2009). "Bone morphogenic protein 2 induces Runx2 and osteopontin expression in human aortic valve interstitial cells: role of Smad1 and extracellular signal-regulated kinase 1/2." J Thorac Cardiovasc Surg **138**(4): 1008-1015.

Zerbino, D. R., P. Achuthan, et al. (2018). "Ensembl 2018." Nucleic Acids Res **46**(D1): D754-D761.

Zhang, K., S. Chang, et al. (2015). "I-GSEA4GWAS v2: a web server for functional analysis of SNPs in trait-associated pathways identified from genome-wide association study." Protein cell **6**(3): 221.

Zhang, K., S. Cui, et al. (2010). "i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study." Nucleic Acids Res **38**(Web Server issue): W90-W95.

Appendix A

GWAS-driven Gene-set Analyses, Genetic and Functional Follow-up Suggest GLIS1 as a Susceptibility Gene for Mitral Valve Prolapse

The supplementary tables of chapter III are presented in this part. The tables include all the enriched pathways/enriched tissue or cell types from the pathway-based analyses, iGSEA4GWAS and DEPICT. We also show the enriched gene sets where GLIS1 was found to be the best ranked gene by i-GSEA4GWAS analysis. All the 111 annotated SNPs in high LD with the most associated SNP in GLIS1 locus are shown also.

Supplementary Table 1. Enriched gene sets identified by i-GSEA4GWAS. We display all 244 confidently enriched pathways /gene sets (FDR < 0.05). We also added the Hedgehog Signaling Pathway result. The number of significantly contributing genes and the total number of genes analyzed are indicated. A short gene set description was provided when available, in addition to the catalog IDs (GO and KEGG). FDR: false discovery rate.

Gene Set Name	Significant genes	Total genes analyzed	P Value	FDR	Gene Set Description
Biocarta: EDG1 Pathway	26	27	0.001	1.00E-03	Phospholipids as signalling intermediaries
GO: Actin Binding	58	76	0.001	1.00E-03	GO:0003779. Interacting selectively with monomeric or multimeric forms of actin, including actin filaments.
GO: Adherens Junction	18	23	0.001	1.00E-03	GO:0005912. A cell junction at which the cytoplasmic face of the plasma membrane is attached to actin filaments.
GO: Anion Transport	24	31	0.001	1.00E-03	GO:0006820. The directed movement of anions, atoms or small molecules with a net negative charge, into, out of, within or between cells.
GO: Basolateral Plasma Membrane	28	35	0.001	1.00E-03	GO:0016323. Part of the plasma membrane that includes the basal end and sides of the cell. Often used in reference to animal polarized epithelial membranes, where the basal membrane is the part attached to the extracellular matrix, or in plant cells, where the basal membrane is defined with respect to the zygotic axis.
GO: Transcription Repressor Activity	94	148	0.001	1.00E-03	GO:0016564. Any transcription regulator activity that prevents or downregulates transcription.
GO: Vasculature Development	41	55	0.001	1.00E-03	GO:0001944. The process whose specific outcome is the progression of the vasculature over time, from its formation to the mature structure.

GO: Cytoskeletal Protein Binding	112	159	0.001	1.10E-03	GO:0008092. Interacting selectively with any protein component of any cytoskeleton (actin, microtubule, or intermediate filament cytoskeleton).
KEGG: Intestinal Immune Network for IGA Production	31	45	0.001	1.11E-03	hsa04672
GO: Ion Transport	131	180	0.001	1.13E-03	GO:0006811. The directed movement of charged atoms or small charged molecules into, out of, within or between cells.
Biocarta: Integrin Pathway	30	38	0.001	1.18E-03	Integrin Signaling Pathway
KEGG: Dilated Cardiomyopathy	68	90	0.001	1.31E-03	hsa05414
GO: Structural Constituent of Cytoskeleton	39	57	0.001	1.33E-03	GO:0005200. The action of a molecule that contributes to the structural integrity of a cytoskeletal structure.
GO: Actin Filament Binding	21	25	0.001	1.35E-03	GO:0051015. Interacting selectively with an actin filament, also known as F-actin, a helical filamentous polymer of globular G-actin subunits.
GO: Sulfur Metabolic Process	31	37	0.001	1.36E-03	GO:0006790. The chemical reactions and pathways involving the nonmetallic element sulfur or compounds that contain sulfur, such as the amino acids methionine and cysteine or the tripeptide glutathione.
GO: Angiogenesis	36	48	0.001	1.37E-03	GO:0001525. Blood vessel formation when new vessels emerge from the proliferation of pre-existing blood vessels.
GO: Cell Cortex	30	39	0.001	1.38E-03	GO:0005938. The region of a cell that lies just beneath the plasma membrane and often, but not always, contains a network of actin filaments and associated proteins.
KEGG: Cell Adhesion Molecules Cams	91	127	0.001	1.38E-03	hsa04514
GO: Positive Regulation of Cell Proliferation	97	145	0.001	1.39E-03	GO:0008284. Any process that activates or increases the rate or extent of cell proliferation.
GO: Anion Transmembrane Transporter Activity	43	59	0.001	1.39E-03	GO:0008509. Catalysis of the transfer of a negatively charged ion from one side of a membrane to the other.
GO: Anatomical Structure formation	42	56	0.001	1.40E-03	GO:0048646. The process pertaining to the initial formation of an anatomical structure from unspecified parts. This process begins with the specific processes that contribute to the appearance of the discrete structure and ends when the structural rudiment is recognizable. An anatomical structure is any biological entity that occupies space and is distinguished from its surroundings. Anatomical structures can be macroscopic such as a carpel, or microscopic such as an acrosome.
Biocarta: SPPA Pathway	20	22	0.001	1.40E-03	Aspirin Blocks Signaling Pathway Involved in Platelet Activation
KEGG: Focal Adhesion	145	198	0.001	1.41E-03	hsa04510
GO: Transcription Corepressor Activity	58	90	0.001	1.42E-03	GO:0003714. The function of a transcription cofactor that represses transcription from a RNA polymerase II promoter; does not bind DNA itself.

GO: DI TRI Valent Inorganic Cation Transmembrane Transporter Activity	18	22	0.001	1.43E-03	GO:0015082. Catalysis of the transfer of inorganic cations with a valency of two or three from one side of the membrane to the other. Inorganic cations are atoms or small molecules with a positive charge that do not contain carbon in covalent linkage.
GO: Positive Regulation of Transcription DNA Dependent	82	118	0.001	1.44E-03	GO:0045893. Any process that activates or increases the frequency, rate or extent of DNA-dependent transcription.
GO: Substrate Specific Channel Activity	108	154	0.001	1.45E-03	GO:0022838. Catalysis of energy-independent facilitated diffusion, mediated by passage of a specific solute through a transmembrane aqueous pore or channel. Stereospecificity is not exhibited but this transport may be specific for a particular molecular species or class of molecules.
GO: Ion Channel Activity	104	147	0.001	1.46E-03	GO:0005216. Catalysis of facilitated diffusion of an ion (by an energy-independent process) by passage through a transmembrane aqueous pore or channel without evidence for a carrier-mediated mechanism.
KEGG: Pentose Phosphate Pathway	20	25	0.001	1.55E-03	hsa00030
GO: Cell Cortex Part	19	24	0.001	1.55E-03	GO:0044448. Any constituent part of the cell cortex, the region of a cell that lies just beneath the plasma membrane and often, but not always, contains a network of actin filaments and associated proteins.
GO: Transcription Activator Activity	112	170	0.001	1.56E-03	GO:0016563. Any transcription regulator activity required for initiation or upregulation of transcription.
GO: Transmembrane Receptor Protein Kinase Activity	41	51	0.001	1.57E-03	GO:0019199.
GO: Positive Regulation of RNA Metabolic Process	83	120	0.001	1.57E-03	GO:0051254. Any process that activates or increases the frequency, rate or extent of the chemical reactions and pathways involving RNA.
GO: Metal Ion Transmembrane Transporter Activity	105	145	0.001	1.59E-03	GO:0046873. Catalysis of the transfer of metal ions from one side of a membrane to the other.
GO: mRNA Metabolic Process	43	82	0.001	1.59E-03	GO:0016071. The chemical reactions and pathways involving mRNA, messenger RNA, which is responsible for carrying the coded genetic 'message', transcribed from DNA, to sites of protein assembly at the ribosomes.
GO: Cytoskeleton Dependent Intracellular Transport	20	26	0.001	1.62E-03	GO:0030705. The directed movement of substances along cytoskeletal elements such as microfilaments or microtubules within a cell.
GO: Sulfotransferase Activity	23	28	0.001	1.64E-03	GO:0008146. Catalysis of the transfer of a sulfate group from 3'-phosphoadenosine 5'-phosphosulfate to the hydroxyl group of an acceptor, producing the sulfated derivative and 3'-phosphoadenosine 5'-phosphate.
GO: mRNA Processing GO 0006397	37	72	0.001	1.68E-03	GO:0006397. Any process involved in the conversion of a primary mRNA transcript into one or more mature mRNA(s) prior to translation into polypeptide.
GO: Enzyme Linked Receptor Protein Signaling Pathway	108	140	0.001	1.74E-03	GO:0007167. Any series of molecular signals initiated by the binding of an extracellular ligand to a receptor on the surface of the target cell, where the receptor possesses catalytic activity or is closely associated with an enzyme such as a protein kinase.
GO: Hemopoietic or Lymphoid Organ Development	53	75	0.001	1.93E-03	GO:0048534. The process whose specific outcome is the progression of any organ involved in hemopoiesis or lymphoid cell activation over time, from its formation to the mature structure. Such development includes

					differentiation of resident cell types (stromal cells) and of migratory cell types dependent on the unique microenvironment afforded by the organ for their proper differentiation.
GO: Lyase Activity	46	68	0.001	1.95E-03	GO:0016829. Catalysis of the cleavage of C-C, C-O, C-N and other bonds by other means than by hydrolysis or oxidation, or conversely adding a group to a double bond. They differ from other enzymes in that two substrates are involved in one reaction direction, but only one in the other direction. When acting on the single substrate, a molecule is eliminated and this generates either a new double bond or a new ring.
KEGG: Hypertrophic Cardiomyopathy HCM	60	83	0.001	1.98E-03	hsa05410
GO: Transmembrane Receptor Protein Tyrosine Kinase Activity	36	43	0.001	1.98E-03	GO:0004714. Catalysis of the reaction: ATP + a protein-L-tyrosine = ADP + a protein-L-tyrosine phosphate, to initiate a change in cell activity.
GO: Striated Muscle Development	26	40	0.001	2.00E-03	GO:0014706. The process whose specific outcome is the progression of a striated muscle over time, from its formation to the mature structure. Striated muscle contain fibers that are divided by transverse bands into striations, and cardiac and skeletal muscle are types of striated muscle. Skeletal muscle myoblasts fuse to form myotubes and eventually multinucleated muscle fibers. The fusion of cardiac cells is very rare and can only form binucleate cells.
GO: Cation Transport	102	143	0.001	2.00E-03	GO:0006812. The directed movement of cations, atoms or small molecules with a net positive charge, into, out of, within or between cells.
GO: Regulation of G Protein Coupled Receptor Protein Signaling Pathway	19	22	0.001	2.02E-03	GO:0008277. Any process that modulates the frequency, rate or extent of G-protein coupled receptor protein signaling pathway activity.
KEGG: Oxidative Phosphorylation	62	113	0.001	2.15E-03	hsa00190
GO: Regulation of Neurotransmitter Levels	17	24	0.001	2.19E-03	GO:0001505. Any process that modulates levels of neurotransmitter.
GO: Endosome Transport	17	23	0.001	2.61E-03	GO:0016197. The directed movement of substances into, out of or mediated by an endosome, a membrane-bound organelle that carries materials newly ingested by endocytosis. It passes many of the materials to lysosomes for degradation.
GO: Organ Morphogenesis	99	144	0.001	2.64E-03	GO:0009887. Morphogenesis of an organ. An organ is defined as a tissue or set of tissues that work together to perform a specific function or functions. Morphogenesis is the process by which anatomical structures are generated and organized. Organs are commonly observed as visibly distinct structures, but may also exist as loosely associated clusters of cells that work together to perform a specific function or functions.
GO: Regulation of Angiogenesis	20	26	0.001	2.75E-03	GO:0045765. Any process that modulates the frequency, rate or extent of angiogenesis.
GO: Deoxyribonuclease Activity	14	22	0.001	2.76E-03	GO:0004536. Catalysis of the hydrolysis of ester linkages within deoxyribonucleic acid.
GO: Immune System Development	55	79	0.001	2.77E-03	GO:0002520. The process whose specific outcome is the progression of an organismal system whose objective is to provide calibrated responses by an organism to a potential internal or invasive threat, over time, from its formation to the mature structure. A system is a regularly interacting or interdependent group of organs or tissues that work together to carry out a given biological process.

GO: DNA Catabolic Process	18	23	0.001	2.78E-03	GO:0006308. The chemical reactions and pathways resulting in the breakdown of DNA, deoxyribonucleic acid, one of the two main types of nucleic acid, consisting of a long unbranched macromolecule formed from one or two strands of linked deoxyribonucleotides, the 3'-phosphate group of each constituent deoxyribonucleotide being joined in 3',5'-phosphodiester linkage to the 5'-hydroxyl group of the deoxyribose moiety of the next one.
GO: Hemopoiesis	51	73	0.001	2.80E-03	GO:0030097. The process whose specific outcome is the progression of the myeloid and lymphoid derived organ/tissue systems of the blood and other parts of the body over time, from formation to the mature structure. The site of hemopoiesis is variable during development, but occurs primarily in bone marrow or kidney in many adult vertebrates.
GO: Structural Constituent of Muscle	22	33	0.002	2.84E-03	GO:0008307. The action of a molecule that contributes to the structural integrity of a muscle fiber.
GO: Detection of Stimulus	35	46	0.001	2.84E-03	GO:0051606. The series of events in which a stimulus is received by a cell and converted into a molecular signal.
GO: Actin Cytoskeleton	84	128	0.001	2.88E-03	GO:0015629. The part of the cytoskeleton (the internal framework of a cell) composed of actin and associated proteins. Includes actin cytoskeleton-associated complexes.
KEGG: Tight Junction	85	127	0.001	3.00E-03	hsa04530
Biocarta: MPR Pathway	25	34	0.001	3.02E-03	How Progesterone Initiates Oocyte Membrane
GO: Positive Regulation of Transcription	97	143	0.001	3.24E-03	GO:0045941. Any process that activates or increases the frequency, rate or extent of transcription.
GO: Di-, Tri-valent Inorganic Cation Transport	26	32	0.001	3.25E-03	GO:0015674. The directed movement of inorganic cations with a valency of two or three into, out of, within or between cells. Inorganic cations are atoms or small molecules with a positive charge which do not contain carbon in covalent linkage.
KEGG: Cardiac Muscle Contraction	48	72	0.001	3.33E-03	hsa04260
GO: RNA Processing	88	166	0.001	3.45E-03	GO:0006396. Any process involved in the conversion of one or more primary RNA transcripts into one or more mature RNA molecules.
GO: Sensory Perception	123	185	0.001	3.80E-03	GO:0007600. The series of events required for an organism to receive a sensory stimulus, convert it to a molecular signal, and recognize and characterize the signal.
Biocarta: METPathway	29	37	0.001	3.81E-03	Signaling of Hepatocyte Growth Factor Receptor
GO: Phosphoric Ester Hydrolase Activity	103	149	0.001	3.85E-03	GO:0042578. Catalysis of the reaction: RPO-R' + H2O = RPOOH + R'H. This reaction is the hydrolysis of any phosphoric ester bond, any ester formed from orthophosphoric acid, O=P(OH)3.
GO: Collagen	21	23	0.001	3.85E-03	GO:0005581. Any of the various assemblies in which collagen chains form a left-handed triple helix; may assemble into higher order structures.
GO: Calcium Ion Transport	22	27	0.001	3.88E-03	GO:0006816. The directed movement of calcium (Ca) ions into, out of, within or between cells.
Biocarta: PDINS Pathway	17	23	0.002	3.89E-03	Phosphoinositides and their downstream targets.

GO: Positive Regulation of Nucleobasenucleosidenucleotide and Nucleic Acid Metabolic Process	103	153	0.001	4.08E-03	GO:0045935. Any process that activates or increases the frequency, rate or extent of the chemical reactions and pathways involving nucleobases, nucleosides, nucleotides and nucleic acids.
GO: Microtubule Based Process	58	81	0.001	4.10E-03	GO:0007017. Any cellular process that depends upon or alters the microtubule cytoskeleton, that part of the cytoskeleton comprising microtubules and their associated proteins.
GO: Phosphoric Monoester Hydrolase Activity	76	108	0.001	4.53E-03	GO:0016791. Catalysis of the hydrolysis of phosphoric monoesters, releasing inorganic phosphate.
GO: Actin Filament Based Process	71	113	0.002	4.66E-03	GO:0030029. Any cellular process that depends upon or alters the actin cytoskeleton, that part of the cytoskeleton comprising actin filaments and their associated proteins.
GO: Regulation of Cell Differentiation	44	60	0.001	4.68E-03	GO:0045595. Any process that modulates the frequency, rate or extent of cell differentiation, the process whereby relatively unspecialized cells acquire specialized structural and functional features.
GO: Regulation of Cytokine Production	18	24	0.001	4.82E-03	GO:0001817. Any process that modulates the frequency, rate, or extent of production of a cytokine.
GO: Cell Migration	69	94	0.001	4.83E-03	GO:0016477. The orderly movement of cells from one site to another, often during the development of a multicellular organism.
GO: Kinase Inhibitor Activity	16	25	0.001	4.84E-03	GO:0019210. Stops, prevents or reduces the activity of a kinase, an enzyme which catalyzes of the transfer of a phosphate group, usually from ATP, to a substrate molecule.
GO: Steroid Metabolic Process	48	71	0.001	4.86E-03	GO:0008202. The chemical reactions and pathways involving steroids, compounds with a 1,2,cyclopentanoperhydrophenanthrene nucleus.
Biocarta: Mcalpain Pathway	19	25	0.001	4.90E-03	mCalpain and friends in Cell motility
GO: Inorganic Cation Transmembrane Transporter Activity	39	55	0.001	4.90E-03	GO:0022890. Catalysis of the transfer of inorganic cations from one side of a membrane to the other. Inorganic cations are atoms or small molecules with a positive charge that do not contain carbon in covalent linkage.
GO: DNA Damage Response signal Transduction	23	33	0.001	4.91E-03	GO:0042770. A cascade of processes induced by the detection of DNA damage within a cell.
KEGG: Leukocyte Transendothelial Migration	73	113	0.002	5.13E-03	hsa04670
KEGG: Glycine Serine and Threonine Metabolism	18	31	0.001	5.14E-03	hsa00260
GO: Transferase Activity Transferring Sulfur Containing Groups	25	32	0.002	5.26E-03	GO:0016782. Catalysis of the transfer of a sulfur-containing group from one compound (donor) to another (acceptor).
GO: Mitochondrial Part	69	139	0.003	5.41E-03	GO:0044429. Any constituent part of a mitochondrion, a semiautonomous, self replicating organelle that occurs in varying numbers, shapes, and sizes in the cytoplasm of virtually all eukaryotic cells. It is notably the

					site of tissue respiration.
Biocarta: ECM Pathway	19	24	0.001	5.44E-03	Erk and PI-3 Kinase Are Necessary for Collagen Binding in Corneal Epithelia
KEGG: Phosphatidylinositol Signaling System	50	76	0.001	5.58E-03	hsa04070
GO: Lipid Transporter Activity	21	27	0.001	5.60E-03	GO:0005319. Enables the directed movement of lipids into, out of, within or between cells.
GO: Sequence Specific DNA Binding	41	58	0.001	5.63E-03	GO:0043565. Interacting selectively with DNA of a specific nucleotide composition, e.g. GC-rich DNA binding, or with a specific sequence motif or type of DNA e.g. promotor binding or rDNA binding.
GO: Heart Development	26	37	0.001	5.66E-03	GO:0007507. The process whose specific outcome is the progression of the heart over time, from its formation to the mature structure. The heart is a hollow, muscular organ, which, by contracting rhythmically, keeps up the circulation of the blood.
KEGG: Glycerophospholipid Metabolism	50	76	0.002	5.66E-03	hsa00564
GO: Secondary Metabolic Process	16	26	0.001	5.67E-03	GO:0019748. The chemical reactions and pathways resulting in many of the chemical changes of compounds that are not necessarily required for growth and maintenance of cells, and are often unique to a taxon. In multicellular organisms secondary metabolism is generally carried out in specific cell types, and may be useful for the organism as a whole. In unicellular organisms, secondary metabolism is often used for the production of antibiotics or for the utilization and acquisition of unusual nutrients.
GO: Positive Regulation of Cell Differentiation	18	25	0.001	5.71E-03	GO:0045597. Any process that activates or increases the frequency, rate or extent of cell differentiation.
GO: Alcohol Metabolic Process	56	87	0.001	5.95E-03	GO:0006066. The chemical reactions and pathways involving alcohols, any of a class of alkyl compounds containing a hydroxyl group.
GO: Response to UV	17	25	0.001	6.00E-03	GO:0009411. A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of an ultraviolet radiation (UV light) stimulus. Ultraviolet radiation is electromagnetic radiation with a wavelength in the range of 10 to 380 nanometers.
GO: Gated Channel Activity	87	121	0.001	6.01E-03	GO:0022836. Catalysis of the transmembrane transfer of a solute by a channel that opens in response to a specific stimulus.
KEGG: Adherens Junction	56	72	0.001	6.04E-03	hsa04520
KEGG: Mtor Signaling Pathway	36	52	0.001	6.08E-03	hsa04150
GO: Metal Ion Transport	81	115	0.001	6.10E-03	GO:0030001. The directed movement of metal ions, any metal ion with an electric charge, into, out of, within or between cells.
Biocarta: Insulin Pathway	15	22	0.001	6.15E-03	Insulin Signaling Pathway
GO: Calcium Channel Activity	26	33	0.001	6.19E-03	GO:0005262. Catalysis of facilitated diffusion of an calcium (by an energy-independent process) involving passage through a transmembrane aqueous pore or channel without evidence for a carrier-mediated mechanism.

GO: Fatty Acid Metabolic Process	41	63	0.001	6.27E-03	GO:0006631. The chemical reactions and pathways involving fatty acids, aliphatic monocarboxylic acids liberated from naturally occurring fats and oils by hydrolysis.
GO: G Protein Coupled Receptor Activity	131	184	0.001	6.30E-03	GO:0004930. A receptor that binds an extracellular ligand and transmits the signal to a heterotrimeric G-protein complex. These receptors are characteristically seven-transmembrane receptors and are made up of hetero- or homodimers.
KEGG: TGF-Beta Signaling Pathway	58	84	0.001	6.38E-03	hsa04350
GO: Protein C Terminus Binding	48	73	0.001	6.40E-03	GO:0008022. Interacting selectively with a protein C-terminus, the end of any peptide chain at which the 1-carboxy function of a constituent amino acid is not attached in peptide linkage to another amino-acid residue.
GO: Transmembrane Receptor Protein Tyrosine Kinase Signaling Pathway	65	83	0.001	7.21E-03	GO:0007169. The series of molecular signals generated as a consequence of a transmembrane receptor tyrosine kinase binding to its physiological ligand.
Biocarta: NOS1 Pathway	15	21	0.001	7.24E-03	Nitric Oxide Signaling Pathway
GO: Cell Cycle Checkpoint	27	47	0.001	7.37E-03	GO:0000075. A point in the eukaryotic cell cycle where progress through the cycle can be halted until conditions are suitable for the cell to proceed to the next stage.
Biocarta: Inflam Pathway	17	29	0.002	7.44E-03	Cytokines and Inflammatory Response
GO: G Protein Coupled Receptor Binding	31	54	0.004	7.50E-03	GO:0001664. Interacting selectively with a G-protein-coupled receptor.
KEGG: Glycerolipid Metabolism	36	49	0.001	8.01E-03	hsa00561
GO: Activation of Protein Kinase Activity	19	26	0.002	8.04E-03	GO:0032147. Any process that initiates the activity of an inactive protein kinase.
GO: Glucose Metabolic Process	17	28	0.002	8.28E-03	GO:0006006. The chemical reactions and pathways involving glucose, the aldohexose gluco-hexose. D-glucose is dextrorotatory and is sometimes known as dextrose; it is an important source of energy for living organisms and is found free as well as combined in homo- and hetero-oligosaccharides and polysaccharides.
KEGG: Glycosaminoglycan Biosynthesis Heparan Sulfate	21	26	0.001	8.33E-03	hsa00534
Biocarta: CSKPathway	17	22	0.002	8.38E-03	Activation of Csk by cAMP-dependent Protein Kinase Inhibits Signaling through the T Cell Receptor
KEGG: Ether Lipid Metabolism	22	33	0.001	8.47E-03	hsa00565
GO: Regulation of Binding	39	57	0.002	8.53E-03	GO:0051098. Any process that modulates the frequency, rate or extent of binding, the selective interaction of a molecule with one or more specific sites on another molecule.
GO: Protein Kinase Inhibitor Activity	15	24	0.001	8.87E-03	GO:0004860. Stops, prevents or reduces the activity of a protein kinase, an enzyme which phosphorylates a protein.
KEGG: Valine Leucine and	30	44	0.001	8.98E-03	hsa00280

Isoleucine Degradation					
GO: Muscle Development	55	93	0.001	9.01E-03	GO:0007517. The process whose specific outcome is the progression of the muscle over time, from its formation to the mature structure. The muscle is an organ consisting of a tissue made up of various elongated cells that are specialized to contract and thus to produce movement and mechanical work.
GO: Glycosaminoglycan Binding	24	33	0.002	9.19E-03	GO:0005539. Interacting selectively with any glycan (polysaccharide) containing a substantial proportion of aminomonosaccharide residues.
GO: Positive Regulation of Transcription From RNA Polymerase ii Promoter	43	65	0.001	9.24E-03	GO:0045944. Any process that activates or increases the frequency, rate or extent of transcription from the RNA polymerase II promoter.
GO: Envelope	91	164	0.004	9.24E-03	GO:0031975. A multilayered structure surrounding all or part of a cell; encompasses one or more lipid bilayers, and may include a cell wall layer, also includes the space between layers.
GO: Organelle Envelope	91	164	0.004	9.24E-03	GO:0031967. A double membrane structure enclosing an organelle, including two lipid bilayers and the region between them. In some cases, an organelle envelope may have more than two membranes.
GO: Protein Serine Threonine Phosphatase Activity	15	22	0.002	9.25E-03	GO:0004722. Catalysis of the reaction: protein serine/threonine phosphate + H2O = protein serine/threonine + phosphate.
KEGG: Insulin Signaling Pathway	84	135	0.001	9.26E-03	hsa04910
Biocarta: PPATA Pathway	41	57	0.002	9.28E-03	Mechanism of Gene Regulation by Peroxisome Proliferators via PPARa(alpha)
GO: Negative Regulation of Cell Differentiation	22	28	0.002	9.56E-03	GO:0045596. Any process that stops, prevents or reduces the frequency, rate or extent of cell differentiation.
Biocarta: NO1 Pathway	22	30	0.002	9.81E-03	Actions of Nitric Oxide in the Heart
GO: DNA Integrity Checkpoint	14	23	0.001	9.92E-03	GO:0031570. Any cell cycle checkpoint that delays or arrests cell cycle progression in response to changes in DNA structure.
GO: Response to Hormone Stimulus	22	32	0.003	1.05E-02	GO:0009725. A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a hormone stimulus.
GO: Membrane Organization and Biogenesis	90	133	0.003	1.10E-02	GO:0016044. A process that is carried out at the cellular level which results in the formation, arrangement of constituent parts, or disassembly of membranes inside and surrounding the cell.
GO: Negative Regulation of Transcription	115	184	0.002	1.14E-02	GO:0016481. Any process that stops, prevents or reduces the frequency, rate or extent of transcription.
GO: Contractile Fiber	15	25	0.005	1.14E-02	GO:0043292. Fibers, composed of actin, myosin, and associated proteins, found in cells of smooth or striated muscle.
GO: Polysaccharide Binding	25	35	0.001	1.15E-02	GO:0030247. Interacting selectively with any polysaccharide.
GO: Actin Filament Organization	13	23	0.003	1.25E-02	GO:0007015. Control of the spatial distribution of actin filaments; includes organizing filaments into meshworks, bundles, or other structures, as by cross-linking.

GO: Cellular Protein Complex Assembly	22	32	0.004	1.26E-02	GO:0043623. The aggregation, arrangement and bonding together of a set of components to form a protein complex, occurring at the level of an individual cell.
GO: Actin Cytoskeleton Organization and Biogenesis	64	103	0.004	1.26E-02	GO:0030036. A process that is carried out at the cellular level which results in the formation, arrangement of constituent parts, or disassembly of cytoskeletal structures comprising actin filaments and their associated proteins.
GO: Transforming Growth Factor Beta Receptor Signaling Pathway	27	36	0.002	1.27E-02	GO:0007179. The series of molecular signals generated as a consequence of a transforming growth factor beta receptor binding to one of its physiological ligands.
GO: Response to Abiotic Stimulus	52	83	0.004	1.27E-02	GO:0009628. A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of an abiotic (non-living) stimulus.
GO: Regulation of Heart Contraction	18	25	0.002	1.27E-02	GO:0008016. Any process that modulates the frequency, rate or extent of heart contraction. Heart contraction is the process by which the heart decreases in volume in a characteristic way to propel blood through the body.
GO: Transcription Coactivator Activity	75	120	0.003	1.27E-02	GO:0003713. The function of a transcription cofactor that activates transcription from a RNA polymerase II promoter; does not bind DNA itself.
GO: Regulation of Cell Migration	22	28	0.003	1.28E-02	GO:0030334. Any process that modulates the frequency, rate or extent of cell migration.
GO: Kinase Regulator Activity	29	45	0.002	1.28E-02	GO:0019207. Modulates the activity of a kinase, an enzyme which catalyzes of the transfer of a phosphate group, usually from ATP, to a substrate molecule.
GO: Response to Light Stimulus	28	44	0.005	1.28E-02	GO:0009416. A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a light stimulus, electromagnetic radiation of wavelengths classified as infrared, visible or ultraviolet light.
GO: Small Protein Conjugating Enzyme Activity	36	52	0.002	1.29E-02	GO:0008639. Catalysis of the covalent attachment of small proteins, such as ubiquitin or ubiquitin-like proteins, to lysine residues on a target protein. This function may be performed alone or in conjunction with an E3, ubiquitin-like protein ligase.
GO: Voltage Gated Channel Activity	52	73	0.002	1.39E-02	GO:0022832. Catalysis of the transmembrane transfer of a solute by a channel whose open state is dependent on the voltage across the membrane in which it is embedded.
GO: Mitosis	48	80	0.004	1.39E-02	GO:0007067. Progression through mitosis, the division of the eukaryotic cell nucleus to produce two daughter nuclei that, usually, contain the identical chromosome complement to their mother.
GO: Electron Carrier Activity	49	78	0.001	1.40E-02	GO:0009055. Any molecular entity that serves as an electron acceptor and electron donor in an electron transport system.
KEGG: Purine Metabolism	95	154	0.004	1.40E-02	hsa00230
KEGG: Fructose and Mannose Metabolism	21	33	0.006	1.47E-02	hsa00051
GO: Rhodopsin Like Receptor Activity	90	128	0.002	1.47E-02	GO:0001584. A G-protein coupled receptor that is structurally/functionally related to the rhodopsin receptor.
GO: Contractile Fiber Part	14	23	0.006	1.52E-02	GO:0044449. Any constituent part of a contractile fiber, a fiber composed of actin, myosin, and associated

					proteins, found in cells of smooth or striated muscle.
GO: Carbohydrate Biosynthetic Process	39	49	0.002	1.53E-02	GO:0016051. The chemical reactions and pathways resulting in the formation of carbohydrates, any of a group of organic compounds based of the general formula Cx(H2O)y.
GO: Cation Channel Activity	83	118	0.002	1.55E-02	GO:0005261. Catalysis of the energy-independent passage of cations across a lipid bilayer down a concentration gradient.
GO: Phosphoprotein Phosphatase Activity	56	79	0.002	1.59E-02	GO:0004721. Catalysis of the reaction: a phosphoprotein + H2O = a protein + phosphate. Together with protein kinases, these enzymes control the state of phosphorylation of cell proteins and thereby provide an important mechanism for regulating cellular activity.
KEGG: Arrhythmogenic Right Ventricular Cardiomyopathy Arvc	55	74	0.001	1.76E-02	hsa05412
GO: Nuclease Activity	30	55	0.004	1.77E-02	GO:0004518. Catalysis of the hydrolysis of ester linkages within nucleic acids.
KEGG: Vascular Smooth Muscle Contraction	76	115	0.003	1.78E-02	hsa04270
GO: Regulation of Organelle Organization and Biogenesis	27	39	0.003	1.83E-02	GO:0033043. Any process that modulates the frequency, rate or extent of the processes involved in the formation, arrangement of constituent parts, or disassembly of an organelle.
Biocarta: Ceramide Pathway	17	22	0.005	1.85E-02	Ceramide Signaling Pathway
GO: Mitochondrial Envelope	46	94	0.005	1.86E-02	GO:0005740. The double lipid bilayer enclosing the mitochondrion and separating its contents from the cell cytoplasm; includes the intermembrane space.
GO: Mitochondrial Lumen	24	46	0.006	1.89E-02	GO:0031980. The volume enclosed by the mitochondrial inner membrane.
GO: Mitochondrial Matrix	24	46	0.006	1.89E-02	GO:0005759. The gel-like material, with considerable fine structure, that lies in the matrix space, or lumen, of a mitochondrion. It contains the enzymes of the tricarboxylic acid cycle and, in some organisms, the enzymes concerned with fatty-acid oxidation.
KEGG: Sphingolipid Metabolism	27	40	0.008	2.02E-02	hsa00600
GO: Excretion	23	36	0.006	2.06E-02	GO:0007588. The elimination by an organism of the waste products that arise as a result of metabolic activity. These products include water, carbon dioxide (CO2), and nitrogenous compounds.
GO: Positive Regulation of Multicellular Organismal Process	43	64	0.003	2.06E-02	GO:0051240. Any process that activates or increases the frequency, rate or extent of an organismal process, the processes pertinent to the function of an organism above the cellular level; includes the integrated processes of tissues and organs.
GO: Regulation of Mitosis	25	40	0.006	2.08E-02	GO:0007088. Any process that modulates the frequency, rate or extent of mitosis.
KEGG: Progesterone Mediated Oocyte Maturation	53	84	0.004	2.11E-02	hsa04914
KEGG: Long Term Depression	52	70	0.002	2.21E-02	hsa04730
GO: Monovalent Inorganic Cation	62	90	0.004	2.21E-02	GO:0015672. The directed movement of inorganic cations with a valency of one into, out of, within or between

Transport					cells. Inorganic cations are atoms or small molecules with a positive charge which do not contain carbon in covalent linkage.
KEGG: Dorso Ventral Axis formation	20	24	0.003	2.22E-02	hsa04320
GO: Growth	50	76	0.004	2.22E-02	GO:0040007. The increase in size or mass of an entire organism, a part of an organism or a cell.
GO: Oxidoreductase Activity Acting on CH-OH Group of Donors	36	62	0.007	2.34E-02	GO:0016614. Catalysis of an oxidation-reduction (redox) reaction in which a CH-OH group act as a hydrogen or electron donor and reduces a hydrogen or electron acceptor.
GO: Regulation of Cell Growth	28	45	0.005	2.35E-02	GO:0001558. Any process that modulates the frequency, rate or extent of cell growth.
GO: Hydro Lyase Activity	17	27	0.006	2.35E-02	GO:0016836. Catalysis of the cleavage of a carbon-oxygen bond by elimination of water.
GO: Lipase Activity	33	50	0.003	2.35E-02	GO:0016298. Catalysis of the hydrolysis of a lipid or phospholipid.
KEGG: Neurotrophin Signaling Pathway	81	126	0.005	2.36E-02	hsa04722
GO: Transmembrane Receptor Protein Serine Threonine Kinase Signaling Pathway	34	47	0.007	2.36E-02	GO:0007178. The series of molecular signals generated as a consequence of a transmembrane receptor serine/threonine kinase binding to its physiological ligand.
Biocarta: Gleevec Pathway	17	23	0.006	2.37E-02	Inhibition of Cellular Proliferation by Gleevec
KEGG: Notch Signaling Pathway	32	47	0.008	2.37E-02	hsa04330
GO: Dephosphorylation	47	69	0.004	2.38E-02	GO:0016311. The process of removing one or more phosphoric (ester or anhydride) residues from a molecule.
Biocarta: NFAT Pathway	35	53	0.005	2.39E-02	NFAT and Hypertrophy of the heart (Transcription in the broken heart)
Biocarta: IGF1 Pathway	15	21	0.004	2.49E-02	IGF-1 Signaling Pathway
Biocarta: TPO Pathway	18	24	0.004	2.49E-02	TPO Signaling Pathway
Biocarta: Chrebp2 Pathway	28	41	0.005	2.50E-02	Regulation And Function Of ChREBP in Liver
Biocarta: BAD Pathway	20	25	0.005	2.55E-02	Regulation of BAD phosphorylation
GO: Positive Regulation of Transport	13	22	0.005	2.56E-02	GO:0051050. Any process that activates or increases the frequency, rate or extent of the directed movement of substances (such as macromolecules, small molecules, ions) into, out of, within or between cells.
GO: Positive Regulation of Protein Modification Process	21	28	0.005	2.56E-02	GO:0031401. Any process that activates or increases the frequency, rate or extent of the covalent alteration of one or more amino acid residues within a protein.
GO: Skeletal Muscle Development	19	31	0.003	2.57E-02	GO:0007519. The developmental sequence of events leading to the formation of adult muscle that occurs in the anima. In vertebrate skeletal muscle the main events are: the fusion of myoblasts to form myotubes that increase in size by further fusion to them of myoblasts, the formation of myofibrils within their cytoplasm and the establishment of functional neuromuscular junctions with motor neurons. At this stage they can be regarded as mature muscle fibers.

GO: Establishment of Protein Localization	108	185	0.007	2.71E-02	GO:0045184. The directed movement of a protein to a specific location.
GO: Transcription Factor Complex	50	88	0.009	2.73E-02	GO:0005667. Any complex, distinct from RNA polymerase, including one or more polypeptides capable of binding DNA at promoters or at cis-acting regulatory sequences, and regulating transcription.
KEGG: Long Term Potentiation	52	69	0.004	2.76E-02	hsa04720
GO: One Carbon Compound Metabolic Process	19	26	0.007	2.79E-02	GO:0006730. The chemical reactions and pathways involving compounds containing a single carbon atom.
GO: Oxidoreductase Activity GO 0016616	31	56	0.01	2.86E-02	GO:0016616. Catalysis of an oxidation-reduction (redox) reaction in which a CH-OH group acts as a hydrogen or electron donor and reduces NAD+ or NADP.
Biocarta: HER2 Pathway	19	22	0.006	2.87E-02	Role of ERBB2 in Signal Transduction and Oncology
GO: Protein Kinase Binding	37	61	0.008	2.93E-02	GO:0019901. Interacting selectively with a protein kinase, any enzyme that catalyzes the transfer of a phosphate group, usually from ATP, to a protein substrate.
Biocarta: DC Pathway	13	22	0.009	2.94E-02	Dendritic cells in regulating TH1 and TH2 Development
GO: Response to Endogenous Stimulus	113	193	0.006	2.95E-02	GO:0009719. A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of an endogenous stimulus.
GO: Regulation of Transport	40	66	0.005	2.98E-02	GO:0051049. Any process that modulates the frequency, rate or extent of the directed movement of substances (such as macromolecules, small molecules, ions) into, out of, within or between cells.
Biocarta: VEGF Pathway	21	28	0.003	2.99E-02	VEGF, Hypoxia, and Angiogenesis
GO: Oxidoreductase Activity Acting on The CH-CH Group of Donors	15	23	0.008	2.99E-02	GO:0016627. Catalysis of an oxidation-reduction (redox) reaction in which a CH-CH group acts as a hydrogen or electron donor and reduces a hydrogen or electron acceptor.
Biocarta: CTCF Pathway	15	23	0.004	3.00E-02	CTCF: First Multivalent Nuclear Factor
Biocarta: CCR3 Pathway	20	23	0.005	3.02E-02	CCR3 signaling in Eosinophils
GO: Negative Regulation of Developmental Process	120	194	0.004	3.05E-02	GO:0051093. Any process that stops, prevents or reduces the rate or extent of development, the biological process whose specific outcome is the progression of an organism over time from an initial condition (e.g. a zygote, or a young adult) to a later condition (e.g. a multicellular animal or an aged adult).
KEGG: Gnrh Signaling Pathway	66	101	0.009	3.18E-02	hsa04912
GO: Muscle Cell Differentiation	16	22	0.003	3.27E-02	GO:0042692. The process whereby a relatively unspecialized cell acquires specialized features of a muscle cell.
GO: Regulation of DNA Metabolic Process	26	45	0.007	3.28E-02	GO:0051052. Any process that modulates the frequency, rate or extent of the chemical reactions and pathways involving DNA.
GO: Receptor Complex	35	54	0.007	3.28E-02	GO:0043235. Any protein group composed of two or more subunits, which may or may not be identical, which undergoes combination with a hormone, neurotransmitter, drug or intracellular messenger to initiate a change in cell function.

GO: Digestion	28	41	0.009	3.58E-02	GO:0007586. The whole of the physical, chemical, and biochemical processes carried out by multicellular organisms to break down ingested nutrients into components that may be easily absorbed and directed into metabolism.
GO: M Phase	65	112	0.01	3.59E-02	GO:0000279. Progression through M phase, the part of the cell cycle comprising nuclear division.
GO: Carbon Oxygen Lyase Activity	18	31	0.011	3.59E-02	GO:0016835. Catalysis of the breakage of a carbon-oxygen bond.
GO: Phosphatase Regulator Activity	17	25	0.006	3.66E-02	GO:0019208. Modulates the activity of a phosphatase, an enzyme which catalyzes of the removal of a phosphate group from a substrate molecule.
KEGG: Axon Guidance	91	127	0.006	3.79E-02	hsa04360
Biocarta: TCR Pathway	31	44	0.008	3.88E-02	T Cell Receptor Signaling Pathway
GO: Regulation of Nucleocytoplasmic Transport	15	22	0.008	3.90E-02	GO:0046822. Any process that modulates the frequency, rate or extent of the directed movement of substances between the nucleus and the cytoplasm.
GO: Intracellular Protein Transport	83	142	0.01	3.95E-02	GO:0006886. The directed movement of proteins in a cell, including the movement of proteins between specific compartments or structures within a cell, such as organelles of a eukaryotic cell.
GO: Membrane Fusion	20	28	0.009	4.00E-02	GO:0006944. The joining of two lipid bilayers to form a single membrane.
GO: Chemokine Receptor Binding	26	43	0.017	4.00E-02	GO:0042379. Interacting selectively with any chemokine receptor.
Biocarta: GPCR Pathway	25	34	0.008	4.07E-02	Signaling Pathway from G-Protein Families
GO: Vesicle	71	121	0.008	4.08E-02	GO:0031982. Any small, fluid-filled, spherical organelle enclosed by membrane or protein.
GO: Protein Tyrosine Kinase Activity	49	63	0.01	4.19E-02	GO:0004713. Catalysis of the reaction: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
GO: Protein Targeting	64	108	0.015	4.24E-02	GO:0006605. The process of targeting specific proteins to particular membrane-bound subcellular organelles. Usually requires an organelle specific protein sequence motif.
Biocarta: PDGF Pathway	23	32	0.01	4.30E-02	PDGF Signaling Pathway
Biocarta: Il2rb Pathway	23	38	0.011	4.42E-02	IL-2 Receptor Beta Chain in T cell Activation
Biocarta: IGF1r Pathway	18	23	0.011	4.50E-02	Multiple antiapoptotic pathways from IGF-1R signaling lead to BAD phosphorylation
KEGG: Taste Transduction	28	48	0.008	4.56E-02	hsa04742
GO: Gtpase Regulator Activity	78	123	0.012	4.56E-02	GO:0030695. Modulates the rate of GTP hydrolysis by a GTPase.
GO: Nuclear Organization and Biogenesis	19	29	0.01	4.65E-02	GO:0006997. A process that is carried out at the cellular level which results in the formation, arrangement of constituent parts, or disassembly of the nucleus.
GO: Nuclear Membrane	33	49	0.016	4.67E-02	GO:0031965. Either of the lipid bilayers that surround the nucleus and form the nuclear envelope; excludes the intermembrane space.
GO: Cell Junction	51	80	0.01	4.67E-02	GO:0030054. A specialized region of connection between two cells or between a cell and the extracellular

					matrix.
GO: Epidermis Development	41	71	0.014	4.74E-02	GO:0008544. The process whose specific outcome is the progression of the epidermis over time, from its formation to the mature structure. The epidermis is the outer epithelial layer of a plant or animal, it may be a single layer that produces an extracellular material (e.g. the cuticle of arthropods) or a complex stratified squamous epithelium, as in the case of many vertebrate species.
GO: Cell Cycle Phase	95	168	0.014	4.74E-02	GO:0022403. A cell cycle process comprising the steps by which a cell progresses through one of the biochemical and morphological phases and events that occur during successive cell replication or nuclear replication events.
GO: Chemokine Activity	25	42	0.015	4.75E-02	GO:0008009. The function of a family of chemotactic pro-inflammatory activation-inducible cytokines acting primarily upon hemopoietic cells in immunoregulatory processes; all chemokines possess a number of conserved cysteine residues involved in intramolecular disulfide bond formation.
GO: Generation of A Signal Involved In Cell Cell Signaling	19	29	0.014	4.75E-02	GO:0003001. The cellular process by which a physical entity or change in state, a signal, is created that originates in one cell and is used to transfer information to another cell. This process begins with the initial formation of the signal and ends with the mature form and placement of the signal.
GO: Regulation of Growth	34	57	0.01	4.92E-02	GO:0040008. Any process that modulates the frequency, rate or extent of the growth of all or part of an organism so that it occurs at its proper speed, either globally or in a specific part of the organism's development.
GO: Regulation of DNA Binding	29	46	0.017	4.93E-02	GO:0051101. Any process that modulates the frequency, rate or extent of DNA binding, selective interaction with deoxyribonucleic acid.
KEGG: Melanogenesis	66	100	0.013	4.93E-02	hsa04916
GO: Protein Complex Binding	37	54	0.011	4.96E-02	GO:0032403. Interacting selectively with any protein complex (a complex of two or more proteins that may include other nonprotein molecules).
KEGG: T Cell Receptor Signaling Pathway	67	107	0.014	4.96E-02	hsa04660
GO: Neuron Differentiation	57	75	0.011	4.98E-02	GO:0030182. The process whereby a relatively unspecialized cell acquires specialized features of a neuron.
GO: Regulation of Protein Modification Process	30	43	0.017	5.00E-02	GO:0031399. Any process that modulates the frequency, rate or extent of the covalent alteration of one or more amino acid residues within a protein.
GO: Protein Tyrosine Phosphatase Activity	38	53	0.013	5.00E-02	GO:0004725. Catalysis of the reaction: protein tyrosine phosphate + H2O = protein tyrosine + phosphate.

Supplementary Table 2: Enriched reconstituted gene sets identified by DEPICT. DEPICT was used to assess whether 50 genes located in associated regions (GWAS P-value $< 1 \times 10^{-5}$) were enriched for any of the reconstituted gene sets defined by DEPICT. We display 36 reconstituted clusters with at least one gene set that reaches suggestive enrichment (P-value < 0.01). For each cluster we provide all gene sets that contribute to this cluster and were nominally enriched (P-value<0.05). Associated regions were constructed by mapping genes to a given region if the genes resided within, or overlapped with linkage

disequilibrium (r2 > 0.5) flanks of a given lead SNP. Ex: Cluster ID and cluster center was defined by Affinity Propagation method which was used to cluster gene sets that were highly correlated and automatically defines independent clusters based on calculated Pearson's correlation matrix (See methods for dertails). The central gene set of a given cluster is the most significant gene set in that cluster (bold). We provide the list of genes near associated loci allocated to each gene set and their respective likelihood to be part of each gene set. We only provide genes with significant likelihood |z-score| > 1.96 corresponding to P-value <0.05. NA: not available, is indicated for gene sets for which no gene presented a significant z-score.

Cluster ID	Gene set ID	Gene set Name	Central gene set of the cluster	Most significant gene set	P value	Genes at associated loci contributing to gene sets
1	MP:0010300	Increased skin tumor incidence	Intestinal adenocarcinoma	Increased skin tumor incidence	0.0001	XPC (3.7), SRR (3.1), SLC7A11 (3.0), BPNT1 (2.9), TMEM40 (2.3), AC068353.34 (2.3), TACR3 (-2.1)
1	MP:0002957	Intestinal adenocarcinoma			0.001	FOXL1 (2.5), SLFN11 (2.5), AC068353.34 (2.2), UFL1 (-2.6)
1	ENSG00000138180	CEP55 PPI subnetwork			0.02	AL450226.1 (2.2)
1	MP:0002404	Increased intestinal adenoma incidence			0.02	SRR (3.2), CAND2 (2.8), XPC (2.7), RASSF3 (2.2), AC068353.34 (2.1), FOXL1 (2.0), LINC00973 (2.0), AL450226.1 (2.0), AKR1C3 (-2.4)
2	GO:0030684	Preribosome	FLG2 PPI subnetwork	Preribosome	0.0004	TSR1 (3.6), SETD4 (3.2), ERCC4 (2.2), SRR (-2.3)
2	ENSG00000065268	WDR18 PPI subnetwork			0.004	SMG6 (3.1), TSR1 (2.9), SLFN11 (2.4), RUNX1 (-2.3)
2	ENSG00000143520	FLG2 PPI subnetwork			0.03	TSR1 (3.0), SMG6 (3.0)
2	ENSG00000136021	SCYL2 PPI subnetwork			0.04	LINC02064 (2.6), RP1L1 (2.2), PEAK1 (-2.4)
3	MP:0001689	Incomplete somite formation	Vasculature development	Incomplete somite formation	0.0004	PTPRM (3.3), LINC00312 (-2.6)
3	GO:0045766	Positive regulation of angiogenesis			0.0006	LMCD1 (2.3), LINC00312 (2.2)
3	MP:0001614	Abnormal blood vessel morphology			0.0009	LMCD1 (2.5), PEAK1 (2.5), PTPRM (2.4), FHL5 (2.0)
3	MP:0004787	Abnormal dorsal aorta morphology			0.007	PTPRM (2.9), RASSF3 (2.8), LMCD1 (2.0), SEZ6L (-2.3)
3	ENSG00000087303	NID2 PPI subnetwork			0.009	PEAK1 (2.7), PTPRM (2.3), LMCD1 (2.1)
3	MP:0006126	Abnormal outflow tract development			0.01	LINC01013 (2.0), ERCC4 (-2.1)

3	GO:0001570	Vasculogenesis			0.01	FOXL1 (4.0), PTPRM (2.9), RASSF3 (2.9), UBTD1 (2.4), FHL5 (2.4), LINC01648 (2.0), AC004944.1 (2.0)
3	MP:0008803	Abnormal placental labyrinth vasculature morphology			0.01	RASSF3 (2.2), FOXL1 (2.1), PTPRM (2.0), LINC00299 (2.0)
3	MP:0004251	Failure of heart looping			0.02	SMG6 (2.8), PBX1 (2.7), RASSF3 (2.5), CAND2 (2.3), PIEZO1 (2.3), PEAK1 (2.1), RAB3GAP2 (2.1), FHL5 (-2.0), LINC00312 (-2.7)
3	ENSG00000142798	HSPG2 PPI subnetwork			0.02	PTPRM (2.8), PEAK1 (2.6), LMCD1 (2.5), MN1 (2.5), TACR3 (2.2)
3	MP:0010392	Prolonged qrs complex duration			0.02	PEAK1 (2.6)
3	MP:0000272	Abnormal aorta morphology			0.02	LMCD1 (3.9), LINC01013 (2.6), PEAK1 (2.5), LINC00312 (2.5), LINC00299 (2.2), SETD4 (2.1), HMG20A (2.0), UBTD1 (-2.1), TMEM40 (-2.4)
3	ENSG00000109072	SEBOX PPI subnetwork			0.02	LMCD1 (2.6), FOXL1 (2.3), LINC00973 (2.2), TACR3 (2.0)
3	MP:0005592	Abnormal vascular smooth muscle morphology			0.02	LMCD1 (4.1), RASSF3 (3.9), FHL5 (3.4), PTPRM (3.2), UBTD1 (2.0), FOXL1 (2.0)
3	MP:0010856	Dilated respiratory conducting tubes			0.03	RUNX1 (2.3), FOXL1 (2.3), PEAK1 (2.0)
3	GO:0045785	Positive regulation of cell adhesion			0.03	AC004944.1 (3.1), PEAK1 (3.0), LMCD1 (2.6), LINC01648 (2.5), AL450226.1 (2.3), LINC00973 (2.0), CAND2 (-2.0)
3	ENSG00000128052	KDR PPI subnetwork			0.03	PTPRM (3.1), RAB3GAP2 (2.9), FHL5 (2.6), UBTD1 (2.3), LINC00973 (2.1)
3	ENSG00000182492	BGN PPI subnetwork			0.03	LMCD1 (2.8), AL450226.1 (2.4)
3	GO:0048514	Blood vessel morphogenesis			0.03	LMCD1 (3.6), UBTD1 (3.2), PTPRM (3.0), LINC00312 (2.8), PEAK1 (2.8), FHL5 (2.7)
3	MP:0003071	Decreased vascular permeability			0.04	PTPRM (3.6), FHL5 (3.4), LINC01013 (2.8), SLFN11 (2.2)
3	GO:0001525	Angiogenesis			0.04	LMCD1 (3.5), UBTD1 (3.2), PEAK1 (2.8), FHL5 (2.7), LINC00312 (2.6), PTPRM (2.4), AC004944.1 (-2.1)
3	KEGG DORSO VENTRAL AXIS	Kegg dorso ventral axis formation			0.04	PEAK1 (3.3), FOXL1 (3.1), PTPRM (2.4),

	FORMATION					MN1 (2.3), RASSF3 (2.1), FHL5 (2.0)
3	GO:0044319	Wound healing, spreading of cells			0.04	LINC02064 (2.4), LINC01013 (2.3), UBTD1 (2.1)
3	GO:0001944	Vasculature development			0.04	LMCD1 (3.6), PEAK1 (3.2), UBTD1 (2.9), PTPRM (2.8), LINC00312 (2.6), FHL5 (2.3), FOXL1 (2.0)
3	GO:0001945	Lymph vessel development			0.04	LINC01921 (3.1), FHL5 (3.0), UBTD1 (2.5), FOXL1 (2.5), AC004944.1 (2.2)
3	GO:0032403	Protein complex binding			0.04	RUNX1 (2.9), PTPRM (2.8)
3	GO:0001568	Blood vessel development			0.04	LMCD1 (3.5), PEAK1 (3.2), PTPRM (2.9), UBTD1 (2.9), LINC00312 (2.8), FHL5 (2.2), FOXL1 (2.0)
3	MP:0004921	Decreased placenta weight			0.04	PIEZO1 (3.0), TMEM40 (2.4), TRIB1 (2.1)
4	GO:0048661	Positive regulation of smooth muscle cell proliferation	Regulation of smooth muscle cell proliferation	Positive regulation of smooth muscle cell proliferation	0.0006	LINC00312 (3.7), FHL5 (3.6), LINC01648 (2.6), LMCD1 (2.2), PTPRM (2.1), ANKRD20A11P (2.1), TRIB1 (2.0)
4	GO:0033002	Muscle cell proliferation			0.004	FHL5 (3.1), LMCD1 (2.8), LINC00312 (2.7), FOXL1 (2.6), PTPRM (2.3), AC068353.34 (2.0), ANKRD20A11P (2.0), CBR1 (-2.2)
4	GO:0048660	Regulation of smooth muscle cell proliferation			0.01	FHL5 (3.8), LINC00312 (3.5), PTPRM (2.4), LMCD1 (2.3), ANKRD20A11P (2.2)
4	GO:0048659	Smooth muscle cell proliferation			0.01	FHL5 (4.0), LINC00312 (3.1), ANKRD20A11P (2.3), PTPRM (2.3), LMCD1 (2.0)
4	GO:0045787	Positive regulation of cell cycle			0.02	HDGFL1 (2.6), LINC01921 (2.4), SLC7A11 (2.2)
4	ENSG00000071539	TRIP13 PPI subnetwork			0.02	AKR1C3 (3.3), LINC00312 (2.0), SIPA1L1 (-2.0)
5	MP:0002747	Abnormal aortic valve morphology	Abnormal semilunar valve morphology	Abnormal aortic valve morphology	0.0006	LINC00299 (2.2), SLFN11 (2.1), TRIB1 (2.1), LINC00312 (2.1), AC068353.34 (2.0), RUNX1 (2.0), CBR1 (-2.1)
5	MP:0002746	Abnormal semilunar valve morphology			0.007	LMCD1 (3.5), PTPRM (3.0), RUNX1 (2.6), LINC00312 (2.3), GLIS1 (2.0)
5	MP:0005269	Abnormal occipital bone morphology			0.02	GLIS1 (2.9), SIPA1L1 (2.8), FOXL1 (2.0), RAB3GAP2 (-2.2)

5	MP:0002754	Dilated heart right ventricle			0.02	LMCD1 (5.9), LINC00312 (4.1), FHL5 (2.5), PTPRM (2.4), PEAK1 (2.0)
6	MP:0008840	Abnormal spike wave discharge	Voltage-gated calcium channel activity	Abnormal spike wave discharge	0.0006	ANKRD20A11P (2.6), SEZ6L (2.3)
6	MP:0003990	Decreased neurotransmitter release			0.005	NRXN3 (3.0), ABLIM2 (2.7), SEZ6L (2.6), PTPRM (2.1), MN1 (2.1), FHL5 (2.0)
6	MP:0003063	Increased coping response			0.02	ANKRD20A11P (3.1), GLIS1 (2.8), CAND2 (2.5), NRXN3 (2.5), SIPA1L1 (2.1), MN1 (2.1), LINC00312 (-2.0)
6	GO:0005245	Voltage-gated calcium channel activity			0.03	ABLIM2 (3.2), SEZ6L (2.8), MN1 (2.3), ANKRD20A11P (2.2), NRXN3 (2.2), XPC (-2.0)
6	GO:0007628	Adult walking behavior			0.04	NRXN3 (2.7), TACR3 (2.4), LINC01648 (2.2), AC004944.1 (2.0)
6	GO:0051924	Regulation of calcium ion transport			0.04	LINC00299 (2.3), ANKRD20A11P (2.2), ABLIM2 (2.1)
7	MP:0009409	Abnormal skeletal muscle fiber type ratio	Abnormal skeletal muscle fiber morphology	Abnormal skeletal muscle fiber type ratio	0.0006	CAND2 (3.5), ABLIM2 (2.6), LINC01648 (2.3)
7	MP:0000005	Increased brown adipose tissue amount			0.01	FOXL1 (3.1), FHL5 (2.7)
7	MP:0009404	Centrally nucleated skeletal muscle fibers			0.01	CAND2 (4.1), LMCD1 (2.5), UBTD1 (2.0)
7	MP:0002841	Impaired skeletal muscle contractility			0.01	CAND2 (4.8), ABLIM2 (4.4), LMCD1 (2.1), RP1L1 (2.0)
7	MP:0004091	Abnormal z lines			0.01	CAND2 (3.2), UBTD1 (2.2), ABLIM2 (2.1), LMCD1 (2.0)
7	MP:0000747	Muscle weakness			0.01	ABLIM2 (2.7), CAND2 (2.4)
7	MP:0004087	Abnormal muscle fiber morphology			0.02	RUNX1 (3.7), MN1 (3.4), LMCD1 (3.2), LINC00312 (2.7), CAND2 (2.4), AKR1C3 (2.0), SRR (-2.0), SETD4 (-2.2)
7	MP:0002279	Abnormal diaphragm morphology			0.02	LMCD1 (4.5), CAND2 (3.2), MN1 (2.4), RUNX1 (2.3), PIEZO1 (2.2), UBTD1 (2.1)
7	MP:0000748	Progressive muscle weakness			0.02	LMCD1 (2.8), PEAK1 (2.3), IARS2 (2.2), UBTD1 (2.0)
7	MP:0003084	Abnormal skeletal muscle fiber morphology			0.03	CAND2 (3.7), LMCD1 (3.5), UBTD1 (2.9), PIEZO1 (2.5), ABLIM2 (2.4), LINC00312 (2.0)

7	MP:0009400	Decreased skeletal muscle fiber size			0.04	TRIB1 (3.0), LMCD1 (2.8), ABLIM2 (2.6), CAND2 (2.6), LINC00312 (2.0)
7	MP:0002332	Abnormal exercise endurance			0.04	LMCD1 (3.8), LINC00312 (3.5), TRIB1 (2.8), ABLIM2 (2.2), CAND2 (2.1), CBR1 (2.0), SRR (-2.2)
8	ENSG00000093167	LRRFIP2 PPI subnetwork	GSTM1 PPI subnetwork	LRRFIP2 PPI subnetwork	0.0009	SMG6 (3.5), AC068353.34 (2.7), LINC00973 (2.4), TMEM40 (2.3)
8	ENSG00000166033	HTRA1 PPI subnetwork			0.007	LMCD1 (4.5), AC068353.34 (3.1), PEAK1 (2.2), GLIS1 (2.0), FOXL1 (2.0)
8	ENSG00000171992	SYNPO PPI subnetwork			0.01	AC068353.34 (2.8), SMG6 (2.6), LMCD1 (2.1), ANKRD20A11P (2.1)
8	ENSG00000197321	SVIL PPI subnetwork			0.01	AC068353.34 (3.4), AKR1C3 (2.2), ANKRD20A11P (2.1), TMEM40 (2.1)
8	ENSG00000106028	SSBP1 PPI subnetwork			0.01	SMG6 (3.5), LMCD1 (3.1), AC068353.34 (2.4), PEAK1 (2.3)
8	ENSG00000167657	DAPK3 PPI subnetwork			0.02	LMCD1 (3.0)
8	ENSG00000196586	MYO6 PPI subnetwork			0.02	PEAK1 (2.7), SMG6 (2.6), LMCD1 (2.4), SEZ6L (2.3), TMEM40 (2.1)
8	ENSG00000134184	GSTM1 PPI subnetwork			0.02	AC068353.34 (2.9), ANKRD20A11P (2.5), SMG6 (2.3), LMCD1 (2.2), AKR1C3 (2.2)
8	ENSG00000071909	MYO3B PPI subnetwork			0.02	AC068353.34 (2.8), LMCD1 (2.4), ANKRD20A11P (2.2), AKR1C3 (2.0)
8	ENSG00000105357	MYH14 PPI subnetwork			0.02	LMCD1 (3.0), ANKRD20A11P (2.7), AC068353.34 (2.6), SMG6 (2.0)
8	ENSG00000198467	TPM2 PPI subnetwork			0.02	LMCD1 (4.4), AC068353.34 (2.5), SMG6 (2.2), LINC01013 (-2.2)
8	ENSG00000134202	GSTM3 PPI subnetwork			0.02	SMG6 (3.1), AC068353.34 (3.0), FOXL1 (2.6), LMCD1 (2.6), RASSF3 (2.6), PEAK1 (2.4), ABLIM2 (2.1), LINC01013 (-2.5)
8	ENSG00000100014	SPECC1L PPI subnetwork			0.02	ANKRD20A11P (2.7), AKR1C3 (2.6), AC068353.34 (2.1), SMG6 (2.1)
8	ENSG00000110880	CORO1C PPI subnetwork			0.03	SMG6 (2.4), AC068353.34 (2.1), LMCD1 (2.0), AKR1C3 (2.0), ANKRD20A11P (2.0)
8	ENSG00000101331	C20ORF160 PPI subnetwork			0.03	LMCD1 (2.3), AC068353.34 (2.2), SMG6

						(2.2), RASSF3 (2.0)
8	ENSG00000120265	PCMT1 PPI subnetwork			0.03	GLIS1 (2.5), PEAK1 (2.0)
8	ENSG00000006747	SCIN PPI subnetwork			0.04	SLC7A11 (2.3), LMCD1 (2.3), SMG6 (2.3), AC068353.34 (2.1)
8	ENSG00000128487	SPECC1 PPI subnetwork			0.04	LMCD1 (2.5), RASSF3 (2.2), SMG6 (2.1), AC068353.34 (2.1), PTPRM (2.0)
8	ENSG00000204361	FAM55B PPI subnetwork			0.04	AC068353.34 (2.1), LMCD1 (2.1), LINC01648 (-2.0)
9	GO:0034405	Response to fluid shear stress	TGF-beta receptor binding	Response to fluid shear stress	0.001	PIEZO1 (3.1), ERCC4 (2.6), AL450226.1 (2.5), LINC02064 (2.5), LINC01921 (2.1), LMCD1 (2.1), RASSF3 (2.0), LINC01648 (2.0)
9	ENSG00000139567	ACVRL1 PPI subnetwork			0.003	RASSF3 (2.1), RUNX1 (2.0)
9	ENSG00000198742	SMURF1 PPI subnetwork			0.008	HMG20A (3.7), PBX1 (2.7), LINC00312 (2.2)
9	GO:0060389	Pathway-restricted smad protein phosphorylation			0.009	LINC02064 (3.2), LMCD1 (2.3), GLIS1 (2.0), SRR (-2.0), TMEM40 (-2.2)
9	GO:0005160	TGF-beta receptor binding			0.02	LMCD1 (3.6), PTPRM (2.6), RASSF3 (2.4), AKR1C3 (2.1), LINC00312 (2.0)
9	ENSG00000120693	SMAD9 PPI subnetwork			0.03	UFL1 (2.3), TRIB1 (2.1)
9	ENSG00000163513	TGFBR2 PPI subnetwork			0.03	AC068353.34 (2.7), LMCD1 (2.5), RUNX1 (2.0), SLFN11 (-2.0)
9	ENSG00000123612	ACVR1C PPI subnetwork			0.03	LINC00973 (3.5), RUNX1 (3.2), LMCD1 (3.0), BPNT1 (-2.1)
9	GO:0017015	Regulation of transforming growth factor beta receptor signaling pathway			0.04	RASSF3 (2.5), LINC00299 (-2.0)
10	MP:0000223	Decreased monocyte cell number	Impaired neutrophil chemotaxis	Decreased monocyte cell number	0.001	BPNT1 (2.0)
10	MP:0002465	Abnormal eosinophil physiology			0.002	LINC01013 (3.3)
10	GO:0005544	Calcium-dependent phospholipid binding			0.005	PIEZO1 (2.5), AKR1C3 (2.3)
10	REACTOME PROSTANOID METABOLISM	Reactome prostanoid metabolism			0.01	AKR1C3 (3.1), SLFN11 (2.4), RUNX1 (2.2), PBX1 (2.2), LINC01013 (2.1)
10	ENSG00000196415	PRTN3 PPI subnetwork			0.02	NA
10	ENSG00000101336	HCK PPI subnetwork			0.02	RUNX1 (2.6), RAB3GAP2 (2.3), UBTD1 (2.1), HMG20A (2.0)

10	MP:0000219	Increased neutrophil cell number			0.03	SLFN11 (2.2), PEAK1 (2.0)
10	ENSG00000138032	PPM1B PPI subnetwork			0.03	PIEZO1 (2.6), RP1L1 (2.3)
10	MP:0008720	Impaired neutrophil chemotaxis			0.04	SLFN11 (2.1)
10	MP:0008657	Increased interleukin-1 beta secretion			0.04	LINC00299 (2.2), TRIB1 (2.0)
10	MP:0000511	Abnormal intestinal mucosa morphology			0.04	LINC01013 (2.6)
11	MP:0000761	Thin diaphragm muscle	Abnormal neuromuscular synapse morphology	Thin diaphragm muscle	0.001	LINC00973 (2.5), ABLIM2 (2.1)
11	MP:0004835	Abnormal miniature endplate potential			0.01	ABLIM2 (4.7), UBTD1 (3.4), LINC00973 (2.4), TACR3 (2.4), PEAK1 (2.3), RUNX1 (2.2), CAND2 (2.0)
11	MP:0002826	Tonic seizures			0.01	NRXN3 (3.2), LINC00312 (2.2), SRR (-2.0)
11	ENSG00000125968	ID1 PPI subnetwork			0.02	ABLIM2 (2.2)
11	MP:0001053	Abnormal neuromuscular synapse morphology			0.02	ABLIM2 (5.2), UBTD1 (3.6), PTPRM (3.0), TACR3 (2.6), NRXN3 (2.5), RUNX1 (2.3)
11	ENSG00000030304	MUSK PPI subnetwork			0.02	UBTD1 (3.6), RUNX1 (2.9), ABLIM2 (2.7), LMCD1 (2.2), TACR3 (2.1)
11	MP:0000753	Paralysis			0.03	RAB3GAP2 (3.8), CBR1 (3.6), NRXN3 (2.7), SLC7A11 (2.4), CAND2 (2.2), SLFN11 (2.1), RASSF3 (-2.7)
11	ENSG00000174469	CNTNAP2 PPI subnetwork			0.03	NRXN3 (2.1)
11	GO:0031594	Neuromuscular junction			0.04	ABLIM2 (3.6), CAND2 (3.0), RUNX1 (2.4)
11	MP:0002914	Abnormal endplate potential			0.04	ABLIM2 (3.9), LINC01013 (3.4), RUNX1 (2.9), UBTD1 (2.9), LINC00973 (2.6), CAND2 (2.2)
12	ENSG00000165702	GFI1B PPI subnetwork	Actomyosin	GFI1B PPI subnetwork	0.001	SMG6 (2.2)
12	GO:0051015	Actin filament binding			0.002	AC068353.34 (2.3), ANKRD20A11P (2.3), PIEZO1 (2.0)
12	GO:0032432	Actin filament bundle			0.009	LMCD1 (4.0), CAND2 (3.2), LINC00312 (2.9), PIEZO1 (2.6), AC068353.34 (2.5), RASSF3 (2.1), SIPA1L1 (2.1)
12	GO:0001725	Stress fiber			0.009	LMCD1 (3.7), CAND2 (3.2), LINC00312 (2.9), AC068353.34 (2.3), RASSF3 (2.3), PIEZO1 (2.1), FHL5 (2.0)

12	ENSG00000163466	ARPC2 PPI subnetwork			0.01	AC068353.34 (2.7), SMG6 (2.5), PEAK1 (2.1), SRR (2.1)
12	GO:0042641	Actomyosin			0.01	LMCD1 (3.6), CAND2 (3.2), LINC00312 (2.6), RASSF3 (2.5), SIPA1L1 (2.1), PIEZO1 (2.1), SMG6 (2.1), AC068353.34 (2.0)
12	ENSG00000135269	TES PPI subnetwork			0.03	RASSF3 (2.9), SIPA1L1 (2.1), CAND2 (2.0), LMCD1 (2.0), PBX1 (2.0), LINC00973 (-2.5)
12	ENSG00000145349	CAMK2D PPI subnetwork			0.03	RAB3GAP2 (2.6), CAND2 (2.4), SIPA1L1 (2.2), RASSF3 (2.0)
12	REACTOME SMOOTH MUSCLE CONTRACTION	Reactome smooth muscle contraction			0.03	FHL5 (4.7), RASSF3 (3.8), LMCD1 (3.6), NRXN3 (2.5), LINC00312 (2.2), LINC01648 (-2.1)
12	ENSG00000150867	PIP4K2A PPI subnetwork			0.03	SIPA1L1 (2.7), AC068353.34 (2.7), PEAK1 (2.3), ABLIM2 (2.2), RAB3GAP2 (2.1), ERCC4 (2.1), CAND2 (2.0), SEZ6L (-2.1)
12	ENSG00000154556	SORBS2 PPI subnetwork			0.03	PEAK1 (2.2)
12	ENSG00000144061	NPHP1 PPI subnetwork			0.04	MN1 (3.0), PIEZO1 (2.8), ERCC4 (-2.0)
12	GO:0005925	Focal adhesion			0.04	AC068353.34 (2.6), LMCD1 (2.6), PIEZO1 (2.6), RASSF3 (2.5), XPC (2.4), LINC00312 (2.0), LINC01648 (-3.0)
12	ENSG00000106976	DNM1 PPI subnetwork			0.04	RAB3GAP2 (2.0)
12	GO:0015629	Actin cytoskeleton			0.04	LMCD1 (3.5), CAND2 (3.3), RASSF3 (2.1)
13	MP:0006354	Abnormal fourth branchial arch artery morphology	Abnormal fourth branchial arch artery morphology	Abnormal fourth branchial arch artery morphology	0.002	FOXL1 (5.3), FHL5 (3.2), PEAK1 (2.8), PIEZO1 (2.1), PTPRM (2.0)
13	MP:0010465	Aberrant origin of the right subclavian artery			0.004	PIEZO1 (2.7), PBX1 (2.3), TACR3 (2.2), CAND2 (2.2), SIPA1L1 (-2.2)
13	MP:0004157	Interrupted aortic arch			0.004	FOXL1 (3.0), ERCC4 (2.6), PIEZO1 (2.3), TACR3 (2.2), RUNX1 (2.2)
13	GO:0031490	Chromatin DNA binding			0.01	FOXL1 (5.4), MN1 (3.2), LINC00973 (2.7), RUNX1 (2.5), FHL5 (2.3), UBTD1 (2.0), SEZ6L (2.0), IARS2 (-2.0)
13	GO:0048844	Artery morphogenesis			0.02	FOXL1 (4.4), GLIS1 (2.3), LMCD1 (2.0)
13	GO:0001569	Patterning of blood vessels			0.04	LINC01648 (3.2), FHL5 (2.6), FOXL1 (2.3), PEAK1 (2.3)

13	MP:0004158	Right aortic arch			0.04	TRIB1 (2.6), UFL1 (2.4), ERCC4 (2.1), RUNX1 (2.0)
14	GO:0043073	Germ cell nucleus	Germ cell nucleus	Germ cell nucleus	0.00238	LINC01921 (4.8), FHL5 (2.9), LINC01013 (2.3), SRR (2.3)
14	GO:0001673	Male germ cell nucleus			0.00313	LINC01921 (4.9), FHL5 (3.3), SRR (2.6), LINC02064 (2.1), LINC01013 (2.1)
15	MP:0009230	Abnormal sperm head morphology	Teratozoospermia	Abnormal sperm head morphology	0.00244	LINC01921 (3.4), FHL5 (3.3), LINC00312 (2.6), HDGFL1 (2.4)
15	MP:0009238	Coiled sperm flagellum			0.01	LINC01921 (6.0), FHL5 (2.3), AL450226.1 (2.0)
15	MP:0004542	Impaired acrosome reaction			0.01	LINC01921 (3.8), FHL5 (3.7), UBTD1 (3.4), AC004944.1 (2.7)
15	MP:0005578	Teratozoospermia			0.02	LINC01921 (5.7), FHL5 (3.7), HDGFL1 (3.6), UBTD1 (2.0)
15	MP:0000242	Impaired fertilization			0.03	LINC01921 (4.5), HDGFL1 (2.7), FHL5 (2.6), AC004944.1 (2.4), XPC (-2.7)
15	MP:0009232	Abnormal sperm nucleus morphology			0.03	LINC01921 (5.6), FHL5 (2.7), HDGFL1 (2.0), AL450226.1 (2.0)
15	GO:0007281	Germ cell development			0.04	HDGFL1 (5.3), LINC01921 (5.2), FHL5 (2.7)
15	MP:0010769	Abnormal survival			0.04	TACR3 (2.4), RASSF3 (2.4)
16	ENSG00000125503	PPP1R12C PPI subnetwork	MYO1G PPI subnetwork	PPP1R12C PPI subnetwork	0.003	PIEZO1 (2.2), AC068353.34 (2.2), ANKRD20A11P (2.1)
16	ENSG00000106992	AK1 PPI subnetwork			0.004	RASSF3 (2.1)
16	ENSG00000136286	MYO1G PPI subnetwork			0.005	RASSF3 (2.2), ANKRD20A11P (2.0)
16	ENSG00000013297	CLDN11 PPI subnetwork			0.005	RASSF3 (2.4), ANKRD20A11P (2.4), PIEZO1 (2.0)
16	ENSG00000166866	MYO1A PPI subnetwork			0.01	AC068353.34 (3.8), RASSF3 (2.7), LMCD1 (2.4), PEAK1 (2.3), LINC01648 (-2.2)
16	ENSG00000118260	CREB1 PPI subnetwork			0.02	UBTD1 (3.2), SMG6 (2.2), LINC01921 (2.1)
16	ENSG00000169220	RGS14 PPI subnetwork			0.02	RASSF3 (2.7)
16	ENSG00000136156	ITM2B PPI subnetwork			0.02	RASSF3 (2.4), AC068353.34 (2.2), ANKRD20A11P (2.0)

16	ENSG00000136068	FLNB PPI subnetwork			0.02	SRR (3.2), AC068353.34 (2.6), SMG6 (2.4), PIEZO1 (2.1), RASSF3 (2.0), UBTD1 (2.0)
16	ENSG00000105376	ICAM5 PPI subnetwork			0.03	RASSF3 (2.1), ANKRD20A11P (2.0)
16	ENSG00000139514	SLC7A1 PPI subnetwork			0.03	ANKRD20A11P (2.3), RASSF3 (2.2), GLIS1 (-2.0)
17	GO:0043500	Muscle adaptation	Muscle system process	Muscle adaptation	0.003	LINC00312 (3.9), RASSF3 (2.5), CAND2 (2.2), AL450226.1 (-2.5)
17	GO:0090257	Regulation of muscle system process			0.01	CAND2 (3.6), TACR3 (2.1)
17	GO:0031143	Pseudopodium			0.01	CAND2 (4.2), LMCD1 (3.7), ABLIM2 (2.1)
17	GO:0003012	Muscle system process			0.02	CAND2 (4.7), ABLIM2 (2.7), RASSF3 (2.5), LMCD1 (2.1), ANKRD20A11P (2.0)
17	GO:0006937	Regulation of muscle contraction			0.03	CAND2 (3.7)
17	ENSG00000109846	CRYAB PPI subnetwork			0.03	CAND2 (2.8), SLC7A11 (2.5), FHL5 (-2.6)
17	ENSG00000135424	ITGA7 PPI subnetwork			0.03	FHL5 (2.9), ABLIM2 (2.5), TACR3 (2.4), UBTD1 (2.4)
17	GO:0055117	Regulation of cardiac muscle contraction			0.04	LINC01013 (3.9), RASSF3 (2.1), AC004944.1 (2.0), LINC00312 (-2.4)
17	GO:0048747	Muscle fiber development			0.04	AL450226.1 (3.2), CAND2 (3.2), ABLIM2 (3.0), RASSF3 (2.2), TACR3 (2.0), LINC01921 (2.0), SLFN11 (-2.4)
17	ENSG00000177791	MYOZ1 PPI subnetwork			0.04	CAND2 (3.8), LMCD1 (2.2)
17	GO:0016460	Myosin ii complex			0.04	CAND2 (4.2), LMCD1 (3.0)
17	MP:0005140	Decreased cardiac muscle contractility			0.04	LMCD1 (3.4), CAND2 (3.2)
17	GO:0000146	Microfilament motor activity			0.04	CAND2 (2.8), LMCD1 (2.5)
17	MP:0005620	Abnormal muscle contractility			0.04	FHL5 (3.0), CAND2 (3.0), NRXN3 (2.5), ABLIM2 (2.2), LINC00312 (2.1), LMCD1 (2.0)
17	GO:0006936	Muscle contraction			0.04	CAND2 (4.8), ABLIM2 (2.7), RASSF3 (2.5), LMCD1 (2.0)
17	GO:0005516	Calmodulin binding			0.04	SIPA1L1 (3.1), ABLIM2 (3.1), LMCD1 (2.2), CAND2 (2.1)

18	ENSG00000122641	INHBA PPI subnetwork	INHBA PPI subnetwork	INHBA PPI subnetwork	0.004	RUNX1 (4.5), LMCD1 (4.0)
18	ENSG00000106991	ENG PPI subnetwork			0.02	LMCD1 (4.2), RUNX1 (2.6), LINC01013 (2.1), SRR (-2.2)
18	MP:0003645	Increased pancreatic beta cell number			0.03	AC068353.34 (2.5), LINC00312 (2.0)
18	ENSG00000175189	INHBC PPI subnetwork			0.03	RUNX1 (5.3), LMCD1 (3.7), LINC00973 (2.6), MN1 (2.2)
18	GO:0001558	Regulation of cell growth			0.04	LMCD1 (3.0), ANKRD20A11P (2.7), AC004944.1 (2.4), GLIS1 (2.2)
18	ENSG00000163083	INHBB PPI subnetwork			0.04	RUNX1 (4.6), LMCD1 (2.8), MN1 (2.6)
19	GO:0034504	Protein localization to nucleus	Protein localization to nucleus	Protein localization to nucleus	0.004	MN1 (2.9), XPC (2.0), PBX1 (2.0)
19	GO:0006606	Protein import into nucleus			0.03	MN1 (2.8), PBX1 (2.1), ANKRD20A11P (-2.0)
19	GO:0051170	Nuclear import			0.04	MN1 (2.8), PBX1 (2.0), RUNX1 (2.0)
19	GO:0033365	Protein localization to organelle			0.04	XPC (2.0)
20	ENSG00000158290	CUL4B PPI subnetwork	VPRBP PPI subnetwork	CUL4B PPI subnetwork	0.004	HMG20A (4.5), RAB3GAP2 (3.5), IARS2 (2.7), XPC (2.6), UFL1 (2.5), AC068353.34 (2.3), CDC5L (2.3), LINC00299 (2.2)
20	ENSG00000003096	KLHL13 PPI subnetwork			0.008	CDC5L (3.3), RAB3GAP2 (2.9), SIPA1L1 (2.5), IARS2 (2.4)
20	ENSG00000108094	CUL2 PPI subnetwork			0.01	HMG20A (2.6)
20	ENSG00000139842	CUL4A PPI subnetwork			0.01	HMG20A (4.0), IARS2 (2.8), XPC (2.5), LINC00299 (2.0), RAB3GAP2 (2.0), SLC7A11 (-2.5)
20	ENSG00000145041	VPRBP PPI subnetwork			0.01	CDC5L (3.2), RAB3GAP2 (3.1), SIPA1L1 (2.5), IARS2 (2.3), CAND2 (2.1)
20	ENSG00000120696	KBTBD7 PPI subnetwork			0.03	CDC5L (3.1), RAB3GAP2 (2.9), UFL1 (2.3), CAND2 (2.0)
21	ENSG00000112559	MDFI PPI subnetwork	TRAF2 PPI subnetwork	MDFI PPI subnetwork	0.004	LINC01921 (2.2), LINC00973 (2.1), UBTD1 (2.1), SETD4 (2.0)
21	ENSG00000127191	TRAF2 PPI subnetwork			0.004	RAB3GAP2 (2.7)
21	ENSG00000064300	NGFR PPI subnetwork			0.02	RUNX1 (2.0)
21	ENSG00000183763	TRAIP PPI subnetwork			0.03	SETD4 (2.3), TRIB1 (-2.1)

21	ENSG00000056972	TRAF3IP2 PPI subnetwork			0.04	SLC7A11 (2.5)
22	ENSG00000164091	WDR82 PPI subnetwork	Reactome transport of mature mrna derived from an intronless transcript	WDR82 PPI subnetwork	0.004	HMG20A (5.0), XPC (3.4), LINC00299 (2.8), RAB3GAP2 (2.1)
22	ENSG00000136997	MYC PPI subnetwork			0.008	HMG20A (2.7), RAB3GAP2 (2.5)
22	ENSG00000139496	NUPL1 PPI subnetwork			0.0095	CDC5L (2.9), RAB3GAP2 (2.6), IARS2 (2.5), TSR1 (2.0)
22	REACTOME PURINE RIBONUCLEOSIDE MONOPHOSPHATE BIOSYNTHESIS	Reactome purine ribonucleoside monophosphate biosynthesis			0.01	IARS2 (4.5), TSR1 (2.3), XPC (2.2), LINC01921 (2.0), ANKRD20A11P (-2.8)
22	REACTOME GLUCOSE TRANSPORT	Reactome glucose transport			0.02	RAB3GAP2 (2.2)
22	ENSG00000093000	NUP50 PPI subnetwork			0.03	RAB3GAP2 (2.5), CDC5L (2.4)
22	REACTOME ISG15 ANTIVIRAL MECHANISM	Reactome isg15 antiviral mechanism			0.03	NA
22	REACTOME ANTIVIRAL MECHANISM BY IFN:STIMULATED GENES	Reactome antiviral mechanism by ifn:stimulated genes			0.03	NA
22	ENSG00000124789	NUP153 PPI subnetwork			0.03	CDC5L (3.4), IARS2 (3.0), TSR1 (2.1), RAB3GAP2 (2.0)
22	REACTOME HEXOSE TRANSPORT	Reactome hexose transport			0.03	HDGFL1 (2.3), RAB3GAP2 (2.3)
22	ENSG00000132341	RAN PPI subnetwork			0.04	CDC5L (3.2), TSR1 (2.5)
22	REACTOME TRANSPORT OF MATURE MRNA DERIVED FROM AN INTRONLESS TRANSCRIPT	Reactome transport of mature mRNA derived from an intronless transcript			0.04	RAB3GAP2 (2.3),CDC5L (2.3),IARS2 (2.3),UBTD1 (-2.3)
22	ENSG00000126883	NUP214 PPI subnetwork			0.04	CDC5L (2.5),IARS2 (2.2),RAB3GAP2 (2.0)
22	ENSG00000124571	XPO5 PPI subnetwork			0.04	CDC5L (2.6),RAB3GAP2 (2.0),ANKRD20A11P (-2.1)
22	ENSG00000101146	RAE1 PPI subnetwork			0.04	CDC5L (3.7),TSR1 (3.4),RAB3GAP2 (2.4)
22	REACTOME TRANSPORT OF MATURE MRNAS DERIVED FROM INTRONLESS TRANSCRIPTS	Reactome transport of mature mRNAs derived from intronless transcripts			0.04	CDC5L (2.4),IARS2 (2.2),RAB3GAP2 (2.1),UBTD1 (-2.2)
23	ENSG00000165156	ZHX1 PPI subnetwork	HIST1H2AG PPI subnetwork	ZHX1 PPI subnetwork	0.004	AKR1C3 (2.5),SLC7A11 (2.1),SLFN11 (2.1),ABLIM2 (-2.3),CAND2 (-3.2)
23	ENSG00000180370	PAK2 PPI subnetwork			0.01	SMG6 (2.3),PIEZO1 (2.2),TACR3 (-2.0)

23	ENSG00000198959	TGM2 PPI subnetwork			0.02	PIEZO1 (3.8),TMEM40 (2.7),RASSF3 (2.4),UFL1 (-2.0)
23	ENSG00000068654	POLR1A PPI subnetwork			0.02	TSR1 (3.6),CDC5L (2.5)
23	ENSG00000177889	UBE2N PPI subnetwork			0.02	XPC (2.6),BPNT1 (2.2)
23	ENSG00000112592	TBP PPI subnetwork			0.03	CDC5L (2.5),ERCC4 (2.4)
23	ENSG00000168274	HIST1H2AE PPI subnetwork			0.03	CDC5L (2.7),TSR1 (2.4)
23	ENSG00000137259	HIST1H2AB PPI subnetwork			0.03	CDC5L (2.7),TSR1 (2.4)
23	ENSG00000143190	POU2F1 PPI subnetwork			0.03	AC068353.34 (2.0)
23	ENSG00000196866	HIST1H2AD PPI subnetwork			0.03	CDC5L (2.9),TSR1 (2.1),PBX1 (2.0)
23	ENSG00000198374	HIST1H2AL PPI subnetwork			0.03	CDC5L (2.9),TSR1 (2.1),PBX1 (2.0)
23	ENSG00000196747	HIST1H2AI PPI subnetwork			0.03	CDC5L (2.9),TSR1 (2.1),PBX1 (2.0)
23	ENSG00000196787	HIST1H2AG PPI subnetwork			0.03	CDC5L (2.9),TSR1 (2.1),PBX1 (2.0)
23	ENSG00000184348	HIST1H2AK PPI subnetwork			0.03	CDC5L (2.9),TSR1 (2.1),PBX1 (2.0)
23	ENSG00000137055	PLAA PPI subnetwork			0.03	XPC (2.2)
23	ENSG00000145216	FIP1L1 PPI subnetwork			0.03	ERCC4 (2.6),HMG20A (2.5),LINC02064 (2.5)
23	ENSG00000153147	SMARCA5 PPI subnetwork			0.04	CDC5L (2.7),RUNX1 (2.4),HMG20A (2.3)
23	ENSG00000183558	HIST2H2AA3 PPI subnetwork			0.04	CDC5L (2.6),TSR1 (2.1),HMG20A (2.0)
23	ENSG00000203812	HIST2H2AA4 PPI subnetwork			0.04	CDC5L (2.6),TSR1 (2.1),HMG20A (2.0)
23	ENSG00000188486	H2AFX PPI subnetwork			0.04	CDC5L (3.1)
24	GO:0045624	Positive regulation of t-helper cell differentiation	Positive regulation of cd4-positive, alpha-beta t cell activation	Positive regulation of t-helper cell differentiation	0.005	RUNX1 (2.0)
24	GO:2000516	Positive regulation of cd4-positive, alpha-beta t cell activation			0.007	NA
24	GO:0043372	Positive regulation of cd4-positive, alpha-beta t cell differentiation			0.007	NA
24	GO:0046637	Regulation of alpha-beta t cell differentiation			0.04	RUNX1 (2.2),LINC00299 (2.1)
24	GO:0045086	Positive regulation of interleukin-2 biosynthetic process			0.04	AC004944.1 (3.7),ERCC4 (2.3)

25	GO:0042744	Hydrogen peroxide catabolic process	Reticulocytosis	Hydrogen peroxide catabolic process	0.005	LINC01921 (3.5),LINC02064 (3.4),AKR1C3 (3.0),ANKRD20A11P (2.3),BPNT1 (2.2)
25	MP:0002640	Reticulocytosis			0.009	AKR1C3 (2.7),BPNT1 (-2.3)
25	GO:0015923	Mannosidase activity			0.01	UFL1 (4.7),PIEZO1 (2.3)
25	GO:0048872	Homeostasis of number of cells			0.03	AKR1C3 (3.2),ERCC4 (2.9)
25	MP:0001585	Hemolytic anemia			0.04	AKR1C3 (3.1)
26	ENSG00000088832	FKBP1A PPI subnetwork	ARF6 PPI subnetwork	FKBP1A PPI subnetwork	0.006	SLC7A11 (2.6)
26	ENSG00000165527	ARF6 PPI subnetwork			0.008	IARS2 (3.9),SLC7A11 (2.8),HDGFL1 (2.8),UFL1 (2.5)
26	ENSG00000184489	PTP4A3 PPI subnetwork			0.01	IARS2 (3.8),UFL1 (3.8),RAB3GAP2 (2.3),TSR1 (2.2),HDGFL1 (2.2),BPNT1 (2.0)
26	REACTOME ASSOCIATION OF TRICCCT WITH TARGET PROTEINS DURING BIOSYNTHESIS	Reactome association of triccct with target proteins during biosynthesis			0.02	IARS2 (2.8),HDGFL1 (2.5),UFL1 (2.4)
26	ENSG00000198824	CHAMP1 PPI subnetwork			0.02	UFL1 (2.8),TSR1 (2.5),RUNX1 (2.4),CDC5L (2.2),FOXL1 (-2.4)
26	ENSG00000139637	C12ORF10 PPI subnetwork			0.02	SLC7A11 (2.4),BPNT1 (2.4),FHL5 (-2.6)
26	ENSG00000028137	TNFRSF1B PPI subnetwork			0.03	IARS2 (2.9),SLC7A11 (2.4),UFL1 (2.3),TSR1 (2.1)
26	ENSG00000180104	EXOC3 PPI subnetwork			0.03	RAB3GAP2 (3.6),UFL1 (2.5),CDC5L (2.5),SEZ6L (2.0)
26	ENSG00000129465	RIPK3 PPI subnetwork			0.03	IARS2 (2.9)
26	ENSG00000132432	SEC61G PPI subnetwork			0.03	UFL1 (5.5)
26	ENSG00000173530	TNFRSF10D PPI subnetwork			0.03	IARS2 (3.3),XPC (2.5),SLC7A11 (2.5),TSR1 (2.2),CDC5L (2.1)
26	ENSG00000141480	ARRB2 PPI subnetwork			0.04	RASSF3 (2.3),SMG6 (2.3),UBTD1 (2.3),AC004944.1 (-2.1)
26	ENSG00000068323	TFE3 PPI subnetwork			0.04	IARS2 (3.7),SLC7A11 (2.8),CDC5L (2.0)
27	GO:0045933	Positive regulation of muscle contraction	Positive regulation of smooth muscle contraction	Positive regulation of muscle contraction	0.007	TACR3 (2.7),LINC01013 (2.6),LINC00312 (2.1),AC004944.1 (2.0)
27	GO:0045987	Positive regulation of smooth muscle contraction			0.03	TACR3 (3.2),LINC00299 (3.0),LINC00312 (2.4),LINC01013 (2.0)

28	ENSG00000078061	ARAF PPI subnetwork	SHCBP1 PPI subnetwork	ARAF PPI subnetwork	0.007	CAND2 (2.4),SIPA1L1 (2.1),AL450226.1 (-2.3)
28	ENSG00000120509	PDZD11 PPI subnetwork			0.008	UBTD1 (2.0),AL450226.1 (-2.7)
28	ENSG00000136153	LMO7 PPI subnetwork			0.01	SIPA1L1 (2.9),TMEM40 (2.9),CAND2 (2.7),RASSF3 (2.6),GLIS1 (2.0)
28	ENSG00000131941	RHPN2 PPI subnetwork			0.02	AC068353.34 (3.0),TMEM40 (2.5),LINC01648 (2.1),PTPRM (2.0)
28	ENSG00000161800	RACGAP1 PPI subnetwork			0.02	AC068353.34 (3.4),LINC00312 (2.2)
28	ENSG00000143375	CGN PPI subnetwork			0.02	TMEM40 (2.3),AC068353.34 (2.3)
28	ENSG00000171241	SHCBP1 PPI subnetwork			0.02	AC068353.34 (2.8),TMEM40 (2.6)
28	ENSG00000168476	REEP4 PPI subnetwork			0.03	AC068353.34 (2.4),TMEM40 (2.4),LINC01648 (2.0),PTPRM (2.0)
28	ENSG00000125731	SH2D3A PPI subnetwork			0.03	TMEM40 (2.7),AC068353.34 (2.6),SETD4 (-2.0)
28	ENSG00000115963	RND3 PPI subnetwork			0.04	TMEM40 (2.8),AC068353.34 (2.7),PTPRM (2.0),SRR (-2.7)
28	ENSG00000189319	FAM53B PPI subnetwork			0.04	TMEM40 (2.5),AC068353.34 (2.5)
28	ENSG00000163362	C1ORF106 PPI subnetwork			0.04	TMEM40 (2.5),AC068353.34 (2.3),PTPRM (2.0)
28	ENSG00000138771	SHROOM3 PPI subnetwork			0.04	AC068353.34 (2.5),TMEM40 (2.5),LINC01648 (2.0),SETD4 (-2.0)
28	ENSG00000197555	SIPA1L1 PPI subnetwork			0.04	AC068353.34 (2.5),TMEM40 (2.4),LINC01648 (2.0)
28	GO:0030216	Keratinocyte differentiation			0.04	TMEM40 (4.9)
28	ENSG00000075413	MARK3 PPI subnetwork			0.04	FHL5 (3.1),AC068353.34 (2.2)
28	ENSG00000171219	CDC42BPG PPI subnetwork			0.04	AC068353.34 (2.8),TMEM40 (2.3),LINC01648 (2.3)
29	GO:0017119	Golgi transport complex	Golgi transport complex	Golgi transport complex	0.008	UFL1 (4.4),RAB3GAP2 (4.2),IARS2 (2.8),BPNT1 (2.3)
29	ENSG00000135775	COG2 PPI subnetwork			0.02	RAB3GAP2 (2.2),LINC01013 (2.1),XPC (2.0)
29	ENSG00000164597	COG5 PPI subnetwork			0.02	RAB3GAP2 (2.6)

29	GO:0007030	Golgi organization			0.04	UFL1 (4.7),RAB3GAP2 (3.6)
29	ENSG00000092108	SCFD1 PPI subnetwork			0.04	UFL1 (3.9),SLFN11 (2.3)
29	ENSG00000103051	COG4 PPI subnetwork			0.04	UFL1 (5.0),RAB3GAP2 (3.2),GLIS1 (2.8)
30	ENSG00000184575	XPOT PPI subnetwork	SPTLC1 PPI subnetwork	XPOT PPI subnetwork	0.008	CDC5L (2.8),PTPRM (2.0)
30	ENSG00000090054	SPTLC1 PPI subnetwork			0.03	IARS2 (3.0),CDC5L (2.7),TSR1 (2.4)
30	ENSG00000137767	SQRDL PPI subnetwork			0.04	PIEZO1 (2.6),UFL1 (2.0)
31	MP:0001147	Small testis	Azoospermia	Small testis	0.008	HDGFL1 (5.1),LINC01921 (2.6),ANKRD20A11P (2.3)
31	MP:0008261	Arrest of male meiosis			0.04	HDGFL1 (5.0)
31	MP:0005159	Azoospermia			0.04	HDGFL1 (5.9),LINC01921 (2.0)
32	ENSG00000082781	ITGB5 PPI subnetwork	ITGB6 PPI subnetwork	ITGB5 PPI subnetwork	0.009	PIEZO1 (2.9),LMCD1 (2.9),LINC00973 (2.6),UBTD1 (2.3),PEAK1 (2.1),PTPRM (2.0),LINC01648 (-2.4),BPNT1 (-2.6)
32	ENSG00000115221	ITGB6 PPI subnetwork			0.01	PIEZO1 (2.9),LINC01648 (-2.1)
32	ENSG00000105855	ITGB8 PPI subnetwork			0.02	PIEZO1 (2.8),FOXL1 (2.3)
32	GO:0008305	Integrin complex			0.04	TMEM40 (3.4),PIEZO1 (2.8),RASSF3 (2.0),LINC01013 (-2.7)
32	MP:0000069	Kyphoscoliosis			0.04	PIEZO1 (2.6),PTPRM (2.0)
33	GO:0050708	Regulation of protein secretion	Regulation of protein secretion	Regulation of protein secretion	0.009	NA
33	GO:0050707	Regulation of cytokine secretion			0.02	LINC00299 (3.8),LINC01013 (2.8),SLFN11 (2.3)
33	GO:0051048	Negative regulation of secretion			0.04	LINC01648 (2.7),SEZ6L (2.1),HMG20A (-2.3)
33	GO:0009306	Protein secretion			0.04	NA
33	GO:0070201	Regulation of establishment of protein localization			0.04	NA
34	ENSG00000197905	TEAD4 PPI subnetwork	Abnormal eye anterior chamber morphology	TEAD4 PPI subnetwork	0.01	FOXL1 (3.2),SLFN11 (2.9),PTPRM (2.2),GLIS1 (2.1),UBTD1 (2.0),TMEM40 (2.0),ABLIM2 (-2.0),ANKRD20A11P (-2.2)
34	MP:0005205	Abnormal eye anterior chamber morphology			0.01	MN1 (3.4),LINC00973 (3.0),LMCD1 (2.1),LINC02064 (2.0)

34	MP:0001322	Abnormal iris morphology			0.02	NA
34	MP:0005542	Corneal vascularization			0.02	FASLG (2.3),HMG20A (-2.1)
34	MP:0004624	Abnormal thoracic cage morphology			0.03	MN1 (2.3),PBX1 (2.2)
35	MP:0004876	Decreased mean systemic arterial blood pressure	Regulation of tube size	Decreased mean systemic arterial blood pressure	0.01	FHL5 (3.8),FOXL1 (2.8),PTPRM (2.4),TACR3 (2.3),LINC00299 (2.3)
35	MP:0008874	Decreased physiological sensitivity to xenobiotic			0.02	SLC7A11 (3.2),FOXL1 (2.6),ABLIM2 (2.0)
35	GO:0035150	Regulation of tube size			0.03	FHL5 (3.5),FOXL1 (3.0)
35	GO:0050880	Regulation of blood vessel size			0.03	FHL5 (3.6),FOXL1 (3.0)
35	GO:0042310	Vasoconstriction			0.03	FHL5 (4.0),FOXL1 (3.2)
35	MP:0001613	Abnormal vasodilation			0.04	FHL5 (2.7),PIEZO1 (2.2),PTPRM (2.1),CBR1 (2.0),HMG20A (-2.0)
36	ENSG00000166930	MS4A5 PPI subnetwork	MS4A5 PPI subnetwork	MS4A5 PPI subnetwork	0.01	NA
36	ENSG00000173702	MUC13 PPI subnetwork			0.01	NA
36	ENSG00000169306	IL1RAPL1 PPI subnetwork			0.01	ANKRD20A11P (2.3)
36	ENSG00000198478	SH3BGRL2 PPI subnetwork			0.03	NA

Supplementary Table 3. Enrichment of tissue and cell types from DEPICT. Suggestive enrichment (FDR < 0.20) was observed for 21 categories of tissues and/or cell types. Two categories where GILS1 was highly expressed were added. Genes listed were annotated to a given tissue/cell types and are within an associated locus. Each gene has a z-score representing the gene's likelihood to be highly expressed in the tissue/cell types. Genes displayed are those with a |z-score| > 1.96 that corresponds to a P-value<0.05 as the minimum threshold, or the gene with higher absolute value of z-score (*). MeSH: Medical Subject Headings, FDR: false discovery rate.

MeSH term	Name	MeSH first level term	MeSH second level term	P value	FDR	Contributing genes
A15.145.846	Serum	Hemic and Immune Systems	Blood	2.61E-03	<0.20	LINC01013 (2.9),UBTD1 (2.6),ERCC4 (2.0),UFL1 (-2.1)
A10.690.467	Muscle Smooth	Tissues	Muscles	3.94E-03	<0.20	FOXL1 (3.2)
A07.231	Blood Vessels	Cardiovascular System	Blood Vessels	4.25E-03	<0.20	LINC01013 (4.1),UBTD1 (2.3)
A11.620.520	Myocytes Smooth Muscle	Cells	Muscle Cells	5.36E-03	<0.20	FOXL1 (4.8),TACR3 (2.6),LINC00973 (2.4),LINC01648 (2.1),SRR (2.0)
A11.620	Muscle Cells	Cells	Muscle Cells	5.36E-03	<0.20	FOXL1 (4.8),TACR3 (2.6),LINC00973 (2.4),LINC01648 (2.1),SRR (2.0)
A11.872.190.260	Embryoid Bodies	Cells	Stem Cells	5.92E-03	<0.20	TACR3 (2.6)
A11.329.228	Fibroblasts	Cells	Connective Tissue Cells	6.72E-03	<0.20	FOXL1 (2.2),SRR (2.1)
A11.329	Connective Tissue Cells	Cells	Connective Tissue Cells	8.20E-03	<0.20	*SLC7A11 (1.3)
A07.231.908	Veins	Cardiovascular System	Blood Vessels	9.70E-03	<0.20	LINC01013 (5.8),UBTD1 (2.6),SLC7A11 (2.0)
A10.165.450.300.425	Keloid	Tissues	Connective Tissue	1.00E-02	<0.20	MN1 (3.4),PEAK1 (3.1),XPC (2.7),HDGFL1 (2.5),LINC02064 (2.3),**GLIS1 (2.3)**,LINC01648 (2.3),SLC7A11 (2.0),TRIB1 (-2.5)
A05.360.444.492	Penis	Urogenital System	Genitalia	1.00E-02	<0.20	*LINC01921 (1.9)
A05.360.444.492.362	Foreskin	Urogenital System	Genitalia	1.00E-02	<0.20	LINC01921 (2.0)
A07.231.908.670.874	Umbilical Veins	Cardiovascular System	Blood Vessels	1.00E-02	<0.20	LINC01013 (6.3),UBTD1 (2.8),SLC7A11 (2.2)
A07.231.908.670	Portal System	Cardiovascular System	Blood Vessels	1.00E-02	<0.20	LINC01013 (6.3),UBTD1 (2.8),SLC7A11 (2.2)
A10.165.450.300	Cicatrix	Tissues	Connective Tissue	1.00E-02	<0.20	MN1 (3.1),PEAK1 (2.9),HDGFL1 (2.4),XPC (2.3),LINC02064 (2.2),LINC01648 (2.0),TRIB1 (-2.3)
A10.165.450	Granulation Tissue	Tissues	Connective Tissue	1.00E-02	<0.20	MN1 (3.1),PEAK1 (2.9),HDGFL1 (2.4),XPC (2.3),LINC02064 (2.2),LINC01648 (2.0),TRIB1 (-2.3)
A11.436.275	Endothelial Cells	Cells	Epithelial Cells	1.00E-02	<0.20	LINC01013 (3.8),UBTD1 (2.4)
A10.690	Muscles	Tissues	Muscles	2.00E-02	<0.20	LINC00312 (3.3),MN1 (2.7),CAND2 (2.4),LMCD1

						(2.4),ABLIM2 (2.4),LINC02064 (2.3)
A07.231.114	Arteries	Cardiovascular System	Blood Vessels	2.00E-02	<0.20	FOXL1 (2.7),FHL5 (2.4),LINC01013 (2.0)
A11.329.629	Osteoblasts	Cells	Connective Tissue Cells	2.00E-02	<0.20	LINC01013 (4.1),ERCC4 (2.9),**GLIS1 (2.8)**,PEAK1 (2.7),SLC7A11 (2.5),LMCD1 (2.3),SIPA1L1 (2.1),PTPRM (2.1),SRR (2.0)
A11.872.580	Mesenchymal Stem Cells	Cells	Stem Cells	3.00E-02	<0.20	SRR (2.3),**GLIS1 (2.3)**,PEAK1 (2.2)

Supplementary Table 4. The 6 enriched gene sets (FDR < 0.05) where *GLIS1* was found to be the best-ranked gene by i-GSEA4GWAS. Enrichment significance (p-value and FDR) of pathways / gene sets are displayed. Genes where SNPs are associated with MVP (P-value < 0.01 from the GWAS meta-analysis) are listed from the most to the less significant.

Gene Set Name	P-value	FDR	Significant genes in the pathway/gene set
GO: Positive Regulation of Transcription DNA Dependent	0.001	0.001	**GLIS1**; RUNX1; SMAD3; MKL2; SQSTM1; NUFIP1; NRIP1; PAX8; GLIS3; TBX5; CREB5; ESRRG; MDFIC; RXRA; MYO6; ARNTL
GO: Positive Regulation of RNA Metabolic Process	0.001	0.002	GLIS1; RUNX1; SMAD3; MKL2; SQSTM1; NUFIP1; NRIP1; PAX8; GLIS3; TBX5; CREB5; ESRRG; MDFIC; RXRA; MYO6; ARNTL
GO: Positive Regulation of Transcription	0.001	0.003	**GLIS1**; RUNX1; SMAD3; HIVEP3; MKL2; IRF4; SQSTM1; NUFIP1; NRIP1; PAX8; GLIS3; TBX5; CREB5; ESRRG; MDFIC; RXRA; NFATC2; MYO6; ARNTL; GLIS2
GO: Positive Regulation of Nucleobasenucleosidenucleotide and Nucleic Acid Metabolic Process	0.001	0.004	**GLIS1**; RUNX1; SMAD3; HIVEP3; MKL2; IRF4; SQSTM1; NUFIP1; NRIP1; PAX8; GLIS3; TBX5; CREB5; ESRRG; MDFIC; ABCA1; RXRA; NFATC2; MYO6; ARNTL; GLIS2
GO: Positive Regulation of Transcription From RNA Polymerase Ii Promoter	0.001	0.009	**GLIS1**; RUNX1; SMAD3; MKL2; SQSTM1; NUFIP1; NRIP1; GLIS3; RXRA; MYO6; ARNTL
GO: Negative Regulation of Transcription	0.002	0.011	**GLIS1**; SMAD3; NFX1; MECP2; NRIP1; TGIF1; GLIS3; BCOR; VDR; TIMELESS; RFX3; ARHGAP35; SUDS3; GLIS2
GO: Negative Regulation of RNA Metabolic Process	0.023	0.089	**GLIS1**; SMAD3; NFX1; MECP2; NRIP1; TGIF1; GLIS3; BCOR
GO: Specific RNA Polymerase Ii Transcription Factor Activity	0.044	0.119	**GLIS1**; GLIS3; SOX9; RXRA

Supplementary Table 5. Functional annotation of 111 SNPs at the GLIS1 locus. We provided functional annotations for all nominally associated (P-value < 0.05) SNPs located within ±500Kb and in linkage desequillibrium (r2 > 0.5) with the top associated SNP rs1879743 at the GLIS1 locus. We used 111 SNPs, all were located in intron X as input to Variant Annotation Integrator (VAI), a tool box avialable at the UCSC Genome Browser to annotate SNPs. We provide location in DNase hypersensitive region (DNASEV), cell types and descriptive and the karyotype of cells, in addition to location at transcription factor binding site (TFBS) using ENCODE data. When available, we indicated if SNPs are expression quantitative trait loci (eQTLs) in atrial appendage tissue using GTeX data.

SNP	Position (bp on Chr1)	DNASEV in cell types	Cell types name for DNASEV	Detail of cell types for DNASEV	Karyotype of DNASEV	TFBS	LD with rs1879734 (r^2)	eQTL. P-value	MVP GWAS P-value
rs28743002	54125615	1	Epithelium	Epithelium cell from a lung carcinoma	Cancer		0.69	-	7.85E-05
rs17383325	54126552	2	Skin	Fetal buttock fibroblast	Normal		0.73	-	6.50E-05
rs12026298	54126754						0.73	-	7.47E-05
rs11206187	54127969	3	Lung	Fetal lung fibroblast	Normal	CTCF	0.73	-	5.24E-05
rs2950241	54129591	20	Blood	Lymphoblastoid	Normal	NR3C1	0.95	0.009	1.79E-05
rs12405835	54129732						0.73	0.039	4.65E-05
rs3006905	54130119						0.95	0.033	1.38E-05
rs4927025	54131753	4	Skin	Toe fibroblast	Normal		0.77	0.024	2.72E-05
rs6658469	54133345						0.77	0.046	3.00E-05
rs67319173	54133723						0.77	0.024	2.69E-05
rs1465037	54135078						0.77	0.025	2.67E-05
rs1554750	54135334	30	Brain	Gliblastoma	Cancer	TCF7L2, BHLHE40, USF1	0.77	0.024	2.66E-05
rs2141080	54138489						1.00	-	6.83E-06
rs2141081	54138552	1	Epithelium	Epithelium cell from a lung carcinoma	Cancer		0.91	-	6.84E-06
rs2141082	54138698						1.00	-	8.79E-06
rs1879734	54138854					GATA2	1.00	0.048	6.80E-06

rs4927026	54140843	2	Skin	Fetal buttock fibroblast	Normal		0.73	-	5.35E-05
rs4927027	54141274	3	Lung	Fetal lung fibroblast	Normal		1.00	-	9.52E-06
rs7543166	54143423						1.00	-	1.24E-05
rs2950244	54143759						0.95	-	1.16E-05
rs55888542	54144297	10	Brain	Neuroblastoma	Cancer		1.00	-	1.21E-05
rs7549345	54145288	15	Blood	Acute megakaryocytic leukemia cell	Cancer		0.65	0.046	3.10E-05
rs2948044	54149577						0.50	0.022	6.73E-04
rs7518284	54150890	68	Epithelium	Renal epithelial cell	Normal	EZH2, SUZ12, ZNF263, TCF7L2, KAP1	0.77	0.045	3.64E-05
rs3006882	54152487						0.95	-	2.53E-05
rs2948042	54152703						0.95	-	2.11E-05
rs882879	54154726					CTCF	0.95	-	2.39E-05
rs3013771	54155902	14	Blood	Chronic lymphocytic leukemia cell	Cancer	JUND, NR2F2, GATA2, MYC, POLR2A, TEAD4, CCNT2, RCOR1, TAL1	1.00	-	2.60E-05
rs3013770	54155959					JUND, MYC, POLR2A, TEAD4, CCNT2	0.86	-	4.66E-05
rs3013769	54156323	10	Brain	Neuroblastoma	Cancer	REST	0.95	-	2.64E-05
rs3013768	54156434	10	Brain	Neuroblastoma	Cancer		0.95	-	2.65E-05
rs2948043	54156443	10	Brain	Neuroblastoma	Cancer		0.95	-	2.65E-05
rs12082358	54156624						0.73	-	3.40E-05

rs3013767	54156757						0.95	-	2.44E-05
rs72660653	54157283	9	Blood vessel	Aortic smooth muscle cells	Normal		0.73	-	3.94E-05
rs72660654	54157287	9	Blood vessel	Aortic smooth muscle cells	Normal		0.73	-	3.94E-05
rs3013766	54157498	2	Skin	Fetal buttock fibroblast	Normal		0.82	-	2.97E-05
rs12086451	54157640						0.73	0.047	4.04E-05
rs12076192	54157742						0.68	0.041	4.10E-05
rs11206191	54158102	2	Skin	Fetal buttock fibroblast	Normal		0.69	0.041	4.20E-05
rs7555617	54158155	2	Skin	Fetal buttock fibroblast	Normal		0.73	-	3.82E-05
rs7555711	54158220						0.73	0.041	4.23E-05
rs7555908	54158462						0.73	0.041	4.28E-05
rs7544346	54158475						0.73		3.89E-05
rs11206192	54158690						0.73	0.041	4.29E-05
rs11206193	54158762						0.65		4.61E-05
rs6658245	54159005					CTCF	0.73	0.041	4.20E-05
rs6691833	54159079	17	Fetal membrane	Chorion cells	Normal	CTCF	0.73	0.046	4.20E-05
rs6694399	54159198	17	Fetal membrane	Chorion cells	Normal	CTCF	0.73	0.041	4.16E-05
rs56368850	54159441						0.73		5.32E-05
rs10888803	54160235						0.73	0.041	4.28E-05
rs12044337	54160558	1	Epithelium	Epithelium cell from a lung carcinoma	Cancer		0.73	0.041	3.74E-04
rs10888804	54160570	1	Epithelium	Epithelium cell from a lung carcinoma	Cancer		0.73	0.041	3.79E-04
rs10888805	54160571	1	Epithelium	Epithelium cell from a lung carcinoma	Cancer		0.73	0.041	3.80E-04
rs61184814	54160709	6	Skin	Skin fibroblasts	Normal		0.73	0.041	4.31E-05
rs6673378	54160812	6	Skin	Skin fibroblasts	Normal		0.69		4.32E-05
rs6665257	54161087						0.73	-	4.09E-05

rs6695509	54161112						0.69	0.041	4.34E-05
rs4927028	54161545	8	Blood vessel	Aortic adventitial fibroblast cells	Normal		0.73	0.041	4.37E-05
rs6661728	54162471						0.73	0.041	5.09E-05
rs59020945	54162635						0.73	-	4.93E-05
rs7554633	54163128	3	Lung	Fetal lung fibroblast	Normal		0.65	0.041	4.89E-05
rs7512093	54163207	3	Lung	Fetal lung fibroblast	Normal		0.73	0.041	4.89E-05
rs7519965	54163251						0.73	0.042	4.59E-05
rs11206195	54164156	45	Epithelium	Esophageal epithelial cells	Normal		0.73	0.041	2.55E-04
rs12041221	54164165	45	Epithelium	Esophageal epithelial cells	Normal		0.73	0.028	3.78E-05
rs1615037	54165541						0.91	-	3.16E-05
rs11206196	54165624						0.73	0.041	4.48E-05
rs11206197	54165831						0.73	0.041	5.09E-05
rs11206198	54165910						0.73	0.028	4.90E-05
rs6588486	54166777						0.73	0.028	4.77E-05
rs6588487	54166790						0.73	0.028	4.35E-05
rs11206200	54168545						0.73	0.016	5.00E-05
rs12040081	54168609	1	Epithelium	Epithelium cell from a lung carcinoma	Cancer		0.69	0.028	4.97E-05
rs4927029	54169455						0.73	0.042	4.57E-05
rs12097598	54170801						0.69	0.028	4.99E-05
rs11206201	54170963	5	Gingival	Gum tissue fibroblast	Normal		0.69	0.028	3.48E-05
rs55786134	54171981	4	Skin	Toe fibroblast	Normal		0.69	0.028	1.04E-04
rs17109178	54173551						0.64	0.049	2.36E-05
rs11206202	54174607						0.64	0.034	2.79E-05
rs2694594	54174846	1	Epithelium	Epithelium cell from a lung carcinoma tissue	Cancer		0.74	-	3.58E-05
rs1780392	54175302						0.82	-	3.56E-05

rs12025857	54175360						0.64	0.034	2.78E-05
rs17109200	54175548	3	Lung	Fetal lung fibroblast	Normal		0.64	0.050	2.54E-05
rs12043690	54175629	3	Lung	Fetal lung fibroblast	Normal		0.64	0.034	2.78E-05
rs1879731	54175861						0.52	0.034	2.78E-05
rs1879732	54175909						0.64	0.034	2.78E-05
rs1879733	54175971						0.64	0.034	2.78E-05
rs7539908	54177279						0.73	-	7.54E-05
rs1780391	54177708	1	Epithelium	Epithelium cell from a lung carcinoma	Cancer		0.82	-	3.57E-05
rs12080903	54178270	9	Blood vessel	Aortic smooth muscle cells	Normal		0.64	0.050	2.59E-05
rs12080993	54178501						0.64	0.050	2.59E-05
rs2186041	54180287						0.64	0.042	2.76E-05
rs1614395	54181668	3	Lung	Fetal lung fibroblast	Normal		0.78	-	3.87E-05
rs1526911	54181942						0.64	-	2.65E-05
rs1526910	54182174						0.64	-	2.86E-05
rs56087009	54182502						0.59	-	2.69E-05
rs4927031	54183191						0.64	-	2.95E-05
rs72660679	54183531					GATA2	0.73	-	9.36E-05
rs12043520	54184097					CTCF	0.64	-	3.03E-05
rs4927032	54186489						0.73	-	3.07E-05
rs17388507	54187575						0.73	-	2.88E-04
rs11206211	54188096	18	Skin	Fibroblast from parkinson's disease	Normal		0.64	-	1.03E-04
rs6658137	54188994						0.64	-	1.11E-04
rs12059719	54190333						0.73	-	1.22E-04
rs12091931	54191103	61	Epithelium	Non-pigment ciliary epithelial cells	Normal	JUND, MAFF, MAFK,	0.64	0.036	1.32E-04

						UBTF, MAZ, CTCF, ZBTB7A, TBP			
rs72660683	54191811						0.73	-	3.11E-04
rs7554636	54193023						0.60	-	1.29E-04
rs910298	54193701	6	Skin	Skin fibroblasts	Normal		0.73	-	3.76E-04
rs702491	54194992	4	Skin	Toe fibroblast	Normal	CTCF	0.51	-	4.27E-03
rs12406120	54195034	4	Skin	Toe fibroblast	Normal		0.61	0.032	4.60E-03

Appendix B

The genetic study of sporadic MVP by GWAS

In chapter V, we performed an updated GWAS using HRC panel. We have identified 6 novel risk loci for MVP. We did functional annotation to the most significant associated SNP and SNPs in strong LD with them. Then, we checked whether the identified associated loci are interacted with other regions in the same chromosomes using FUMA. If there are chromatin interactions, genes are shown when SNPs are mapped to genes whose promoter regions overlap with interactions region. Finally, the genes that have P-value $< 10^{-4}$ in the gene-based association test using MAGMA are presented.

Supplementary Table 1. Functional annotation of variants in the genomic loci associated with MVP

Note: Associated genomic loci were defined by all SNPs in LD (r2>0.6) with an independent genome-wide significant (Bonferroni-corrected P < 1e-5) SNP in the sample-size weighted meta-analysis. CADD: Combined Annotation Dependent Depletion score; RDB: Regulome DB score; DB: GETx version; Tissue: studied tissue for eQTLs; tested allele: tested allele for for eQTLs; symble: the SNP is eQTL for the specific gene; Coloumn "DB, Tissue, Gene, tested allele, symble" have been replecated for multiple times that the SNP is eQTL in several tissue/cell types of diverse genes. Table sorted by gwasP and POS_B37

chr	SNP	POS_B38	non_effect_allele	effect_allele	GWAS_P	r2	IndSigSNP	nearestGene	func	CADD	RDB	eQTLs: DB(Tissue,gene)
8	rs56028519	10341263	G	A	7.97E-08	1.00	rs56028519	MSRA	intronic	0.02	7	GTEx_v7(Artery_Aorta,RP11-981G7.6); GTEx_v7(Artery_Tibial,RP11-981G7.6);
8	rs78821248	10340964	A	G	1.33E-07	0.98	rs56028519	MSRA	intronic	2.30	5	GTEx_v7(Artery_Tibial,RP11-981G7.6);
8	rs147818026	10340458	T	C	1.39E-07	0.98	rs56028519	MSRA	intronic	1.14	7	GTEx_v7(Artery_Tibial,RP11-981G7.6);
8	rs55656427	10336073	C	G	1.98E-07	0.97	rs56028519	MSRA	intronic	0.13	5	GTEx_v7(Artery_Tibial,RP11-981G7.6);

8	rs74699996	10340129	T	C	2.01E-07	0.98	rs56028519	MSRA	intronic	7.45	5	GTEx_v7(Artery_Tibial,RP11-981G7.6);
8	rs56141232	10338908	A	C	2.05E-07	0.96	rs56028519	MSRA	intronic	NA	7	GTEx_v7(Artery_Tibial,RP11-981G7.6);
8	rs10903325	10343604	T	G	2.23E-07	0.96	rs56028519	MSRA	intronic	6.31	4	GTEx_v7(Artery_Tibial,RP11-981G7.6);
8	rs11249991	10343376	A	G	2.32E-07	0.95	rs56028519	MSRA	intronic	9.01	2b	GTEx_v7(Artery_Tibial,RP11-981G7.6);
8	rs145292201	10337406	T	TC	2.85E-07	0.97	rs56028519	MSRA:RP11-212F11.1	ncRNA_intronic	1.18	NA	GTEx_v7(Artery_Tibial,RP11-981G7.6);
8	rs144415113	10337407	G	C	2.93E-07	0.97	rs56028519	MSRA:RP11-212F11.1	ncRNA_intronic	1.72	7	GTEx_v7(Artery_Tibial,RP11-981G7.6);
8	rs11777018	10342572	A	G	8.20E-07	0.88	rs56028519	MSRA	intronic	1.64	5	GTEx_v7(Artery_Tibial,RP11-981G7.6); GTEx_v7(Heart_Atrial_Appendage,AF131215.2); GTEx_v7(Heart_Atrial_Appendage,AF131215.9); GTEx_v6(Heart_Left_Ventricle,AF131215.9);
8	rs35886914	10346048	G	GT	9.56E-07	0.86	rs56028519	MSRA	intronic	6.50	NA	GTEx_v7(Artery_Tibial,RP11-981G7.6); GTEx_v7(Heart_Atrial_Appendage,AF131215.2); GTEx_v7(Heart_Atrial_Appendage,AF131215.9); GTEx_v6(Heart_Left_Ventricle,AF131215.9);
8	rs11785954	10344505	T	C	9.81E-07	0.87	rs56028519	MSRA	intronic	0.57	7	GTEx_v7(Artery_Tibial,RP11-981G7.6); GTEx_v6(Heart_Left_Ventricle,AF131215.9);
8	rs11778030	10343854	T	G	9.85E-07	0.86	rs56028519	MSRA	intronic	0.58	5	GTEx_v7(Artery_Tibial,RP11-981G7.6); GTEx_v7(Heart_Atrial_Appendage,AF131215.2); GTEx_v6(Artery_Aorta,AF131215.9); GTEx_v6(Heart_Left_Ventricle,AF131215.9);
8	rs4841326	10347052	T	C	1.11E-06	0.87	rs56028519	MSRA	intronic	2.65	7	GTEx_v7(Artery_Tibial,RP11-981G7.6); GTEx_v7(Heart_Atrial_Appendage,AF131215.2); GTEx_v7(Heart_Atrial_Appendage,AF131215.9); GTEx_v6(Heart_Left_Ventricle,AF131215.9);
8	rs4840479	10347118	A	G	1.13E-06	0.87	rs56028519	MSRA	intronic	3.11	7	GTEx_v7(Artery_Tibial,RP11-981G7.6); GTEx_v7(Heart_Atrial_Appendage,AF131215.2); GTEx_v7(Heart_Atrial_Appendage,AF131215.9);

												GTEx_v6(Heart_Left_Ventricle,AF131215.9);
8	rs4841325	10346877	T	C	1.33E-06	0.86	rs56028519	MSRA	intronic	3.05	7	GTEx_v7(Artery_Tibial,RP11-981G7.6); GTEx_v7(Heart_Atrial_Appendage,AF131215.2); GTEx_v6(Artery_Aorta,AF131215.9); GTEx_v6(Heart_Left_Ventricle,AF131215.9);
8	rs11781529	10334956	T	C	1.5E-06	0.86	rs56028519	MSRA	intronic	7.32	2b	GTEx_v7(Artery_Tibial,RP11-981G7.6); GTEx_v7(Heart_Atrial_Appendage,AF131215.9); GTEx_v6(Artery_Tibial,RP11-981G7.6);
8	rs6989061	10349779	A	G	2.15E-06	0.90	rs56028519	MSRA	intronic	0.26	5	GTEx_v7(Artery_Tibial,RP11-981G7.6);
8	rs71516569	10348644	G	A	2.22E-06	0.87	rs56028519	MSRA	intronic	6.25	6	GTEx_v7(Artery_Aorta,RP11-981G7.6); GTEx_v7(Artery_Tibial,RP11-981G7.6); GTEx_v7(Heart_Atrial_Appendage,AF131215.2); GTEx_v7(Heart_Atrial_Appendage,AF131215.9);
8	rs36105340	10335010	C	CT	2.26E-06	0.87	rs56028519	MSRA	intronic	6.77	NA	GTEx_v7(Artery_Tibial,RP11-981G7.6);
8	rs17767108	10417378	T	C	1.37E-05	0.63	rs56028519	MSRA	intronic	6.14	4	GTEx_v7(Artery_Tibial,RP11-981G7.6); GTEx_v7(Heart_Atrial_Appendage,AF131215.2); GTEx_v7(Heart_Atrial_Appendage,AF131215.9); GTEx_v7(Heart_Atrial_Appendage,RP11-981G7.6);
8	rs10091197	10340290	C	G	0.000308	0.71	rs56028519	MSRA	intronic	3.15	7	no
8	rs6601441	10340958	G	T	0.000352	0.71	rs56028519	MSRA	intronic	0.95	5	no
11	rs10895596	104161315	A	G	5.84E-06	1.00	rs10895596	PDGFD	intronic	4.09	7	no
11	rs7943011	104161432	C	T	5.97E-06	1.00	rs10895596	PDGFD	intronic	2.53	7	no
11	rs7950273	104160870	G	C	6.99E-06	0.99	rs10895596	PDGFD	intronic	3.25	7	no
11	rs1483506	104161749	G	A	7.05E-06	1.00	rs10895596	PDGFD	intronic	NA	6	no
11	rs4237610	104163081	G	A	7.06E-06	1.00	rs10895596	PDGFD	intronic	4.07	5	no
11	rs1483507	104161830	C	T	7.08E-06	1.00	rs10895596	PDGFD	intronic	0.24	5	no
11	rs7115797	104158149	A	G	7.87E-06	1.00	rs10895596	PDGFD	intronic	0.69	7	no
11	rs1456762	101023356	A	G	7.26E-06	1.00	rs1456762	PGR	intergenic	0.36	NA	no
11	rs4754719	101022878	T	C	7.28E-06	0.99	rs1456762	PGR	intergenic	1.31	5	no

11	rs4754718	101022867	A	C	7.28E-06	0.99	rs1456762	PGR	intergenic	1.43	5	no
11	rs4478959	101023695	G	A	7.29E-06	0.99	rs1456762	PGR	intergenic	1.64	7	no
11	rs11224557	101029578	C	T	7.97E-06	0.99	rs1456762	PGR	downstream	0.50	7	no
11	rs7115959	101026905	C	T	8.12E-06	0.99	rs1456762	PGR	intergenic	10.36	6	no
11	rs7126922	101026565	A	C	8.28E-06	0.97	rs1456762	PGR	intergenic	NA	5	no
11	rs7947125	101030102	C	A	8.4E-06	0.99	rs1456762	PGR	UTR3	7.72	5	no
11	rs34060113	101029499	C	CA	8.4E-06	0.99	rs1456762	PGR	downstream	8.47	NA	no
11	rs10895050	101024724	T	C	8.42E-06	0.99	rs1456762	PGR	intergenic	3.82	7	no
11	rs10895051	101024929	C	T	8.45E-06	0.99	rs1456762	PGR	intergenic	7.79	6	no
11	rs4753995	101027449	G	A	8.48E-06	0.99	rs1456762	PGR	intergenic	1.71	7	no
11	rs12281111	101024454	T	C	8.48E-06	0.99	rs1456762	PGR	intergenic	1.08	6	no
11	rs7939025	101028083	C	A	8.62E-06	0.99	rs1456762	PGR	intergenic	0.13	6	no
11	rs1046982	101029985	C	T	8.7E-06	0.99	rs1456762	PGR	UTR3	1.55	NA	no
11	rs11224554	101027283	T	C	8.73E-06	0.99	rs1456762	PGR	intergenic	2.56	6	no
11	rs4753994	101027415	G	A	8.73E-06	0.99	rs1456762	PGR	intergenic	0.11	6	no
11	rs4753993	101027330	C	T	8.73E-06	0.99	rs1456762	PGR	intergenic	0.98	6	no
11	rs11224552	101027177	T	C	8.73E-06	0.99	rs1456762	PGR	intergenic	3.57	6	no
11	rs4753996	101027654	A	C	8.74E-06	0.99	rs1456762	PGR	intergenic	0.23	7	no
11	rs1824125	101031322	A	C	8.86E-06	0.99	rs1456762	PGR	UTR3	5.46	7	no
11	rs11224560	101032614	T	C	2.68E-05	0.91	rs1456762	PGR	UTR3	NA	7	no
11	rs3740751	101030868	A	G	2.7E-05	0.84	rs1456762	PGR	UTR3	1.62	2c	no
11	rs201207801	101028696	G	A	2.9E-05	0.84	rs1456762	PGR	downstream	7.95	5	no
11	rs34552164	101018163	TG	T	3.36E-05	0.91	rs1456762	PGR	intergenic	1.70	NA	no
11	rs1379132	101017901	G	A	3.37E-05	0.90	rs1456762	PGR	intergenic	1.39	7	no
11	rs35296941	101018158	G	T	3.37E-05	0.91	rs1456762	PGR	intergenic	4.38	6	no
11	rs2186787	101018197	A	G	3.39E-05	0.91	rs1456762	PGR	intergenic	10.12	6	no

11	rs1379133	101017908	G	C	3.4E-05	0.90	rs1456762	PGR	intergenic	2.34	7	no
11	rs143742902	101028699	G	GAA	3.51E-05	0.82	rs1456762	PGR	downstream	3.44	NA	no
11	rs11224558	101030247	T	C	3.57E-05	0.92	rs1456762	PGR	UTR3	1.85	5	no
11	rs2212999	101020845	C	T	3.58E-05	0.91	rs1456762	PGR	intergenic	0.20	NA	no
11	rs11224548	101021103	G	T	3.59E-05	0.92	rs1456762	PGR	intergenic	3.40	6	no
11	rs4753992	101019850	C	T	3.64E-05	0.92	rs1456762	PGR	intergenic	0.97	7	no
11	rs4754717	101022763	G	A	3.69E-05	0.93	rs1456762	PGR	intergenic	0.59	5	no
11	rs10895047	101019581	A	G	3.7E-05	0.91	rs1456762	PGR	intergenic	0.99	7	no
11	rs10895046	101019569	T	A	3.84E-05	0.91	rs1456762	PGR	intergenic	2.16	7	no
11	rs11224550	101021805	T	C	3.84E-05	0.92	rs1456762	PGR	intergenic	0.18	6	no
11	rs12222890	101019234	G	A	3.87E-05	0.91	rs1456762	PGR	intergenic	NA	6	no
11	rs12223660	101020972	G	A	4.05E-05	0.92	rs1456762	PGR	intergenic	2.55	6	no
11	rs10895048	101019666	C	G	4.05E-05	0.92	rs1456762	PGR	intergenic	0.45	7	no
11	rs10895053	101033504	T	C	4.14E-05	0.92	rs1456762	PGR	UTR3	2.86	7	no
11	rs11224549	101021221	A	T	4.17E-05	0.92	rs1456762	PGR	intergenic	4.94	6	no
11	rs10501970	101020350	G	A	4.17E-05	0.92	rs1456762	PGR	intergenic	2.47	6	no
11	rs2219337	101020514	G	A	4.18E-05	0.92	rs1456762	PGR	intergenic	6.27	NA	no
11	rs4753991	101019734	C	A	4.2E-05	0.92	rs1456762	PGR	intergenic	0.09	6	no
11	rs4753990	101019598	T	G	4.21E-05	0.91	rs1456762	PGR	intergenic	0.77	7	no
11	rs11224546	101020262	G	A	4.28E-05	0.92	rs1456762	PGR	intergenic	2.04	7	no
11	rs11224556	101028882	G	A	4.51E-05	0.91	rs1456762	PGR	downstream	7.78	5	no
11	rs17093553	101026921	G	A	4.96E-05	0.92	rs1456762	PGR	intergenic	3.28	7	no
11	rs10895040	101009587	A	T	5.31E-05	0.92	rs1456762	TMEM133	intergenic	7.53	5	no
11	rs1840768	101008176	A	G	9.42E-05	0.79	rs1456762	TMEM133	intergenic	8.99	NA	no
11	rs11224561	101034325	T	C	0.001959	0.84	rs1456762	PGR	UTR3	1.43	2b	no
12	rs4762110	65636947	C	T	5.51E-06	1.00	rs4762110	RP11-230G5.2	ncRNA_intronic	3.80	7	no

12	rs2254518	65635099	C	T	5.87E-06	1.00	rs4762110	RP11-230G5.2	ncRNA_intronic	2.19	2b	no
12	rs2453267	65637932	A	T	6.14E-06	1.00	rs4762110	RP11-230G5.2	ncRNA_intronic	0.99	3a	no
12	rs11175857	65671356	T	A	6.95E-06	0.84	rs4762110	RP11-221N13.4	intergenic	0.32	5	no
12	rs7968812	65633435	T	C	8.19E-06	0.85	rs4762110	RP11-230G5.2	ncRNA_intronic	3.34	7	no
12	rs73128080	65669187	G	A	9.59E-06	0.82	rs4762110	RP11-221N13.4	intergenic	0.26	4	no
12	rs2453269	65659912	A	G	1.19E-05	0.83	rs4762110	RP11-221N13.4	ncRNA_intronic	2.71	NA	no
12	rs4762108	65633034	C	T	1.25E-05	0.83	rs4762110	RP11-230G5.2	ncRNA_intronic	3.71	5	no
12	rs7296711	65626680	T	C	1.29E-05	0.83	rs4762110	RP11-230G5.2	ncRNA_exonic	1.37	7	no
12	rs4762107	65630579	A	G	1.34E-05	0.83	rs4762110	RP11-230G5.2	ncRNA_intronic	18.56	5	no
12	rs7306171	65632267	G	C	1.4E-05	0.83	rs4762110	RP11-230G5.2	ncRNA_intronic	2.27	4	no
12	rs7305981	65632367	A	G	1.48E-05	0.83	rs4762110	RP11-230G5.2	ncRNA_intronic	4.34	4	no
12	rs4474486	65637358	G	A	1.49E-05	0.83	rs4762110	RP11-230G5.2	ncRNA_intronic	4.53	5	no
12	rs4365098	65622310	G	C	1.52E-05	0.83	rs4762110	RP11-230G5.2	ncRNA_intronic	1.33	3a	no
12	rs2453270	65659486	C	T	1.52E-05	0.85	rs4762110	RP11-221N13.4	ncRNA_intronic	3.13	NA	no
12	rs2731502	65607982	G	A	1.54E-05	0.83	rs4762110	RP11-230G5.2:RP11-221N13.3	ncRNA_intronic	3.89	4	no
12	rs3923635	65627294	T	C	1.58E-05	0.83	rs4762110	RP11-230G5.2	ncRNA_intronic	1.51	NA	no
12	rs4762111	65638891	C	T	1.62E-05	0.85	rs4762110	RP11-230G5.2	ncRNA_intronic	2.06	7	no
12	rs4575324	65615117	C	T	1.63E-05	0.83	rs4762110	RP11-230G5.2	ncRNA_intron	4.92	5	no

									ic			
12	rs141630124	65646804	A	ATCTT	1.67E-05	0.83	rs4762110	RP11-221N13.4	ncRNA_intronic	1.26	NA	no
12	rs10748021	65610719	A	C	1.7E-05	0.83	rs4762110	RP11-230G5.2:RP11-221N13.3	ncRNA_intronic	NA	6	no
12	rs2468386	65613834	C	G	0.000017	0.83	rs4762110	RP11-230G5.2	ncRNA_intronic	6.66	2c	no
12	rs2468387	65613359	C	T	1.72E-05	0.83	rs4762110	RP11-230G5.2	ncRNA_intronic	4.58	NA	no
12	rs2453266	65610024	G	A	1.81E-05	0.83	rs4762110	RP11-230G5.2:RP11-221N13.3	ncRNA_intronic	0.25	7	no
12	rs4508249	65603333	C	T	1.92E-05	0.83	rs4762110	RP11-230G5.2:RP11-221N13.3	ncRNA_intronic	5.93	4	no
12	rs10878303	65604197	T	C	2.04E-05	0.83	rs4762110	RP11-230G5.2:RP11-221N13.3	ncRNA_intronic	3.24	3a	no
12	rs2453268	65664765	A	T	2.16E-05	0.82	rs4762110	RP11-221N13.4	intergenic	0.01	6	no
12	rs7300688	65636663	A	G	2.26E-05	0.83	rs4762110	RP11-230G5.2	ncRNA_intronic	3.08	5	no
12	rs4762109	65636071	T	C	2.46E-05	0.83	rs4762110	RP11-230G5.2	ncRNA_intronic	0.19	6	no
12	rs2246251	65657381	A	T	2.46E-05	0.83	rs4762110	RP11-221N13.4	ncRNA_intronic	4.89	3a	no
12	rs2453271	65656571	T	A	2.5E-05	0.83	rs4762110	RP11-221N13.4	ncRNA_intronic	NA	5	no
12	rs4499051	65627132	T	C	2.57E-05	0.83	rs4762110	RP11-230G5.2	ncRNA_intronic	2.49	6	no
12	rs7979521	65657768	A	T	2.58E-05	0.83	rs4762110	RP11-221N13.4	ncRNA_intronic	1.02	4	no
12	rs34479637	65645703	TA	T	2.83E-05	0.83	rs4762110	RP11-221N13.4	ncRNA_intronic	NA	NA	no

12	rs10784483	65599936	G	A	3.43E-05	0.82	rs4762110	RP11-230G5.2	ncRNA_intronic	9.49	5	no
12	rs7296059	65596105	T	G	3.45E-05	0.82	rs4762110	RP11-230G5.2	ncRNA_intronic	NA	6	no
12	rs10784482	65598186	T	C	3.47E-05	0.82	rs4762110	RP11-230G5.2	ncRNA_intronic	0.15	6	no
12	rs4489796	65598621	C	T	3.5E-05	0.82	rs4762110	RP11-230G5.2	ncRNA_intronic	1.14	7	no
12	rs11175842	65633776	A	C	3.51E-05	0.87	rs4762110	RP11-230G5.2	ncRNA_intronic	1.14	6	no
12	rs4508248	65600777	T	C	3.51E-05	0.82	rs4762110	RP11-230G5.2	ncRNA_intronic	2.86	5	no
12	rs12302900	65601625	C	G	3.55E-05	0.82	rs4762110	RP11-230G5.2	ncRNA_intronic	NA	6	no
12	rs57541501	65623082	A	AG	4.33E-05	0.82	rs4762110	RP11-230G5.2	ncRNA_intronic	0.49	NA	no
12	rs56051830	65667644	A	ACT	4.47E-05	0.82	rs4762110	RP11-221N13.4	intergenic	0.39	NA	no
12	rs2468382	65634717	T	C	4.57E-05	0.87	rs4762110	RP11-230G5.2	ncRNA_intronic	0.90	7	no
12	rs4237905	65628548	G	T	5.29E-05	0.85	rs4762110	RP11-230G5.2	ncRNA_intronic	1.44	7	no
12	rs4581496	65628670	T	A	5.3E-05	0.85	rs4762110	RP11-230G5.2	ncRNA_intronic	4.01	7	no
12	rs10459262	65628215	G	T	5.43E-05	0.85	rs4762110	RP11-230G5.2	ncRNA_intronic	2.04	5	no
12	rs10784479	65590527	A	T	5.84E-05	0.80	rs4762110	RP11-230G5.2	ncRNA_intronic	6.27	2b	no
12	rs4762102	65591339	A	G	6.07E-05	0.80	rs4762110	RP11-230G5.2	ncRNA_intronic	2.32	7	no
12	rs10784476	65580893	A	G	6.18E-05	0.79	rs4762110	RP11-230G5.2	ncRNA_intronic	1.29	6	no
12	rs2468384	65623681	C	G	6.53E-05	0.85	rs4762110	RP11-230G5.2	ncRNA_intronic	1.88	4	no

12	rs4307765	65593709	A	C	6.55E-05	0.80	rs4762110	RP11-230G5.2	ncRNA_intronic	0.38	5	no
12	rs7301758	65564097	A	G	7.55E-05	0.79	rs4762110	RP11-230G5.2	ncRNA_intronic	12.78	5	no
12	rs4762106	65624693	G	A	7.63E-05	0.85	rs4762110	RP11-230G5.2	ncRNA_intronic	2.83	3a	no
12	rs4556592	65585379	G	C	8.1E-05	0.79	rs4762110	RP11-230G5.2	ncRNA_intronic	0.62	7	GTEx_v7(Heart_Atrial_Appendage,RP11-221N13.3);
12	rs4477481	65593757	T	G	8.18E-05	0.80	rs4762110	RP11-230G5.2	ncRNA_intronic	0.08	5	GTEx_v7(Heart_Atrial_Appendage,RP11-221N13.3);
12	rs10784475	65577783	T	A	0.000082	0.79	rs4762110	RP11-230G5.2	ncRNA_intronic	3.98	6	no
12	rs7308817	65577992	C	T	9.83E-05	0.79	rs4762110	RP11-230G5.2	ncRNA_intronic	0.99	7	no
12	rs4346013	65562677	C	G	0.000104	0.76	rs4762110	RP11-230G5.2	ncRNA_intronic	3.58	4	GTEx_v7(Heart_Atrial_Appendage,RP11-221N13.3);
12	rs7134553	65558011	A	G	0.000118	0.76	rs4762110	RP11-230G5.2	ncRNA_intronic	2.70	4	GTEx_v7(Heart_Atrial_Appendage,RP11-221N13.3);
12	rs4129000	65559754	C	T	0.000118	0.76	rs4762110	RP11-230G5.2:RP11-221N13.2	ncRNA_intronic	1.89	7	GTEx_v7(Heart_Atrial_Appendage,RP11-221N13.3);
12	rs4285921	65558128	T	C	0.000138	0.76	rs4762110	RP11-230G5.2	ncRNA_intronic	7.27	4	GTEx_v7(Heart_Atrial_Appendage,RP11-221N13.3);
12	rs4129001	65559959	G	C	0.00014	0.76	rs4762110	RP11-230G5.2:RP11-221N13.2	ncRNA_intronic	0.18	6	GTEx_v7(Heart_Atrial_Appendage,RP11-221N13.3);
12	rs12424058	65550114	C	A	0.000187	0.75	rs4762110	RP11-230G5.2	ncRNA_intronic	1.82	6	GTEx_v7(Heart_Atrial_Appendage,RP11-221N13.3);
12	rs2252112	65548856	C	G	0.000197	0.75	rs4762110	RP11-230G5.2	ncRNA_intronic	1.22	6	GTEx_v7(Heart_Atrial_Appendage,RP11-221N13.3);
12	rs2252111	65548866	C	T	0.000197	0.75	rs4762110	RP11-230G5.2	ncRNA_intronic	2.90	7	GTEx_v7(Heart_Atrial_Appendage,RP11-221N13.3);
12	rs11175838	65611904	C	T	0.000231	0.71	rs4762110	RP11-230G5.2:RP11-	ncRNA_intronic	1.55	5	no

								221N13.3				
12	rs2250519	65644210	C	T	0.000241	0.62	rs4762110	RP11-221N13.4	upstream	5.59	2b	no
12	rs11504196	65551752	A	T	0.00114	0.64	rs4762110	RP11-230G5.2	ncRNA_intronic	NA	7	no
12	rs9669406	65554301	C	G	0.001265	0.64	rs4762110	RP11-230G5.2	ncRNA_intronic	0.60	5	no
12	rs2251304	65555680	G	A	0.001276	0.64	rs4762110	RP11-230G5.2	ncRNA_intronic	0.10	6	no
12	rs2252122	65548590	A	G	0.001379	0.64	rs4762110	RP11-230G5.2	ncRNA_intronic	1.96	6	no
12	rs2255312	65546259	A	G	0.001556	0.63	rs4762110	RP11-230G5.2	ncRNA_intronic	0.88	7	no
12	rs1494506	65546110	A	G	0.002	0.63	rs4762110	RP11-230G5.2	ncRNA_intronic	3.52	5	no
12	rs2255593	65543332	G	A	0.002042	0.62	rs4762110	RP11-230G5.2	ncRNA_intronic	3.13	7	no
19	rs4802272	45732692	G	A	3.17E-07	1.00	rs4802272	AC074212.3	upstream	0.34	7	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs62111759	45792972	T	A	3.58E-07	0.94	rs4802272	DMWD	upstream	12.93	4	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs7246377	45789001	A	G	3.76E-07	0.94	rs4802272	DMWD	intronic	NA	5	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs10419334	45755735	T	G	5.01E-07	0.95	rs4802272	AC074212.3	intronic	0.09	7	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs10404730	45754100	C	A	5.11E-07	0.95	rs4802272	AC074212.3	intronic	0.56	7	GTEx_v7(Whole_Blood,DMPK);

												GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs635299	45770913	C	A	5.78E-07	0.94	rs4802272	AC074212.6:DMPK	ncRNA_intronic	2.54	NA	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs4802276	45757490	C	T	5.98E-07	0.95	rs4802272	AC074212.3	intronic	6.41	7	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs915915	45771714	C	A	7.14E-07	0.93	rs4802272	AC074212.6:DMPK	ncRNA_exonic	5.82	NA	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs1865116	45791252	A	G	7.79E-07	0.94	rs4802272	DMWD	intronic	6.02	4	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs35732815	45720898	G	A	7.80E-07	0.99	rs4802272	FBXO46	intronic	3.26	4	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs34388571	45722694	A	G	8.17E-07	0.99	rs4802272	FBXO46	intronic	1.40	5	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs522769	45781262	A	G	8.57E-07	0.94	rs4802272	DMPK:AC011530.4	intronic	10.93	5	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs10401487	45748202	C	T	8.70E-07	0.95	rs4802272	AC074212.3	intronic	1.58	7	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD);

19	rs672348	45781076	G	T	8.87E-07	0.94	rs4802272	DMPK:AC011530.4	intronic	12.43	NA	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs617988	45783942	G	C	9.29E-07	0.94	rs4802272	AC011530.4:DMWD	UTR3	6.87	5	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs57947504	45724622	C	T	9.34E-07	0.99	rs4802272	FBXO46	intronic	3.25	7	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs11083781	45751881	T	C	1.07E-06	0.95	rs4802272	AC074212.3	intronic	1.97	7	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs16939	45772798	A	C	1.1E-06	0.94	rs4802272	DMPK	intronic	5.47	4	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs10416076	45751817	A	G	1.11E-06	0.95	rs4802272	AC074212.3	intronic	1.62	7	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs12611358	45748104	A	G	1.17E-06	0.95	rs4802272	AC074212.3	intronic	1.19	7	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs2341096	45758504	G	A	1.17E-06	0.94	rs4802272	AC074212.3	intronic	2.66	7	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs4803854	45774030	A	C	1.2E-06	0.94	rs4802272	DMPK	intronic	2.34	5	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD);

												GTEx_v6(Artery_Tibial,DMWD);
19	rs4802277	45757989	A	G	1.23E-06	0.94	rs4802272	AC074212.3	intronic	0.01	4	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs3745802	45767919	T	G	1.23E-06	0.94	rs4802272	AC074212.5:SIX5: AC074212.6	ncRNA_exonic	8.56	4	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs10401971	45758724	T	C	1.3E-06	0.94	rs4802272	AC074212.3	intronic	0.77	7	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs7260651	45756987	A	G	1.31E-06	0.94	rs4802272	AC074212.3	intronic	0.56	5	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs4802275	45757117	T	C	1.33E-06	0.94	rs4802272	AC074212.3	intronic	5.94	5	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs2041975	45735297	C	T	1.35E-06	0.95	rs4802272	AC074212.3	intronic	3.03	7	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs8112282	45761771	T	C	1.35E-06	0.95	rs4802272	AC074212.3	exonic	11.16	5	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs2014576	45765818	A	G	1.35E-06	0.84	rs4802272	AC074212.5:SIX5	exonic	9.42	NA	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD);

												GTEx_v6(Artery_Tibial,DMWD); GTEx_v6(Muscle_Skeletal,DMWD);
19	rs10422883	45744383	G	A	1.36E-06	0.92	rs4802272	AC074212.3	intronic	3.47	7	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs35226705	45798198	C	A	1.45E-06	0.63	rs4802272	RSPH6A	intronic	0.73	6	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,AC074212.6); GTEx_v6(Muscle_Skeletal,DMWD);
19	rs12610941	45737937	C	T	1.62E-06	0.95	rs4802272	AC074212.3	intronic	1.99	6	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs10415988	45743446	T	C	1.99E-06	0.92	rs4802272	AC074212.3	intronic	3.89	4	GTEx_v7(Whole_Blood,DMPK); GTEx_v7(Artery_Aorta,DMWD); GTEx_v7(Artery_Tibial,DMPK); GTEx_v7(Artery_Tibial,DMWD); GTEx_v6(Artery_Tibial,DMWD);
19	rs547948198	38733096	TG	T	1.05E-06	1.00	rs547948198	CAPN12	intronic	0.78	NA	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12);
19	rs150394753	38753329	T	G	1.17E-06	0.98	rs547948198	CAPN12	intronic	1.49	7	GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs138217270	38753330	T	C	1.25E-06	0.98	rs54794819	CAPN12	intronic	3.18	7	GTEx_v6(Artery_Aorta,CAPN12);

							8					GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs10416718	38756336	A	G	1.52E-06	0.90	rs547948198	CAPN12	intronic	0.71	6	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs55873140	38751919	C	A	1.98E-06	0.97	rs547948198	CAPN12	intronic	1.78	7	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs12978100	38751284	C	T	2.01E-06	0.72	rs547948198	CAPN12	intronic	2.98	5	GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12);

												GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Tibial,CAPN12);
19	rs55945853	38739406	A	G	2.12E-06	0.93	rs547948198	CAPN12:CTD-2540F13.2	ncRNA_exonic	2.55	6	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs62120083	38741341	A	G	2.22E-06	0.93	rs547948198	CAPN12	intronic	3.10	7	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12);

												GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs10407371	38758666	C	T	2.22E-06	0.87	rs547948198	CAPN12	intronic	0.93	7	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs10418728	38756662	C	T	2.98E-06	0.86	rs547948198	CAPN12	intronic	0.58	6	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs11673323	38757546	C	G	2.99E-06	0.86	rs547948198	CAPN12	intronic	0.21	6	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2);

												GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs8107416	38747310	A	C	3.26E-06	0.92	rs547948198	CAPN12	intronic	6.13	5	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs62120087	38749064	C	G	3.26E-06	0.75	rs547948198	CAPN12	intronic	0.16	5	GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs59162138	38677379	A	AACAG	3.63E-06	0.87	rs547948198	ACTN4	intronic	9.20	NA	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4);

												GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs55923191	38749579	T	C	3.74E-06	0.88	rs547948198	CAPN12	intronic	1.53	7	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs62121816	38673824	T	C	3.77E-06	0.86	rs547948198	ACTN4	intronic	1.01	6	GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs9304578	38759697	A	G	3.96E-06	0.91	rs547948198	CAPN12	intronic	4.98	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12);

												GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs3217235	38704666	A	AGGAG	4.3E-06	0.90	rs547948198	ACTN4	intronic	2.26	NA	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs56172450	38749629	C	G	4.44E-06	0.88	rs547948198	CAPN12	intronic	NA	7	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4);

												GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs12462270	38710848	T	C	4.49E-06	0.91	rs547948198	ACTN4	intronic	3.85	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs56113315	38693526	T	C	4.88E-06	0.90	rs547948198	ACTN4:CTB-186G2.4	ncRNA_exonic	6.51	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Muscle_Skeletal,ACTN4);

												GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs2279144	38711141	T	C	4.97E-06	0.91	rs547948198	ACTN4	intronic	7.90	5	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs3786848	38716689	T	C	5.23E-06	0.91	rs547948198	ACTN4	intronic	5.25	5	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs883433	38715648	T	C	5.25E-06	0.91	rs547948198	ACTN4	intronic	14.49	NA	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12);

												GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs3786845	38704443	G	C	5.29E-06	0.90	rs547948198	ACTN4	intronic	0.86	5	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs3745859	38706105	T	C	5.58E-06	0.91	rs547948198	ACTN4	exonic	18.21	5	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12);

												GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs10415646	38689818	T	C	5.79E-06	0.90	rs547948198	ACTN4:CTB-186G2.4	ncRNA_intronic	0.90	7	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs2287729	38704674	G	T	6.03E-06	0.90	rs547948198	ACTN4	intronic	5.94	NA	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12);

19	rs10403540	38714206	A	G	6.09E-06	0.91	rs547948198	ACTN4	intronic	3.45	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs13343794	38688631	A	G	6.13E-06	0.90	rs547948198	ACTN4:CTB-186G2.4	ncRNA_intronic	7.38	7	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs3786851	38724693	T	C	6.16E-06	0.91	rs547948198	ACTN4	intronic	4.13	2b	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4);

												GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs62121814	38669089	T	C	6.48E-06	0.87	rs547948198	ACTN4	intronic	0.82	7	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs11083473	38689294	G	A	6.69E-06	0.72	rs547948198	ACTN4:CTB-186G2.4	ncRNA_intronic	0.07	3a	GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs8111746	38722219	T	C	6.88E-06	0.91	rs54794819	ACTN4	intronic	5.66	4	GTEx_v7(Whole_Blood,CAPN12);

							8					GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs7254100	38682846	C	A	6.93E-06	0.87	rs547948198	ACTN4	intronic	2.71	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs28624276	38712257	A	G	7.16E-06	0.92	rs547948198	ACTN4	intronic	1.32	2b	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12);

												GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs62121818	38677120	C	T	7.17E-06	0.88	rs547948198	ACTN4	intronic	1.55	6	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs883394	38714973	T	C	7.59E-06	0.92	rs547948198	ACTN4	intronic	0.96	5	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4);

												GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs10406778	38679833	G	A	7.85E-06	0.88	rs547948198	ACTN4	intronic	4.40	5	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs3786837	38681402	C	G	7.95E-06	0.95	rs547948198	ACTN4	intronic	0.96	2b	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12);

19	rs979971	38653604	T	C	8.89E-06	0.87	rs547948198	ACTN4	intronic	6.27	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs7252589	38680268	A	G	9.14E-06	0.88	rs547948198	ACTN4	intronic	1.48	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs11338400	38722479	A	AG	9.48E-06	0.92	rs547948198	ACTN4	intronic	0.17	NA	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12);

												GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs28488032	38712248	C	G	9.73E-06	0.92	rs547948198	ACTN4	intronic	0.39	2b	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs1808382	38660394	T	G	9.98E-06	0.87	rs547948198	ACTN4	intronic	0.52	NA	GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4);

												GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs3786835	38665240	A	G	1.04E-05	0.87	rs547948198	ACTN4	intronic	5.94	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs7251903	38694171	A	G	1.07E-05	0.91	rs547948198	ACTN4	intronic	4.53	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-

												2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs13344413	38652649	A	G	1.09E-05	0.94	rs547948198	ACTN4	intronic	6.77	2b	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs28377909	38670798	T	C	1.1E-05	0.73	rs547948198	ACTN4	intronic	0.65	5	GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs16972767	38662404	A	G	1.11E-05	0.87	rs547948198	ACTN4	intronic	3.28	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12);

												GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs10421560	38695973	A	G	1.21E-05	0.91	rs547948198	ACTN4	intronic	1.35	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs3786839	38691925	T	C	1.27E-05	0.92	rs547948198	ACTN4:CTB-186G2.4	ncRNA_intronic	0.06	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2);

												GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs10426809	38696310	A	G	1.34E-05	0.73	rs547948198	ACTN4	intronic	1.68	4	GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs11083475	38697472	A	G	1.37E-05	0.91	rs547948198	ACTN4	intronic	8.35	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs12460849	38653244	A	G	1.45E-05	0.87	rs547948198	ACTN4	intronic	7.20	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12);

												GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs10415219	38659595	G	A	1.56E-05	0.87	rs547948198	ACTN4	intronic	9.55	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs979972	38653527	G	A	2.12E-05	0.88	rs547948198	ACTN4	intronic	3.88	NA	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12);

												GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)
19	rs55876653	38656140	C	G	2.42E-05	0.88	rs547948198	ACTN4	intronic	0.19	4	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs56365761	38657463	G	A	3.25E-05	0.88	rs547948198	ACTN4	intronic	3.53	7	GTEx_v7(Whole_Blood,CAPN12); GTEx_v7(Artery_Aorta,ACTN4); GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Coronary,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,ACTN4); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,ACTN4); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12)

19	19:39239966:CA:CAA	38749326	CA	CAA	3.49E-05	0.80	rs547948198	CAPN12	intronic	NA	NA	no
19	rs11665964	38705688	C	T	4.25E-05	0.61	rs547948198	ACTN4	intronic	6.17	2b	GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Tibial,CAPN12);
19	rs2303040	38647968	C	T	5.13E-05	0.90	rs547948198	ACTN4	intronic	6.79	NA	GTEx_v6(Artery_Aorta,CAPN12); GTEx_v6(Artery_Tibial,CAPN12); GTEx_v6(Heart_Left_Ventricle,CTD-2540F13.2); GTEx_v6(Muscle_Skeletal,ACTN4); GTEx_v6(Muscle_Skeletal,CAPN12);
19	rs755690	38676720	A	G	5.83E-05	0.61	rs547948198	ACTN4	intronic	1.17	NA	GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Tibial,CAPN12);
19	rs749701	38699106	T	C	8.46E-05	0.61	rs547948198	ACTN4	intronic	2.89	1f	GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Tibial,CAPN12);
19	rs11670636	38672803	A	C	8.8E-05	0.61	rs547948198	ACTN4	intronic	0.60	6	GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12);

												GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Tibial,CAPN12);
19	rs10424568	38648308	A	G	0.000165	0.62	rs547948198	ACTN4	intronic	5.87	4	GTEx_v7(Artery_Aorta,CAPN12); GTEx_v7(Artery_Tibial,ACTN4); GTEx_v7(Artery_Tibial,CAPN12); GTEx_v7(Artery_Tibial,CTD-2540F13.2); GTEx_v7(Heart_Atrial_Appendage,CAPN12); GTEx_v7(Heart_Left_Ventricle,CAPN12); GTEx_v7(Muscle_Skeletal,ACTN4); GTEx_v7(Muscle_Skeletal,CAPN12); GTEx_v6(Artery_Tibial,CAPN12);

Supplementary Table 2. Chromatin interaction regions linking the 6 significant GWAS loci from a meta-analysis of MVP to implicated genes.

Note: Interaction regions (Chromosome:Start base - end base) include one region overlapping one or more GWS SNP in an enhancer and one region overlapping a gene promoter region. FDR = false discovery rate. Position in the table is B37.

CHR	region1	region2	FDR	type	DB	tissue/cell	inter/ intra	Genes
8	8:10160001-10200000	8:9240001-9280000	6.94E-09	HiC	GSE87112	Left_Ventricle	intra	nan
8	8:10160001-10200000	8:10080001-10120000	1.41E-30	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
8	8:10160001-10200000	8:10120001-10160000	1.93E-23	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
8	8:10160001-10200000	8:10200001-10240000	5.12E-33	HiC	GSE87112	Mesendoderm	intra	nan
8	8:10160001-10200000	8:10240001-10280000	4.09E-58	HiC	GSE87112	Mesendoderm	intra	nan
8	8:10160001-10200000	8:10280001-10320000	3.42E-12	HiC	GSE87112	Mesendoderm	intra	ENSG00000261451
8	8:10160001-10200000	8:10600001-10640000	1.67E-08	HiC	GSE87112	Mesendoderm	intra	nan
8	8:10160001-10200000	8:9600001-9640000	5.05E-10	HiC	GSE87112	Mesendoderm	intra	nan
8	8:10160001-10200000	8:9640001-9680000	5.29E-08	HiC	GSE87112	Mesendoderm	intra	nan
8	8:10160001-10200000	8:9720001-9760000	1.23E-46	HiC	GSE87112	Mesendoderm	intra	ENSG00000239065
8	8:10160001-10200000	8:9760001-9800000	3.41E-81	HiC	GSE87112	Mesendoderm	intra	ENSG00000253230:ENSG00000238496
8	8:10160001-10200000	8:9800001-9840000	3.29E-144	HiC	GSE87112	Mesendoderm	intra	nan
8	8:10160001-10200000	8:9840001-9880000	8.67E-67	HiC	GSE87112	Mesendoderm	intra	nan
8	8:10160001-10200000	8:9880001-9920000	3.00E-97	HiC	GSE87112	Mesendoderm	intra	ENSG00000260093:ENSG00000175806
8	8:10160001-10200000	8:9920001-9960000	2.75E-11	HiC	GSE87112	Mesendoderm	intra	ENSG00000221320
8	8:10160001-10200000	8:9960001-10000000	5.13E-30	HiC	GSE87112	Mesendoderm	intra	nan
8	8:10160001-10200000	8:10000001-10040000	3.08E-25	HiC	GSE87112	Mesendoderm	intra	nan
8	8:10160001-10200000	8:10080001-10120000	7.79E-30	HiC	GSE87112	Mesendoderm	intra	nan
8	8:10160001-10200000	8:10120001-10160000	1.89E-42	HiC	GSE87112	Mesendoderm	intra	nan
8	8:10160001-10200000	8:10200001-10240000	3.08E-24	HiC	GSE87112	hESC	intra	nan
8	8:10160001-10200000	8:10240001-10280000	3.94E-54	HiC	GSE87112	hESC	intra	nan
8	8:10160001-10200000	8:9600001-9640000	8.15E-10	HiC	GSE87112	hESC	intra	nan
8	8:10160001-10200000	8:9720001-9760000	2.53E-43	HiC	GSE87112	hESC	intra	ENSG00000239065
8	8:10160001-10200000	8:9760001-9800000	1.41E-35	HiC	GSE87112	hESC	intra	ENSG00000253230:ENSG00000238496
8	8:10160001-10200000	8:9800001-9840000	2.61E-103	HiC	GSE87112	hESC	intra	nan

8	8:10160001-10200000	8:9840001-9880000	1.17E-35	HiC	GSE87112	hESC	intra	nan
8	8:10160001-10200000	8:9880001-9920000	6.97E-34	HiC	GSE87112	hESC	intra	ENSG00000260093:ENSG00000175806
8	8:10160001-10200000	8:9960001-10000000	2.03E-17	HiC	GSE87112	hESC	intra	nan
8	8:10160001-10200000	8:10000001-10040000	1.71E-23	HiC	GSE87112	hESC	intra	nan
8	8:10160001-10200000	8:10040001-10080000	8.17E-09	HiC	GSE87112	hESC	intra	nan
8	8:10160001-10200000	8:10000001-10040000	1.41E-21	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
8	8:10160001-10200000	8:9960001-10000000	2.23E-22	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
8	8:10160001-10200000	8:9920001-9960000	7.46E-11	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000221320
8	8:10160001-10200000	8:9880001-9920000	1.13E-118	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000260093:ENSG00000175806
8	8:10160001-10200000	8:9280001-9320000	4.98E-07	HiC	GSE87112	Left_Ventricle	intra	ENSG00000254237:ENSG00000249050
8	8:10160001-10200000	8:9800001-9840000	6.13E-08	HiC	GSE87112	Left_Ventricle	intra	nan
8	8:10160001-10200000	8:9880001-9920000	9.01E-60	HiC	GSE87112	Left_Ventricle	intra	ENSG00000260093:ENSG00000175806
8	8:10160001-10200000	8:9880001-9920000	1.23E-12	HiC	GSE87112	GM12878	intra	ENSG00000260093:ENSG00000175806
8	8:10160001-10200000	8:10200001-10240000	7.27E-33	HiC	GSE87112	IMR90	intra	nan
8	8:10160001-10200000	8:10240001-10280000	1.73E-52	HiC	GSE87112	IMR90	intra	nan
8	8:10160001-10200000	8:10280001-10320000	5.04E-07	HiC	GSE87112	IMR90	intra	ENSG00000261451
8	8:10160001-10200000	8:8880001-8920000	3.67E-12	HiC	GSE87112	IMR90	intra	ENSG00000263407:ENSG00000252543
8	8:10160001-10200000	8:9240001-9280000	1.18E-07	HiC	GSE87112	IMR90	intra	nan
8	8:10160001-10200000	8:9600001-9640000	1.03E-13	HiC	GSE87112	IMR90	intra	nan
8	8:10160001-10200000	8:9680001-9720000	1.11E-10	HiC	GSE87112	IMR90	intra	nan
8	8:10160001-10200000	8:9720001-9760000	3.55E-38	HiC	GSE87112	IMR90	intra	ENSG00000239065
8	8:10160001-10200000	8:9760001-9800000	6.58E-51	HiC	GSE87112	IMR90	intra	ENSG00000253230:ENSG00000238496
8	8:10160001-10200000	8:9800001-9840000	1.26E-73	HiC	GSE87112	IMR90	intra	nan
8	8:10160001-10200000	8:10080001-10120000	3.67E-38	HiC	GSE87112	hESC	intra	nan
8	8:10160001-10200000	8:9840001-9880000	1.33E-47	HiC	GSE87112	IMR90	intra	nan
8	8:10160001-10200000	8:9920001-9960000	1.94E-14	HiC	GSE87112	IMR90	intra	ENSG00000221320
8	8:10160001-10200000	8:9960001-10000000	1.63E-12	HiC	GSE87112	IMR90	intra	nan

8	8:10160001-10200000	8:10000001-10040000	2.50E-20	HiC	GSE87112	IMR90	intra	nan
8	8:10160001-10200000	8:10080001-10120000	7.81E-36	HiC	GSE87112	IMR90	intra	nan
8	8:10160001-10200000	8:10120001-10160000	2.78E-30	HiC	GSE87112	IMR90	intra	nan
8	8:10160001-10200000	8:10200001-10240000	2.85E-13	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
8	8:10160001-10200000	8:10240001-10280000	5.70E-20	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
8	8:10160001-10200000	8:11160001-11200000	6.32E-08	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000246477:ENSG00000177710:ENSG00000154316
8	8:10160001-10200000	8:9080001-9120000	2.29E-07	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000254091:ENSG00000254235:ENSG00000253887
8	8:10160001-10200000	8:9240001-9280000	3.43E-15	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
8	8:10160001-10200000	8:9720001-9760000	5.13E-37	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000239065
8	8:10160001-10200000	8:9760001-9800000	2.91E-23	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000253230:ENSG00000238496
8	8:10160001-10200000	8:9800001-9840000	3.26E-61	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
8	8:10160001-10200000	8:9840001-9880000	1.22E-64	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
8	8:10160001-10200000	8:9880001-9920000	9.25E-120	HiC	GSE87112	IMR90	intra	ENSG00000260093:ENSG00000175806
8	8:10160001-10200000	8:10120001-10160000	1.38E-35	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100920001-100960000	1.51E-26	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:100880001-100920000	11:102360001-102400000	9.22E-10	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:102800001-102840000	1.11E-07	HiC	GSE87112	Mesendoderm	intra	ENSG00000137745
11	11:100880001-100920000	11:98080001-98120000	1.75E-07	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:98400001-98440000	2.51E-07	HiC	GSE87112	Mesendoderm	intra	ENSG00000254939
11	11:100880001-100920000	11:98480001-98520000	1.61E-07	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:98640001-98680000	1.39E-10	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:99040001-99080000	5.25E-09	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:99200001-99240000	2.90E-15	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:99280001-99320000	2.14E-12	HiC	GSE87112	Mesendoderm	intra	ENSG00000266616
11	11:100880001-100920000	11:99480001-99520000	1.20E-13	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:99520001-99560000	3.03E-07	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:99640001-99680000	3.71E-16	HiC	GSE87112	Mesendoderm	intra	nan

11	11:100880001-100920000	11:99680001-99720000	3.61E-21	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:102040001-102080000	3.36E-09	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:99720001-99760000	2.58E-33	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:99800001-99840000	1.95E-23	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:99840001-99880000	7.17E-29	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:99880001-99920000	1.98E-36	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:99920001-99960000	4.35E-18	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:99960001-100000000	1.56E-11	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:100000001-100040000	1.39E-28	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:100040001-100080000	8.23E-44	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:100080001-100120000	3.97E-17	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:100120001-100160000	6.19E-14	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:100160001-100200000	2.49E-33	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:100200001-100240000	1.66E-18	HiC	GSE87112	Mesendoderm	intra	ENSG00000255547
11	11:100880001-100920000	11:100240001-100280000	4.07E-25	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:100280001-100320000	7.39E-11	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:99760001-99800000	4.38E-57	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:100320001-100360000	2.44E-20	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:101840001-101880000	1.72E-15	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:101760001-101800000	2.31E-11	HiC	GSE87112	Mesendoderm	intra	ENSG00000187151:ENSG00000110318
11	11:100880001-100920000	11:100200001-100240000	1.94E-34	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000255547
11	11:100880001-100920000	11:100240001-100280000	2.67E-36	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:100280001-100320000	1.45E-10	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:100320001-100360000	2.72E-22	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:100360001-100400000	6.94E-10	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:100400001-100440000	2.68E-35	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:100440001-100480000	3.18E-24	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan

11	11:100880001-100920000	11:100480001-100520000	6.84E-08	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000242165
11	11:100880001-100920000	11:100560001-100600000	2.60E-09	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:100800001-100840000	1.07E-21	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000238388
11	11:100880001-100920000	11:100840001-100880000	5.09E-63	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000170647
11	11:100880001-100920000	11:100920001-100960000	3.16E-192	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:100960001-101000000	1.10E-33	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:101800001-101840000	2.81E-17	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:101000001-101040000	5.58E-32	HiC	GSE87112	Mesendoderm	intra	ENSG00000082175:ENSG00000255504
11	11:100880001-100920000	11:101080001-101120000	1.08E-29	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:101120001-101160000	6.38E-58	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:101160001-101200000	1.10E-26	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:101200001-101240000	1.69E-70	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:101240001-101280000	1.89E-75	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:101280001-101320000	1.08E-29	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:101320001-101360000	1.38E-16	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:101360001-101400000	2.70E-16	HiC	GSE87112	Mesendoderm	intra	ENSG00000263885
11	11:100880001-100920000	11:101440001-101480000	1.52E-11	HiC	GSE87112	Mesendoderm	intra	ENSG00000254506
11	11:100880001-100920000	11:101480001-101520000	1.29E-21	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:101600001-101640000	1.86E-13	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:101680001-101720000	3.76E-27	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:101720001-101760000	1.06E-21	HiC	GSE87112	Mesendoderm	intra	ENSG00000137672
11	11:100880001-100920000	11:101040001-101080000	3.40E-171	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:100360001-100400000	5.71E-07	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:100400001-100440000	3.52E-29	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:100440001-100480000	9.61E-40	HiC	GSE87112	Mesendoderm	intra	nan
11	11:100880001-100920000	11:99200001-99240000	4.99E-11	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:99280001-99320000	3.80E-08	HiC	GSE87112	hESC	intra	ENSG00000266616

11	11:100880001-100920000	11:99480001-99520000	6.83E-12	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:99640001-99680000	5.28E-13	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:99680001-99720000	1.17E-16	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:99720001-99760000	1.86E-38	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:99760001-99800000	8.55E-34	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:99800001-99840000	3.99E-32	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:99840001-99880000	2.65E-14	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:99880001-99920000	2.22E-24	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:99920001-99960000	2.28E-18	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:99960001-100000000	3.27E-12	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100000001-100040000	7.67E-27	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:99160001-99200000	2.13E-07	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100040001-100080000	3.80E-38	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100120001-100160000	7.13E-16	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100160001-100200000	4.68E-44	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100200001-100240000	1.76E-42	HiC	GSE87112	hESC	intra	ENSG00000255547
11	11:100880001-100920000	11:100240001-100280000	7.60E-26	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100280001-100320000	3.34E-18	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100320001-100360000	7.89E-17	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100360001-100400000	1.23E-12	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100400001-100440000	1.24E-38	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100440001-100480000	2.10E-53	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100480001-100520000	8.47E-16	HiC	GSE87112	hESC	intra	ENSG00000242165
11	11:100880001-100920000	11:100520001-100560000	1.88E-09	HiC	GSE87112	hESC	intra	ENSG00000255059:ENSG00000248027:ENSG00000165895
11	11:100880001-100920000	11:100560001-100600000	5.64E-09	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100720001-100760000	1.37E-08	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100080001-100120000	2.64E-09	HiC	GSE87112	hESC	intra	nan

11	11:100880001-100920000	11:99080001-99120000	8.25E-08	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:98920001-98960000	7.04E-08	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:98800001-98840000	3.25E-08	HiC	GSE87112	hESC	intra	ENSG00000254830
11	11:100880001-100920000	11:100520001-100560000	8.85E-13	HiC	GSE87112	Mesendoderm	intra	ENSG00000255059:ENSG00000248027:ENSG00000165895
11	11:100880001-100920000	11:100800001-100840000	1.57E-48	HiC	GSE87112	Mesendoderm	intra	ENSG00000238388
11	11:100880001-100920000	11:100920001-100960000	1.23E-132	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100960001-101000000	3.47E-18	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:101000001-101040000	1.32E-31	HiC	GSE87112	hESC	intra	ENSG00000082175:ENSG00000255504
11	11:100880001-100920000	11:101040001-101080000	1.15E-149	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:101080001-101120000	4.59E-18	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:101120001-101160000	3.24E-49	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:101160001-101200000	2.92E-32	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:101200001-101240000	1.54E-62	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:101240001-101280000	1.67E-63	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:101280001-101320000	4.31E-21	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:101320001-101360000	4.52E-19	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:101360001-101400000	1.15E-24	HiC	GSE87112	hESC	intra	ENSG00000263885
11	11:100880001-100920000	11:101400001-101440000	7.01E-10	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:101440001-101480000	5.01E-08	HiC	GSE87112	hESC	intra	ENSG00000254506
11	11:100880001-100920000	11:101480001-101520000	2.17E-22	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:101600001-101640000	1.32E-31	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:101640001-101680000	9.22E-09	HiC	GSE87112	hESC	intra	ENSG00000254534
11	11:100880001-100920000	11:101680001-101720000	6.85E-38	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:101720001-101760000	7.71E-10	HiC	GSE87112	hESC	intra	ENSG00000137672
11	11:100880001-100920000	11:101760001-101800000	1.54E-12	HiC	GSE87112	hESC	intra	ENSG00000187151:ENSG00000110318
11	11:100880001-100920000	11:101800001-101840000	1.13E-12	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:101840001-101880000	1.37E-20	HiC	GSE87112	hESC	intra	nan

11	11:100880001-100920000	11:102040001-102080000	7.39E-07	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:102440001-102480000	2.56E-07	HiC	GSE87112	hESC	intra	ENSG00000256916
11	11:100880001-100920000	11:102560001-102600000	1.12E-07	HiC	GSE87112	hESC	intra	ENSG00000137675:ENSG00000118113
11	11:100880001-100920000	11:98240001-98280000	3.83E-07	HiC	GSE87112	hESC	intra	nan
11	11:100880001-100920000	11:100160001-100200000	2.80E-50	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:100120001-100160000	1.31E-17	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:100080001-100120000	6.33E-18	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:100040001-100080000	8.30E-59	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99160001-99200000	5.33E-18	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:99200001-99240000	6.10E-24	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:99280001-99320000	6.64E-23	HiC	GSE87112	IMR90	intra	ENSG00000266616
11	11:100880001-100920000	11:99360001-99400000	1.44E-08	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:99400001-99440000	9.98E-10	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:99520001-99560000	4.32E-12	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:99600001-99640000	3.52E-08	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:99680001-99720000	4.48E-21	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:99720001-99760000	8.53E-21	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:99760001-99800000	3.29E-30	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:99800001-99840000	1.19E-09	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:99840001-99880000	6.25E-10	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:99880001-99920000	2.65E-09	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:99040001-99080000	1.03E-13	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:99920001-99960000	2.24E-10	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:100000001-100040000	2.03E-23	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:100040001-100080000	1.35E-07	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:100120001-100160000	6.24E-10	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:100160001-100200000	2.29E-27	HiC	GSE87112	IMR90	intra	nan

11	11:100880001-100920000	11:100200001-100240000	1.04E-33	HiC	GSE87112	IMR90	intra	ENSG00000255547
11	11:100880001-100920000	11:100240001-100280000	2.50E-17	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:100280001-100320000	4.35E-09	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:100320001-100360000	4.97E-09	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:100400001-100440000	2.84E-20	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:100440001-100480000	3.24E-11	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:100840001-100880000	3.97E-08	HiC	GSE87112	IMR90	intra	ENSG00000170647
11	11:100880001-100920000	11:100920001-100960000	3.80E-162	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:101000001-101040000	5.85E-20	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000082175:ENSG00000255504
11	11:100880001-100920000	11:99960001-100000000	1.78E-19	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:99000001-99040000	6.41E-11	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:98960001-99000000	1.55E-10	HiC	GSE87112	IMR90	intra	ENSG00000223269
11	11:100880001-100920000	11:98920001-98960000	5.79E-08	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:100960001-101000000	3.74E-09	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:100880001-100920000	11:101000001-101040000	2.97E-10	HiC	GSE87112	Left_Ventricle	intra	ENSG00000082175:ENSG00000255504
11	11:100880001-100920000	11:101040001-101080000	2.72E-16	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:100880001-100920000	11:101200001-101240000	6.62E-12	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:100880001-100920000	11:101240001-101280000	4.32E-30	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:100880001-100920000	11:101280001-101320000	1.11E-09	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:100880001-100920000	11:101360001-101400000	6.92E-12	HiC	GSE87112	Left_Ventricle	intra	ENSG00000263885
11	11:100880001-100920000	11:101720001-101760000	1.38E-10	HiC	GSE87112	Left_Ventricle	intra	ENSG00000137672
11	11:100880001-100920000	11:102040001-102080000	9.34E-08	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:100880001-100920000	11:102080001-102120000	5.59E-13	HiC	GSE87112	Left_Ventricle	intra	ENSG00000254422
11	11:100880001-100920000	11:99000001-99040000	1.59E-08	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:100880001-100920000	11:99160001-99200000	8.75E-10	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:100880001-100920000	11:99640001-99680000	2.82E-07	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:100880001-100920000	11:99680001-99720000	5.47E-07	HiC	GSE87112	Left_Ventricle	intra	nan

11	11:100880001-100920000	11:100360001-100400000	2.12E-07	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:100880001-100920000	11:100480001-100520000	1.91E-11	HiC	GSE87112	Left_Ventricle	intra	ENSG00000242165
11	11:100880001-100920000	11:100840001-100880000	3.74E-15	HiC	GSE87112	Left_Ventricle	intra	ENSG00000170647
11	11:100880001-100920000	11:100920001-100960000	3.19E-09	HiC	GSE87112	GM12878	intra	nan
11	11:100880001-100920000	11:100920001-100960000	1.10E-21	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:101040001-101080000	4.78E-09	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:97920001-97960000	1.02E-12	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:98080001-98120000	5.20E-12	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:98240001-98280000	6.03E-11	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:98400001-98440000	3.90E-08	HiC	GSE87112	IMR90	intra	ENSG00000254939
11	11:100880001-100920000	11:98520001-98560000	3.38E-08	HiC	GSE87112	IMR90	intra	ENSG00000254599
11	11:100880001-100920000	11:98640001-98680000	1.77E-12	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:98680001-98720000	3.69E-07	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:98760001-98800000	8.98E-11	HiC	GSE87112	IMR90	intra	nan
11	11:100880001-100920000	11:98880001-98920000	2.63E-10	HiC	GSE87112	IMR90	intra	ENSG00000149972
11	11:100880001-100920000	11:101040001-101080000	3.69E-80	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:100800001-100840000	5.97E-55	HiC	GSE87112	hESC	intra	ENSG00000238388
11	11:100880001-100920000	11:101120001-101160000	3.58E-11	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:101240001-101280000	6.03E-33	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98920001-98960000	9.64E-45	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98960001-99000000	3.82E-19	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000223269
11	11:100880001-100920000	11:99000001-99040000	3.96E-38	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99040001-99080000	1.12E-31	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99120001-99160000	7.07E-15	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99160001-99200000	1.76E-52	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99200001-99240000	5.96E-97	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99240001-99280000	5.16E-32	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan

11	11:100880001-100920000	11:99280001-99320000	1.31E-53	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000266616
11	11:100880001-100920000	11:99320001-99360000	1.05E-13	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99360001-99400000	1.53E-26	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99400001-99440000	4.34E-27	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98880001-98920000	1.70E-36	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000149972
11	11:100880001-100920000	11:99440001-99480000	3.10E-20	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99520001-99560000	1.26E-45	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99560001-99600000	6.57E-07	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99600001-99640000	1.33E-20	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99640001-99680000	1.72E-21	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99680001-99720000	1.70E-54	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99720001-99760000	2.58E-68	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99760001-99800000	3.90E-78	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99800001-99840000	2.27E-42	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99840001-99880000	5.27E-40	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99880001-99920000	1.78E-56	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99920001-99960000	1.02E-36	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99960001-100000000	2.10E-14	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:100000001-100040000	1.16E-46	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:99480001-99520000	4.94E-58	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98840001-98880000	6.53E-10	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98800001-98840000	2.23E-32	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000254830
11	11:100880001-100920000	11:98760001-98800000	1.69E-11	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:101280001-101320000	4.52E-10	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:101320001-101360000	3.77E-11	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:101360001-101400000	2.67E-15	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000263885
11	11:100880001-100920000	11:101400001-101440000	9.01E-08	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan

11	11:100880001-100920000	11:101480001-101520000	3.28E-09	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:101600001-101640000	1.96E-21	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:101640001-101680000	2.37E-10	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000254534
11	11:100880001-100920000	11:101680001-101720000	2.15E-19	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:101720001-101760000	2.89E-08	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000137672
11	11:100880001-100920000	11:101800001-101840000	4.65E-07	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:101840001-101880000	1.92E-13	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:97880001-97920000	2.98E-18	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:97920001-97960000	3.31E-15	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98040001-98080000	2.39E-10	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98080001-98120000	5.34E-12	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98120001-98160000	2.19E-07	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98160001-98200000	6.75E-09	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98200001-98240000	1.17E-16	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98240001-98280000	3.61E-29	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98280001-98320000	1.84E-09	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98360001-98400000	1.74E-13	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98400001-98440000	1.29E-12	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000254939
11	11:100880001-100920000	11:98480001-98520000	3.16E-12	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98520001-98560000	2.30E-20	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000254599
11	11:100880001-100920000	11:98560001-98600000	4.85E-10	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98600001-98640000	1.13E-11	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98640001-98680000	5.12E-14	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98680001-98720000	3.62E-17	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:98720001-98760000	2.83E-08	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:101200001-101240000	1.00E-17	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:100880001-100920000	11:100840001-100880000	1.12E-58	HiC	GSE87112	hESC	intra	ENSG00000170647

11	11:104000001-104040000	11:104040001-104080000	4.94E-25	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:104000001-104040000	11:105720001-105760000	1.20E-64	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:105680001-105720000	2.18E-26	HiC	GSE87112	Mesendoderm	intra	ENSG00000222663
11	11:104000001-104040000	11:105640001-105680000	1.98E-37	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:105600001-105640000	6.13E-17	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:105560001-105600000	2.24E-18	HiC	GSE87112	Mesendoderm	intra	ENSG00000255403
11	11:104000001-104040000	11:105520001-105560000	1.23E-08	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:105760001-105800000	1.76E-20	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:105480001-105520000	8.87E-22	HiC	GSE87112	Mesendoderm	intra	ENSG00000152578
11	11:104000001-104040000	11:105400001-105440000	1.03E-16	HiC	GSE87112	Mesendoderm	intra	ENSG00000254998
11	11:104000001-104040000	11:105280001-105320000	1.65E-12	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:105240001-105280000	3.22E-09	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:105160001-105200000	1.36E-11	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:105120001-105160000	3.32E-12	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:105040001-105080000	3.12E-20	HiC	GSE87112	Mesendoderm	intra	ENSG00000254767
11	11:104000001-104040000	11:105440001-105480000	7.14E-17	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:104960001-105000000	2.45E-10	HiC	GSE87112	Mesendoderm	intra	ENSG00000137752:ENSG00000204397:ENSG00000255221:ENSG00000255430
11	11:104000001-104040000	11:105800001-105840000	1.70E-16	HiC	GSE87112	Mesendoderm	intra	ENSG00000263371
11	11:104000001-104040000	11:105880001-105920000	4.25E-24	HiC	GSE87112	Mesendoderm	intra	ENSG00000170903
11	11:104000001-104040000	11:103440001-103480000	9.53E-18	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103400001-103440000	4.49E-13	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103360001-103400000	4.10E-12	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103320001-103360000	1.91E-19	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103280001-103320000	4.95E-11	HiC	GSE87112	Mesendoderm	intra	ENSG00000254824
11	11:104000001-104040000	11:103200001-103240000	2.98E-52	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:105840001-105880000	1.18E-28	HiC	GSE87112	Mesendoderm	intra	ENSG00000252081

11	11:104000001-104040000	11:103160001-103200000	3.50E-45	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103080001-103120000	2.16E-40	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103040001-103080000	2.68E-24	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103000001-103040000	3.27E-12	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:102920001-102960000	6.65E-10	HiC	GSE87112	Mesendoderm	intra	ENSG00000260966
11	11:104000001-104040000	11:105960001-106000000	2.80E-07	HiC	GSE87112	Mesendoderm	intra	ENSG00000254433
11	11:104000001-104040000	11:105920001-105960000	1.51E-12	HiC	GSE87112	Mesendoderm	intra	ENSG00000182359:ENSG00000149313
11	11:104000001-104040000	11:103120001-103160000	8.63E-37	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:104920001-104960000	1.15E-10	HiC	GSE87112	Mesendoderm	intra	ENSG00000254750
11	11:104000001-104040000	11:104880001-104920000	4.16E-10	HiC	GSE87112	Mesendoderm	intra	ENSG00000137757
11	11:104000001-104040000	11:104840001-104880000	1.53E-09	HiC	GSE87112	Mesendoderm	intra	ENSG00000196954
11	11:104000001-104040000	11:103440001-103480000	4.57E-26	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103400001-103440000	5.99E-26	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103360001-103400000	9.70E-19	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103320001-103360000	1.34E-23	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103280001-103320000	1.07E-27	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000254824
11	11:104000001-104040000	11:103200001-103240000	2.15E-34	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103480001-103520000	1.09E-32	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103160001-103200000	1.08E-75	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103080001-103120000	3.33E-58	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103040001-103080000	2.17E-39	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103000001-103040000	8.25E-18	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:106680001-106720000	4.07E-13	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000213252
11	11:104000001-104040000	11:106600001-106640000	9.80E-08	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:106320001-106360000	2.44E-08	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103120001-103160000	3.24E-41	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103520001-103560000	2.66E-70	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan

11	11:104000001-104040000	11:103560001-103600000	6.59E-93	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103600001-103640000	2.21E-39	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:104600001-104640000	3.53E-08	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:104520001-104560000	8.95E-17	HiC	GSE87112	Mesendoderm	intra	ENSG00000270449
11	11:104000001-104040000	11:104400001-104440000	3.49E-17	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:104320001-104360000	1.75E-10	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:104240001-104280000	6.73E-09	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:104200001-104240000	5.51E-09	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:104040001-104080000	6.07E-81	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103920001-103960000	1.15E-14	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103880001-103920000	2.75E-58	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000170967
11	11:104000001-104040000	11:103840001-103880000	2.50E-35	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103800001-103840000	2.67E-113	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000255548
11	11:104000001-104040000	11:103760001-103800000	1.44E-89	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000254987
11	11:104000001-104040000	11:103720001-103760000	9.95E-97	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000264200
11	11:104000001-104040000	11:103680001-103720000	1.68E-123	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103640001-103680000	1.38E-85	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103480001-103520000	5.97E-36	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:106080001-106120000	4.05E-10	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103520001-103560000	6.44E-47	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103600001-103640000	2.28E-25	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103080001-103120000	1.00E-37	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:103040001-103080000	2.49E-36	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:103000001-103040000	1.38E-09	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:102960001-103000000	2.02E-07	HiC	GSE87112	hESC	intra	ENSG00000137692:ENSG00000187240
11	11:104000001-104040000	11:102920001-102960000	4.26E-08	HiC	GSE87112	hESC	intra	ENSG00000260966
11	11:104000001-104040000	11:106880001-106920000	1.54E-08	HiC	GSE87112	hESC	intra	ENSG00000152402

11	11:104000001-104040000	11:103120001-103160000	3.83E-31	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:106320001-106360000	9.56E-11	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:105920001-105960000	1.55E-14	HiC	GSE87112	hESC	intra	ENSG00000182359:ENSG00000149313
11	11:104000001-104040000	11:105880001-105920000	5.31E-15	HiC	GSE87112	hESC	intra	ENSG00000170903
11	11:104000001-104040000	11:105840001-105880000	8.13E-27	HiC	GSE87112	hESC	intra	ENSG00000252081
11	11:104000001-104040000	11:105800001-105840000	8.50E-25	HiC	GSE87112	hESC	intra	ENSG00000263371
11	11:104000001-104040000	11:105760001-105800000	1.58E-13	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:105720001-105760000	3.39E-60	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:105960001-106000000	4.71E-14	HiC	GSE87112	hESC	intra	ENSG00000254433
11	11:104000001-104040000	11:105680001-105720000	1.39E-30	HiC	GSE87112	hESC	intra	ENSG00000222663
11	11:104000001-104040000	11:103160001-103200000	3.00E-34	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:103280001-103320000	3.57E-21	HiC	GSE87112	hESC	intra	ENSG00000254824
11	11:104000001-104040000	11:103840001-103880000	2.54E-17	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:103800001-103840000	2.96E-58	HiC	GSE87112	hESC	intra	ENSG00000255548
11	11:104000001-104040000	11:103760001-103800000	2.82E-36	HiC	GSE87112	hESC	intra	ENSG00000254987
11	11:104000001-104040000	11:103720001-103760000	5.72E-63	HiC	GSE87112	hESC	intra	ENSG00000264200
11	11:104000001-104040000	11:103680001-103720000	2.53E-69	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:103640001-103680000	1.61E-47	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:103200001-103240000	5.05E-24	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:103520001-103560000	1.38E-25	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:103480001-103520000	3.23E-28	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:103440001-103480000	7.89E-17	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:103400001-103440000	8.30E-09	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:103360001-103400000	2.78E-22	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:103320001-103360000	4.03E-13	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:103560001-103600000	9.60E-59	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:105640001-105680000	1.11E-36	HiC	GSE87112	hESC	intra	nan

11	11:104000001-104040000	11:105600001-105640000	3.93E-07	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:105560001-105600000	2.22E-15	HiC	GSE87112	hESC	intra	ENSG00000255403
11	11:104000001-104040000	11:104560001-104600000	3.23E-13	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:104520001-104560000	1.23E-12	HiC	GSE87112	hESC	intra	ENSG00000270449
11	11:104000001-104040000	11:104400001-104440000	4.63E-14	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:104320001-104360000	2.09E-11	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:104200001-104240000	4.41E-09	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:104040001-104080000	8.64E-91	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:104600001-104640000	1.75E-09	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:103920001-103960000	3.11E-12	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103840001-103880000	1.71E-25	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103800001-103840000	7.17E-114	HiC	GSE87112	Mesendoderm	intra	ENSG00000255548
11	11:104000001-104040000	11:103760001-103800000	4.96E-48	HiC	GSE87112	Mesendoderm	intra	ENSG00000254987
11	11:104000001-104040000	11:103720001-103760000	8.97E-79	HiC	GSE87112	Mesendoderm	intra	ENSG00000264200
11	11:104000001-104040000	11:103680001-103720000	1.04E-88	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103640001-103680000	1.62E-55	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103880001-103920000	5.06E-49	HiC	GSE87112	Mesendoderm	intra	ENSG00000170967
11	11:104000001-104040000	11:104640001-104680000	1.88E-07	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:104720001-104760000	6.84E-09	HiC	GSE87112	hESC	intra	ENSG00000254569
11	11:104000001-104040000	11:104800001-104840000	3.00E-16	HiC	GSE87112	hESC	intra	ENSG00000196954
11	11:104000001-104040000	11:105520001-105560000	1.76E-08	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:105480001-105520000	1.84E-17	HiC	GSE87112	hESC	intra	ENSG00000152578
11	11:104000001-104040000	11:105440001-105480000	2.05E-19	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:105400001-105440000	5.65E-18	HiC	GSE87112	hESC	intra	ENSG00000254998
11	11:104000001-104040000	11:105360001-105400000	8.09E-07	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:105280001-105320000	6.15E-15	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:105240001-105280000	2.19E-10	HiC	GSE87112	hESC	intra	nan

11	11:104000001-104040000	11:105200001-105240000	1.17E-16	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:105160001-105200000	2.63E-22	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:105120001-105160000	3.54E-14	HiC	GSE87112	hESC	intra	nan
11	11:104000001-104040000	11:105040001-105080000	4.48E-18	HiC	GSE87112	hESC	intra	ENSG00000254767
11	11:104000001-104040000	11:105000001-105040000	3.87E-07	HiC	GSE87112	hESC	intra	ENSG00000255501
11	11:104000001-104040000	11:104920001-104960000	4.29E-14	HiC	GSE87112	hESC	intra	ENSG00000254750
11	11:104000001-104040000	11:104880001-104920000	1.75E-24	HiC	GSE87112	hESC	intra	ENSG00000137757
11	11:104000001-104040000	11:104840001-104880000	8.88E-22	HiC	GSE87112	hESC	intra	ENSG00000196954
11	11:104000001-104040000	11:103560001-103600000	5.24E-45	HiC	GSE87112	Mesendoderm	intra	nan
11	11:104000001-104040000	11:103880001-103920000	2.54E-29	HiC	GSE87112	hESC	intra	ENSG00000170967
11	11:104000001-104040000	11:105960001-106000000	6.12E-13	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000254433
11	11:104000001-104040000	11:105880001-105920000	1.80E-24	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000170903
11	11:104000001-104040000	11:105160001-105200000	4.46E-16	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:105120001-105160000	2.09E-16	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:105040001-105080000	4.20E-38	HiC	GSE87112	IMR90	intra	ENSG00000254767
11	11:104000001-104040000	11:104960001-105000000	4.30E-09	HiC	GSE87112	IMR90	intra	ENSG00000137752:ENSG00000204397:ENSG00000255221:ENSG00000255430
11	11:104000001-104040000	11:104920001-104960000	8.22E-13	HiC	GSE87112	IMR90	intra	ENSG00000254750
11	11:104000001-104040000	11:104880001-104920000	9.61E-31	HiC	GSE87112	IMR90	intra	ENSG00000137757
11	11:104000001-104040000	11:105200001-105240000	5.59E-13	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:104840001-104880000	2.58E-16	HiC	GSE87112	IMR90	intra	ENSG00000196954
11	11:104000001-104040000	11:104760001-104800000	2.50E-10	HiC	GSE87112	IMR90	intra	ENSG00000204403:ENSG00000235505
11	11:104000001-104040000	11:104680001-104720000	7.20E-07	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:104640001-104680000	1.58E-10	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:104600001-104640000	3.67E-10	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:104560001-104600000	1.62E-24	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:104520001-104560000	6.35E-28	HiC	GSE87112	IMR90	intra	ENSG00000270449

11	11:104000001-104040000	11:104800001-104840000	2.13E-13	HiC	GSE87112	IMR90	intra	ENSG00000196954
11	11:104000001-104040000	11:104480001-104520000	4.72E-12	HiC	GSE87112	IMR90	intra	ENSG00000256422
11	11:104000001-104040000	11:105240001-105280000	4.97E-10	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:105320001-105360000	4.21E-12	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:105880001-105920000	1.22E-18	HiC	GSE87112	IMR90	intra	ENSG00000170903
11	11:104000001-104040000	11:105840001-105880000	1.19E-17	HiC	GSE87112	IMR90	intra	ENSG00000252081
11	11:104000001-104040000	11:105800001-105840000	4.48E-14	HiC	GSE87112	IMR90	intra	ENSG00000263371
11	11:104000001-104040000	11:105760001-105800000	5.19E-17	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:105720001-105760000	2.62E-56	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:105680001-105720000	6.82E-41	HiC	GSE87112	IMR90	intra	ENSG00000222663
11	11:104000001-104040000	11:105280001-105320000	1.43E-09	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:105640001-105680000	1.14E-32	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:105560001-105600000	6.47E-19	HiC	GSE87112	IMR90	intra	ENSG00000255403
11	11:104000001-104040000	11:105520001-105560000	9.67E-18	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:105480001-105520000	4.13E-46	HiC	GSE87112	IMR90	intra	ENSG00000152578
11	11:104000001-104040000	11:105440001-105480000	2.84E-17	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:105400001-105440000	7.75E-22	HiC	GSE87112	IMR90	intra	ENSG00000254998
11	11:104000001-104040000	11:105360001-105400000	4.29E-08	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:105600001-105640000	8.66E-21	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:104400001-104440000	7.71E-28	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:104320001-104360000	2.58E-23	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:104240001-104280000	4.01E-17	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103040001-103080000	3.28E-09	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:104000001-104040000	11:103000001-103040000	5.43E-08	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:104000001-104040000	11:105720001-105760000	3.18E-18	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:104000001-104040000	11:105680001-105720000	1.48E-09	HiC	GSE87112	Left_Ventricle	intra	ENSG00000222663
11	11:104000001-104040000	11:105640001-105680000	1.43E-13	HiC	GSE87112	Left_Ventricle	intra	nan

11	11:104000001-104040000	11:105600001-105640000	4.39E-09	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:104000001-104040000	11:103080001-103120000	1.34E-11	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:104000001-104040000	11:105480001-105520000	9.78E-08	HiC	GSE87112	Left_Ventricle	intra	ENSG00000152578
11	11:104000001-104040000	11:105280001-105320000	3.62E-13	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:104000001-104040000	11:105120001-105160000	6.44E-07	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:104000001-104040000	11:104880001-104920000	1.05E-14	HiC	GSE87112	Left_Ventricle	intra	ENSG00000137757
11	11:104000001-104040000	11:104520001-104560000	6.58E-18	HiC	GSE87112	Left_Ventricle	intra	ENSG00000270449
11	11:104000001-104040000	11:104480001-104520000	1.12E-08	HiC	GSE87112	Left_Ventricle	intra	ENSG00000256422
11	11:104000001-104040000	11:104400001-104440000	2.69E-14	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:104000001-104040000	11:105400001-105440000	8.12E-08	HiC	GSE87112	Left_Ventricle	intra	ENSG00000254998
11	11:104000001-104040000	11:103120001-103160000	2.14E-08	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:104000001-104040000	11:103280001-103320000	1.55E-11	HiC	GSE87112	Left_Ventricle	intra	ENSG00000254824
11	11:104000001-104040000	11:104200001-104240000	3.10E-10	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:104040001-104080000	6.40E-48	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103760001-103800000	1.00E-20	HiC	GSE87112	GM12878	intra	ENSG00000254987
11	11:104000001-104040000	11:103680001-103720000	1.98E-07	HiC	GSE87112	GM12878	intra	nan
11	11:104000001-104040000	11:103560001-103600000	1.08E-09	HiC	GSE87112	GM12878	intra	nan
11	11:104000001-104040000	11:103520001-103560000	4.61E-07	HiC	GSE87112	GM12878	intra	nan
11	11:104000001-104040000	11:104040001-104080000	7.31E-08	HiC	GSE87112	GM12878	intra	nan
11	11:104000001-104040000	11:103880001-103920000	4.09E-15	HiC	GSE87112	Left_Ventricle	intra	ENSG00000170967
11	11:104000001-104040000	11:103800001-103840000	1.26E-34	HiC	GSE87112	Left_Ventricle	intra	ENSG00000255548
11	11:104000001-104040000	11:103760001-103800000	1.67E-10	HiC	GSE87112	Left_Ventricle	intra	ENSG00000254987
11	11:104000001-104040000	11:103720001-103760000	9.44E-16	HiC	GSE87112	Left_Ventricle	intra	ENSG00000264200
11	11:104000001-104040000	11:103680001-103720000	8.34E-38	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:104000001-104040000	11:103640001-103680000	2.78E-15	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:104000001-104040000	11:103480001-103520000	2.48E-17	HiC	GSE87112	Left_Ventricle	intra	nan
11	11:104000001-104040000	11:103360001-103400000	3.63E-12	HiC	GSE87112	Left_Ventricle	intra	nan

11	11:104000001-104040000	11:105920001-105960000	1.71E-15	HiC	GSE87112	IMR90	intra	ENSG00000182359:ENSG00000149313
11	11:104000001-104040000	11:105920001-105960000	2.90E-21	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000182359:ENSG00000149313
11	11:104000001-104040000	11:105960001-106000000	1.89E-07	HiC	GSE87112	IMR90	intra	ENSG00000254433
11	11:104000001-104040000	11:106680001-106720000	4.84E-07	HiC	GSE87112	IMR90	intra	ENSG00000213252
11	11:104000001-104040000	11:105120001-105160000	2.02E-24	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:105080001-105120000	1.66E-09	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000255336
11	11:104000001-104040000	11:105040001-105080000	1.45E-36	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000254767
11	11:104000001-104040000	11:105000001-105040000	4.90E-09	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000255501
11	11:104000001-104040000	11:104960001-105000000	1.82E-33	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000137752:ENSG00000204397:ENSG00000255221:ENSG00000255430
11	11:104000001-104040000	11:104920001-104960000	2.62E-18	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000254750
11	11:104000001-104040000	11:105160001-105200000	4.24E-51	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:104880001-104920000	3.14E-55	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000137757
11	11:104000001-104040000	11:104800001-104840000	1.86E-22	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000196954
11	11:104000001-104040000	11:104760001-104800000	6.41E-12	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000204403:ENSG00000235505
11	11:104000001-104040000	11:104720001-104760000	8.60E-16	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000254569
11	11:104000001-104040000	11:104680001-104720000	9.26E-19	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:104640001-104680000	3.42E-19	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:104600001-104640000	5.37E-29	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:104840001-104880000	1.24E-17	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000196954
11	11:104000001-104040000	11:104560001-104600000	1.49E-26	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:105200001-105240000	5.34E-28	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:105280001-105320000	5.94E-27	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:105840001-105880000	1.87E-54	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000252081
11	11:104000001-104040000	11:105800001-105840000	6.63E-36	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000263371
11	11:104000001-104040000	11:105760001-105800000	4.99E-41	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:105720001-105760000	2.23E-219	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan

11	11:104000001-104040000	11:105680001-105720000	2.52E-52	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000222663
11	11:104000001-104040000	11:105640001-105680000	1.02E-102	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:105240001-105280000	2.76E-16	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:105600001-105640000	4.77E-35	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:105520001-105560000	3.76E-33	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:105480001-105520000	2.47E-52	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000152578
11	11:104000001-104040000	11:105440001-105480000	1.03E-38	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:105400001-105440000	9.25E-34	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000254998
11	11:104000001-104040000	11:105360001-105400000	3.35E-15	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:105320001-105360000	6.95E-21	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:105560001-105600000	1.78E-40	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000255403
11	11:104000001-104040000	11:104520001-104560000	8.20E-36	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000270449
11	11:104000001-104040000	11:104480001-104520000	3.81E-13	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000256422
11	11:104000001-104040000	11:104400001-104440000	2.21E-39	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103400001-103440000	2.02E-07	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103360001-103400000	9.33E-14	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103320001-103360000	8.27E-20	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103280001-103320000	2.91E-12	HiC	GSE87112	IMR90	intra	ENSG00000254824
11	11:104000001-104040000	11:103200001-103240000	2.87E-33	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103160001-103200000	1.41E-52	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103440001-103480000	1.45E-20	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103120001-103160000	2.17E-26	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103040001-103080000	1.41E-46	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103000001-103040000	5.67E-15	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:102960001-103000000	2.65E-11	HiC	GSE87112	IMR90	intra	ENSG00000137692:ENSG00000187240
11	11:104000001-104040000	11:102920001-102960000	1.74E-13	HiC	GSE87112	IMR90	intra	ENSG00000260966
11	11:104000001-104040000	11:102880001-102920000	1.86E-08	HiC	GSE87112	IMR90	intra	ENSG00000239861

11	11:104000001-104040000	11:106880001-106920000	1.83E-07	HiC	GSE87112	IMR90	intra	ENSG00000152402
11	11:104000001-104040000	11:103080001-103120000	7.61E-42	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103480001-103520000	5.47E-56	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103520001-103560000	3.51E-41	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103560001-103600000	4.08E-47	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:104320001-104360000	1.78E-27	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:104280001-104320000	9.59E-16	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:104240001-104280000	3.02E-27	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:104200001-104240000	3.56E-17	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:104120001-104160000	2.59E-13	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000252251
11	11:104000001-104040000	11:104040001-104080000	2.00E-99	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
11	11:104000001-104040000	11:103920001-103960000	4.05E-18	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103880001-103920000	2.46E-29	HiC	GSE87112	IMR90	intra	ENSG00000170967
11	11:104000001-104040000	11:103840001-103880000	4.84E-29	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103800001-103840000	3.88E-81	HiC	GSE87112	IMR90	intra	ENSG00000255548
11	11:104000001-104040000	11:103760001-103800000	3.43E-41	HiC	GSE87112	IMR90	intra	ENSG00000254987
11	11:104000001-104040000	11:103720001-103760000	3.63E-42	HiC	GSE87112	IMR90	intra	ENSG00000264200
11	11:104000001-104040000	11:103680001-103720000	2.03E-69	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103640001-103680000	1.82E-68	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103600001-103640000	7.13E-27	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:106320001-106360000	2.42E-10	HiC	GSE87112	IMR90	intra	nan
11	11:104000001-104040000	11:103920001-103960000	4.65E-10	HiC	GSE87112	hESC	intra	nan
12	12:66000001-66040000	12:66080001-66120000	8.64E-07	HiC	GSE87112	Left_Ventricle	intra	nan
12	12:66000001-66040000	12:66480001-66520000	2.65E-67	HiC	GSE87112	Mesendoderm	intra	nan
12	12:66000001-66040000	12:66440001-66480000	1.76E-31	HiC	GSE87112	Mesendoderm	intra	ENSG00000251857
12	12:66000001-66040000	12:66400001-66440000	4.08E-58	HiC	GSE87112	Mesendoderm	intra	ENSG00000256259:ENSG00000213343
12	12:66000001-66040000	12:66360001-66400000	1.15E-75	HiC	GSE87112	Mesendoderm	intra	nan

12	12:66000001-66040000	12:66320001-66360000	1.68E-55	HiC	GSE87112	Mesendoderm	intra	ENSG00000256083
12	12:66000001-66040000	12:66280001-66320000	9.73E-31	HiC	GSE87112	Mesendoderm	intra	ENSG00000267942
12	12:66000001-66040000	12:66240001-66280000	1.02E-18	HiC	GSE87112	Mesendoderm	intra	ENSG00000197301:ENSG00000211577
12	12:66000001-66040000	12:66200001-66240000	4.41E-31	HiC	GSE87112	Mesendoderm	intra	ENSG00000241749:ENSG00000149948
12	12:66000001-66040000	12:66160001-66200000	5.54E-12	HiC	GSE87112	Mesendoderm	intra	nan
12	12:66000001-66040000	12:66120001-66160000	8.25E-10	HiC	GSE87112	Mesendoderm	intra	nan
12	12:66000001-66040000	12:66760001-66800000	1.10E-09	HiC	GSE87112	Mesendoderm	intra	nan
12	12:66000001-66040000	12:66080001-66120000	1.52E-36	HiC	GSE87112	Mesendoderm	intra	nan
12	12:66000001-66040000	12:65920001-65960000	1.08E-82	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000255866
12	12:66000001-66040000	12:65880001-65920000	1.89E-32	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
12	12:66000001-66040000	12:65840001-65880000	4.59E-16	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
12	12:66000001-66040000	12:65800001-65840000	1.47E-82	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000215208
12	12:66000001-66040000	12:65760001-65800000	4.71E-93	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
12	12:66000001-66040000	12:65720001-65760000	3.58E-25	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
12	12:66000001-66040000	12:65680001-65720000	6.00E-165	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
12	12:66000001-66040000	12:65640001-65680000	4.74E-08	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000174099:ENSG00000250280
12	12:66000001-66040000	12:65600001-65640000	4.65E-39	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
12	12:66000001-66040000	12:65560001-65600000	7.83E-35	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000174106
12	12:66000001-66040000	12:65960001-66000000	1.69E-15	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000256268
12	12:66000001-66040000	12:65280001-65320000	6.73E-14	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
12	12:66000001-66040000	12:66800001-66840000	5.21E-12	HiC	GSE87112	Mesendoderm	intra	nan
12	12:66000001-66040000	12:65560001-65600000	5.98E-12	HiC	GSE87112	Mesendoderm	intra	ENSG00000174106
12	12:66000001-66040000	12:65760001-65800000	9.54E-23	HiC	GSE87112	hESC	intra	nan
12	12:66000001-66040000	12:65680001-65720000	6.30E-50	HiC	GSE87112	hESC	intra	nan
12	12:66000001-66040000	12:65600001-65640000	3.58E-08	HiC	GSE87112	hESC	intra	nan
12	12:66000001-66040000	12:65560001-65600000	2.47E-12	HiC	GSE87112	hESC	intra	ENSG00000174106
12	12:66000001-66040000	12:65280001-65320000	1.88E-16	HiC	GSE87112	hESC	intra	nan

12	12:66000001-66040000	12:65240001-65280000	9.50E-10	HiC	GSE87112	hESC	intra	ENSG00000243020:ENSG00000255693
12	12:66000001-66040000	12:66480001-66520000	2.87E-73	HiC	GSE87112	hESC	intra	nan
12	12:66000001-66040000	12:66440001-66480000	1.85E-52	HiC	GSE87112	hESC	intra	ENSG00000251857
12	12:66000001-66040000	12:66400001-66440000	1.79E-74	HiC	GSE87112	hESC	intra	ENSG00000256259:ENSG00000213343
12	12:66000001-66040000	12:66360001-66400000	1.47E-69	HiC	GSE87112	hESC	intra	nan
12	12:66000001-66040000	12:65240001-65280000	1.47E-14	HiC	GSE87112	Mesendoderm	intra	ENSG00000243020:ENSG00000255693
12	12:66000001-66040000	12:66320001-66360000	1.09E-57	HiC	GSE87112	hESC	intra	ENSG00000256083
12	12:66000001-66040000	12:66240001-66280000	1.28E-12	HiC	GSE87112	hESC	intra	ENSG00000197301:ENSG00000211577
12	12:66000001-66040000	12:66200001-66240000	6.32E-22	HiC	GSE87112	hESC	intra	ENSG00000241749:ENSG00000149948
12	12:66000001-66040000	12:66120001-66160000	6.37E-07	HiC	GSE87112	hESC	intra	nan
12	12:66000001-66040000	12:66080001-66120000	1.50E-13	HiC	GSE87112	hESC	intra	nan
12	12:66000001-66040000	12:65920001-65960000	3.80E-17	HiC	GSE87112	Mesendoderm	intra	ENSG00000255866
12	12:66000001-66040000	12:65880001-65920000	1.32E-11	HiC	GSE87112	Mesendoderm	intra	nan
12	12:66000001-66040000	12:65800001-65840000	1.65E-18	HiC	GSE87112	Mesendoderm	intra	ENSG00000215208
12	12:66000001-66040000	12:65760001-65800000	2.85E-21	HiC	GSE87112	Mesendoderm	intra	nan
12	12:66000001-66040000	12:65680001-65720000	8.12E-73	HiC	GSE87112	Mesendoderm	intra	nan
12	12:66000001-66040000	12:65600001-65640000	1.47E-15	HiC	GSE87112	Mesendoderm	intra	nan
12	12:66000001-66040000	12:66280001-66320000	1.95E-26	HiC	GSE87112	hESC	intra	ENSG00000267942
12	12:66000001-66040000	12:65240001-65280000	1.65E-08	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000243020:ENSG00000255693
12	12:66000001-66040000	12:66480001-66520000	3.31E-101	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
12	12:66000001-66040000	12:66440001-66480000	1.37E-78	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000251857
12	12:66000001-66040000	12:66160001-66200000	4.45E-22	HiC	GSE87112	IMR90	intra	nan
12	12:66000001-66040000	12:66120001-66160000	4.84E-42	HiC	GSE87112	IMR90	intra	nan
12	12:66000001-66040000	12:66080001-66120000	4.70E-89	HiC	GSE87112	IMR90	intra	nan
12	12:66000001-66040000	12:66040001-66080000	6.33E-93	HiC	GSE87112	IMR90	intra	ENSG00000213344
12	12:66000001-66040000	12:66400001-66440000	6.68E-07	HiC	GSE87112	GM12878	intra	ENSG00000256259:ENSG00000213343
12	12:66000001-66040000	12:65680001-65720000	8.25E-10	HiC	GSE87112	Left_Ventricle	intra	nan

12	12:66000001-66040000	12:65600001-65640000	3.51E-10	HiC	GSE87112	Left_Ventricle	intra	nan
12	12:66000001-66040000	12:65520001-65560000	7.67E-09	HiC	GSE87112	Left_Ventricle	intra	ENSG00000255700
12	12:66000001-66040000	12:65480001-65520000	3.12E-10	HiC	GSE87112	Left_Ventricle	intra	ENSG00000156076:ENSG00000207099
12	12:66000001-66040000	12:65440001-65480000	4.12E-07	HiC	GSE87112	Left_Ventricle	intra	nan
12	12:66000001-66040000	12:66200001-66240000	2.22E-119	HiC	GSE87112	IMR90	intra	ENSG00000241749:ENSG00000149948
12	12:66000001-66040000	12:65400001-65440000	1.33E-11	HiC	GSE87112	Left_Ventricle	intra	ENSG00000198671
12	12:66000001-66040000	12:65320001-65360000	3.92E-10	HiC	GSE87112	Left_Ventricle	intra	ENSG00000221564
12	12:66000001-66040000	12:65280001-65320000	9.62E-15	HiC	GSE87112	Left_Ventricle	intra	nan
12	12:66000001-66040000	12:64280001-64320000	2.63E-08	HiC	GSE87112	Left_Ventricle	intra	ENSG00000256399
12	12:66000001-66040000	12:66760001-66800000	2.41E-09	HiC	GSE87112	Left_Ventricle	intra	nan
12	12:66000001-66040000	12:66480001-66520000	1.64E-22	HiC	GSE87112	Left_Ventricle	intra	nan
12	12:66000001-66040000	12:66440001-66480000	2.29E-24	HiC	GSE87112	Left_Ventricle	intra	ENSG00000251857
12	12:66000001-66040000	12:66400001-66440000	2.09E-25	HiC	GSE87112	Left_Ventricle	intra	ENSG00000256259:ENSG00000213343
12	12:66000001-66040000	12:66360001-66400000	3.07E-26	HiC	GSE87112	Left_Ventricle	intra	nan
12	12:66000001-66040000	12:66280001-66320000	4.07E-18	HiC	GSE87112	Left_Ventricle	intra	ENSG00000267942
12	12:66000001-66040000	12:66160001-66200000	1.31E-08	HiC	GSE87112	Left_Ventricle	intra	nan
12	12:66000001-66040000	12:65360001-65400000	1.15E-17	HiC	GSE87112	Left_Ventricle	intra	ENSG00000248995
12	12:66000001-66040000	12:66240001-66280000	7.27E-47	HiC	GSE87112	IMR90	intra	ENSG00000197301:ENSG00000211577
12	12:66000001-66040000	12:66280001-66320000	1.76E-50	HiC	GSE87112	IMR90	intra	ENSG00000267942
12	12:66000001-66040000	12:66320001-66360000	2.71E-61	HiC	GSE87112	IMR90	intra	ENSG00000256083
12	12:66000001-66040000	12:66400001-66440000	2.25E-131	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000256259:ENSG00000213343
12	12:66000001-66040000	12:66360001-66400000	1.57E-104	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
12	12:66000001-66040000	12:66320001-66360000	2.93E-114	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000256083
12	12:66000001-66040000	12:66280001-66320000	1.82E-27	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000267942
12	12:66000001-66040000	12:66240001-66280000	3.64E-35	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000197301:ENSG00000211577
12	12:66000001-66040000	12:66200001-66240000	2.12E-38	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000241749:ENSG00000149948
12	12:66000001-66040000	12:66160001-66200000	1.49E-12	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan

12	12:66000001-66040000	12:66080001-66120000	3.55E-32	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
12	12:66000001-66040000	12:66040001-66080000	1.87E-45	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000213344
12	12:66000001-66040000	12:65960001-66000000	8.36E-11	HiC	GSE87112	IMR90	intra	ENSG00000256268
12	12:66000001-66040000	12:65920001-65960000	9.60E-96	HiC	GSE87112	IMR90	intra	ENSG00000255866
12	12:66000001-66040000	12:65880001-65920000	4.80E-31	HiC	GSE87112	IMR90	intra	nan
12	12:66000001-66040000	12:65840001-65880000	2.16E-21	HiC	GSE87112	IMR90	intra	nan
12	12:66000001-66040000	12:65800001-65840000	7.83E-67	HiC	GSE87112	IMR90	intra	ENSG00000215208
12	12:66000001-66040000	12:65760001-65800000	8.06E-45	HiC	GSE87112	IMR90	intra	nan
12	12:66000001-66040000	12:65720001-65760000	1.43E-18	HiC	GSE87112	IMR90	intra	nan
12	12:66000001-66040000	12:65680001-65720000	2.05E-95	HiC	GSE87112	IMR90	intra	nan
12	12:66000001-66040000	12:65600001-65640000	8.86E-22	HiC	GSE87112	IMR90	intra	nan
12	12:66000001-66040000	12:65560001-65600000	9.97E-12	HiC	GSE87112	IMR90	intra	ENSG00000174106
12	12:66000001-66040000	12:66480001-66520000	1.76E-61	HiC	GSE87112	IMR90	intra	nan
12	12:66000001-66040000	12:66400001-66440000	1.22E-70	HiC	GSE87112	IMR90	intra	ENSG00000256259:ENSG00000213343
12	12:66000001-66040000	12:66360001-66400000	2.63E-41	HiC	GSE87112	IMR90	intra	nan
12	12:66000001-66040000	12:65800001-65840000	2.70E-26	HiC	GSE87112	hESC	intra	ENSG00000215208
12	12:66000001-66040000	12:65920001-65960000	8.32E-15	HiC	GSE87112	hESC	intra	ENSG00000255866
19	19:46200001-46240000	19:46880001-46920000	1.79E-08	HiC	GSE87112	Left_Ventricle	intra	ENSG00000268810:ENSG00000269151:ENSG00000169515
19	19:46200001-46240000	19:46040001-46080000	1.58E-08	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	nan
19	19:46200001-46240000	19:46720001-46760000	5.28E-08	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000269637:ENSG00000269745:ENSG00000188293
19	19:46200001-46240000	19:46600001-46640000	6.70E-17	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000267922:ENSG00000268053:ENSG00000188624
19	19:46200001-46240000	19:46280001-46320000	9.80E-12	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000104936:ENSG00000268434:ENSG00000185800:ENSG00000104941
19	19:46200001-46240000	19:46240001-46280000	1.60E-37	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000259605:ENSG00000177045:ENSG00000267395
19	19:46200001-46240000	19:46160001-46200000	9.84E-20	HiC	GSE87112	IMR90	intra	ENSG00000010310:ENSG00000207773:ENSG00000125743:ENSG00000011478
19	19:46200001-46240000	19:46040001-46080000	1.78E-10	HiC	GSE87112	IMR90	intra	nan
19	19:46200001-46240000	19:45960001-46000000	3.42E-12	HiC	GSE87112	IMR90	intra	ENSG00000012061:ENSG00000125740:ENSG00000125744:

								ENSG00000213889
19	19:46200001-46240000	19:45880001-45920000	2.32E-07	HiC	GSE87112	IMR90	intra	ENSG00000104881:ENSG00000117877
19	19:46200001-46240000	19:46160001-46200000	5.38E-16	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000010310:ENSG00000207773:ENSG00000125743:ENSG00000011478
19	19:46200001-46240000	19:47240001-47280000	1.78E-12	HiC	GSE87112	IMR90	intra	ENSG00000090372:ENSG00000222614:ENSG00000181027
19	19:46200001-46240000	19:46280001-46320000	1.21E-30	HiC	GSE87112	IMR90	intra	ENSG00000104936:ENSG00000268434:ENSG00000185800:ENSG00000104941
19	19:46200001-46240000	19:46240001-46280000	2.20E-20	HiC	GSE87112	IMR90	intra	ENSG00000259605:ENSG00000177045:ENSG00000267395
19	19:46200001-46240000	19:45960001-46000000	3.18E-11	HiC	GSE87112	GM12878	intra	ENSG00000012061:ENSG00000125740:ENSG00000125744:ENSG00000213889
19	19:46200001-46240000	19:46280001-46320000	2.12E-09	HiC	GSE87112	GM12878	intra	ENSG00000104936:ENSG00000268434:ENSG00000185800:ENSG00000104941
19	19:46200001-46240000	19:46240001-46280000	2.02E-09	HiC	GSE87112	GM12878	intra	ENSG00000259605:ENSG00000177045:ENSG00000267395
19	19:46200001-46240000	19:46040001-46080000	3.77E-08	HiC	GSE87112	Left_Ventricle	intra	nan
19	19:46200001-46240000	19:45960001-46000000	1.55E-16	HiC	GSE87112	Left_Ventricle	intra	ENSG00000012061:ENSG00000125740:ENSG00000125744:ENSG00000213889
19	19:46200001-46240000	19:47360001-47400000	3.25E-09	HiC	GSE87112	Left_Ventricle	intra	nan
19	19:46200001-46240000	19:47240001-47280000	9.21E-14	HiC	GSE87112	Left_Ventricle	intra	ENSG00000090372:ENSG00000222614:ENSG00000181027
19	19:46200001-46240000	19:46360001-46400000	3.44E-16	HiC	GSE87112	IMR90	intra	ENSG00000125755:ENSG00000170608:ENSG00000170604
19	19:46200001-46240000	19:46240001-46280000	1.76E-08	HiC	GSE87112	Mesendoderm	intra	ENSG00000259605:ENSG00000177045:ENSG00000267395
19	19:39240001-39280000	19:39760001-39800000	3.64E-07	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000182393
19	19:39240001-39280000	19:39720001-39760000	9.27E-10	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000268510:ENSG00000197110:ENSG00000272395:ENSG00000213926:ENSG00000272311:ENSG00000183709
19	19:39240001-39280000	19:39400001-39440000	2.12E-29	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000262484:ENSG00000104835:ENSG00000269547:ENSG00000128626:ENSG00000269486
19	19:39240001-39280000	19:39840001-39880000	2.11E-07	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000241983
19	19:39240001-39280000	19:39000001-39040000	4.22E-17	HiC	GSE87112	GM12878	intra	ENSG00000269445:ENSG00000268055
19	19:39240001-39280000	19:39280001-39320000	3.45E-25	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000178934:ENSG00000207296:ENSG00000171747:ENSG00000268756
19	19:39240001-39280000	19:39400001-39440000	3.98E-12	HiC	GSE87112	hESC	intra	ENSG00000262484:ENSG00000104835:ENSG00000269547:ENSG00000128626:ENSG00000269486

19	19:39240001-39280000	19:39280001-39320000	1.76E-08	HiC	GSE87112	Mesendoderm	intra	ENSG00000178934:ENSG00000207296:ENSG00000171747:ENSG00000268756
19	19:39240001-39280000	19:39320001-39360000	4.23E-08	HiC	GSE87112	Mesendoderm	intra	ENSG00000104823:ENSG00000268083:ENSG00000104824:ENSG00000269688
19	19:39240001-39280000	19:39400001-39440000	1.28E-15	HiC	GSE87112	Mesendoderm	intra	ENSG00000262484:ENSG00000104835:ENSG00000269547:ENSG00000128626:ENSG00000269486
19	19:39240001-39280000	19:39400001-39440000	7.08E-11	HiC	GSE87112	GM12878	intra	ENSG00000262484:ENSG00000104835:ENSG00000269547:ENSG00000128626:ENSG00000269486
19	19:39240001-39280000	19:39400001-39440000	1.42E-08	HiC	GSE87112	Left_Ventricle	intra	ENSG00000262484:ENSG00000104835:ENSG00000269547:ENSG00000128626:ENSG00000269486
19	19:39240001-39280000	19:42040001-42080000	4.16E-08	HiC	GSE87112	Left_Ventricle	intra	ENSG00000270164:ENSG00000268027:ENSG00000007129
19	19:39240001-39280000	19:39320001-39360000	1.63E-16	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000104823:ENSG00000268083:ENSG00000104824:ENSG00000269688
19	19:39240001-39280000	19:39880001-39920000	5.70E-07	HiC	GSE87112	IMR90	intra	ENSG00000006712:ENSG00000063322:ENSG00000128016:ENSG00000266559:ENSG00000090924:ENSG00000269410
19	19:39240001-39280000	19:39000001-39040000	1.43E-33	HiC	GSE87112	Mesendoderm	intra	ENSG00000269445:ENSG00000268055
19	19:39240001-39280000	19:39320001-39360000	2.02E-20	HiC	GSE87112	IMR90	intra	ENSG00000104823:ENSG00000268083:ENSG00000104824:ENSG00000269688
19	19:39240001-39280000	19:39840001-39880000	1.41E-07	HiC	GSE87112	IMR90	intra	ENSG00000241983
19	19:39240001-39280000	19:39600001-39640000	4.55E-07	HiC	GSE87112	IMR90	intra	ENSG00000130669:ENSG00000269172
19	19:39240001-39280000	19:39400001-39440000	7.56E-13	HiC	GSE87112	IMR90	intra	ENSG00000262484:ENSG00000104835:ENSG00000269547:ENSG00000128626:ENSG00000269486
19	19:39240001-39280000	19:39000001-39040000	9.14E-12	HiC	GSE87112	hESC	intra	ENSG00000269445:ENSG00000268055
19	19:39240001-39280000	19:36320001-36360000	2.59E-11	HiC	GSE87112	Left_Ventricle	intra	ENSG00000161270:ENSG00000126259:ENSG00000105290
19	19:39240001-39280000	19:38760001-38800000	6.40E-08	HiC	GSE87112	Mesendoderm	intra	ENSG00000267748:ENSG00000200209:ENSG00000167644
19	19:39240001-39280000	19:38720001-38760000	2.11E-07	HiC	GSE87112	Mesendoderm	intra	ENSG00000011332:ENSG00000167642:ENSG00000167641
19	19:39240001-39280000	19:39000001-39040000	8.36E-12	HiC	GSE87112	IMR90	intra	ENSG00000269445:ENSG00000268055
19	19:39240001-39280000	19:39000001-39040000	5.85E-53	HiC	GSE87112	Mesenchymal_Stem_Cell	intra	ENSG00000269445:ENSG00000268055

Supplementary Table 3. Genes significantly associated with MVP in gene-based association tests (GWGAS) using MAGMA.

Note: GWGAS tests were performed in MAGMA and represent the aggregate effect of all SNPs in a gene (defined by GRCh37 reference genome locations). Bonferroni threshold of significance for 18,456 genes was 2.709e-06. Here we showed genes that have P-value < 10-4 . Position in the table is B37.

CHR	Start BP	End BP	NSNPS	NPARAM	Z Score	P	Gene name
17	2206677	2228554	33	8	5.6229	9.39E-09	SRR
19	46236509	46267792	48	9	5.5084	1.81E-08	AC074212.3
19	46213887	46234162	23	9	5.3763	3.80E-08	FBXO46
17	48712138	48769613	133	32	5.0849	1.84E-07	ABCC3
8	9911778	10286401	1569	69	5.0298	2.45E-07	MSRA
19	46282695	46289231	11	4	5.0269	2.49E-07	AC011530.4
1	218519577	218617961	222	39	4.94	3.91E-07	TGFB2
19	46286205	46296060	21	7	4.9244	4.23E-07	DMWD
19	46268043	46272484	5	2	4.8046	7.75E-07	SIX5
19	46272975	46285810	15	4	4.7579	9.78E-07	DMPK
19	46298968	46318577	65	12	4.5962	2.15E-06	RSPH6A
17	1963133	2207065	631	23	4.5957	2.16E-06	SMG6
11	2323227	2339430	65	9	4.5254	3.01E-06	TSPAN32
2	42396490	42559688	474	21	4.4908	3.55E-06	EML4

3	57994127	58157982	392	30	4.338	7.19E-06	FLNB
14	64970430	65000408	95	10	4.2589	1.03E-05	ZBTB1
14	64932217	64941221	12	5	4.2506	1.07E-05	AKAP5
3	189674517	189840226	590	31	4.2247	1.20E-05	LEPREL1
20	20033158	20341346	995	59	4.204	1.31E-05	C20orf26
7	120628731	120937498	761	31	4.1327	1.79E-05	CPED1
3	182840001	182881627	168	20	4.0986	2.08E-05	LAMP3
14	64915824	64971931	134	11	4.0977	2.09E-05	ZBTB25
1	53971910	54199877	599	38	4.0611	2.44E-05	GLIS1
4	3443614	3451211	35	12	3.9885	3.32E-05	HGFAC
10	99079022	99081672	6	2	3.9662	3.65E-05	FRAT1
2	178477720	178483694	38	7	3.9472	3.95E-05	TTC30A
1	6694228	6761984	140	16	3.9416	4.05E-05	DNAJC11
10	112404155	112599227	673	57	3.9362	4.14E-05	RBM20
1	218683438	218699320	58	4	3.935	4.16E-05	C1orf143
17	2225797	2240801	22	7	3.9272	4.30E-05	TSR1
17	1367392	1396106	86	26	3.9243	4.35E-05	MYO1C
15	67356101	67487533	385	52	3.892	4.97E-05	SMAD3
3	15602211	15643338	67	15	3.861	5.65E-05	HACL1
2	42560686	42652228	187	18	3.8608	5.65E-05	COX7A2L
3	57231944	57260549	42	8	3.8462	6.00E-05	HESX1

3	66429221	66551687	299	30	3.8253	6.53E-05	LRIG1
3	53880607	53899827	65	13	3.8064	7.05E-05	IL17RB
7	2443223	2474242	50	7	3.7991	7.26E-05	CHST12
11	116648436	116658766	29	8	3.7943	7.40E-05	ZNF259
17	41717756	41739322	66	10	3.7901	7.53E-05	MEOX1
3	196366646	196462878	380	66	3.7646	8.34E-05	PIGX
17	21279509	21323179	58	17	3.7572	8.59E-05	KCNJ12
1	25664408	25688852	23	5	3.749	8.88E-05	TMEM50A
15	85198360	85201794	12	3	3.7256	9.74E-05	NMB

Appendix C

Article

Katelynn Toomer*, **Mengyao Yu*,** Diana Fulmer*, Lilong Guo, Kelsey Moore, Reece Moore, Ka'la Drayton, Janiece Glover, Neal Peterson, Sandra Ramos-Ortiz, Alex Drohan, Breiona J. Catching, Rebecca Stairley, Andy Wessels, Joshua H. Lipschutz, Francesca N. Delling, Xavier Jeunemaitre, Christian Dina, Ryan L. Collins, Harrison Brand, Michael E. Talkowski, Federica del Monte, Rupak Mukherjee, Alexander Awgulewitsch, Simon Body, Gary Hardiman, Starr E. Hazard, Willian Da Silveira, Baolin Wang, Maire Leyne, Ronen Durst, Roger Markwald, Solena Le Scouarnec, Albert Hagege, Thierry Le Tourneau, Peter Kohl, Eva Rog-Zielinska, Patrick T. Ellinor, Robert A. Levine, David Milan, Jean-Jacques Schott,Nabila Bouatia-Naji, Susan Slaugenhaupt, Russell A. Norris. **Primary Cilia Defects Cause Mitral Valve Prolapse**; Science Translational Medicine; accepted in principle.

Introduction and my contribution

Primary cilia are microtubule structure, which coordinates signaling pathways, such as WNT (Corbit et al. 2008; Hulin et al. 2017), Hedgehog (Liu et al. 2005) and TGF-beta (Clement et al. 2013) to respond to the external environment. In autosomal dominant polycystic kidney disease (ADPKD), mitral valve prolapsed (MVP) was found in 26% of patients, a ten-fold increase above background (Lumiaho et al. 2001). The comorbidity of MVP and ADPKD, together with the pathological similarity of the disease phenotype suggest that MVP might be caused by the defects of cilia genes in certain individuals.

Using our previous GWAS on MVP (Dina et al. 2015), we tested for a global enrichment for significant association among a large list including 278 different genes reported as involved in cilia biology, mainly as harboring rare mutations in a diverse panel of ciliopathies. We used the SNP test ratio method (O'dushlaine et al. 2009)to compare the association results from MVP GWAS with association results from 1000 simulated non-MVP GWAS. Those simulated GWAS share the same genotype with the MVP GWAS, but using random phenotype. Then, according to a pre-defined p-value threshold, the method counted the number of times the pathway tested

included a higher ratio of significant to non-significant SNPs, compared to the 1000 simulated GWAS. As recommended for this method, we used the KEGG pathways catalog, in addition to our customized list of cilia genes. Given the limited number of highly associated SNPs in the current GWAS that includes ~1500 cases and 2500 controls, we set significance at P-value<0.05 to allow more SNPs to contribute to the analyses. We also applied i-GSEA4GWAS (Zhang et al. 2010; Zhang et al. 2015) using 6.6M SNPs from the previous meta-analysis as input. In this pathway analysis, we added the 278 known cilia ciliopathy genes as a pre-defined gene set (Cilia Gene Set). Threshold FDR<0.05 was considered as statistically significant.

Among nominally associated SNPs in the GWAS (SNPs<0.05), we found that SNPs that map in the cilia genes list are more likely (trend P=0.03) to be enriched for associated SNPs with MVP using SRT. Of note, cardiac enriched pathways were highlighted in the top 10 enriched pathways, including hypertrophic cardiomyopathy, dilated cardiomyopathy and cardiac muscle contractionwere (Table 1). The significant enrichment of MVP-associated variants in this cilia gene set (p=0.009, FDR= 0.024) was found using i-GSEA4GWAS as well (Table 2). Those pathway analyses confirmed that primary cilia genes are involved in MVP. Besides, taking advantage of previous GWAS analysis, we found that variants in the ±500 kb of *DZIP1* did not show association with MVP. However, a deleterious missense mutation in *DZIP1* was related to the impaired ciliogenesis during development and resulted in adult myxomatous valve disease.

Table 1. Enriched pathways for nominally associated common genetic variants in MVP GWAS. We provide the pathway ID according to the KEGG database nomenclature, the number of time the 1000 simulated GWAS with permuted phenotypes provided a ratio of significant to non-significant SNPs in this pathway higher than the real data, an empirical p-value of the ratio comparison, and the descriptive of the pathway tested. Ex: for the cilia enriched pathway, for 1000 simulated GWAS and one GWAS based on real phenotypes, there were 34 times where the simulated data provided higher rations, and only once for the most enriched pathway phagosome.

Pathway ID in KEGG	Ratio significant to non-significant SNPs higher in simulated data	Total number of analyses	Empirical p-value	Descriptive of the pathway
hsa04145	1	1001	0.001	Phagosome
hsa05416	1	1001	0.001	Viral myocarditis
hsa05410	2	1001	0.002	Hypertrophic cardiomyopathy (HCM)
hsa05414	2	1001	0.002	Dilated cardiomyopathy
hsa00190	3	1001	0.003	Oxidative phosphorylation
hsa05340	3	1001	0.003	Primary immunodeficiency
hsa00010	4	1001	0.004	Glycolysis / Gluconeogenesis
hsa04612	4	1001	0.004	Antigen processing and presentation
hsa00051	5	1001	0.005	Fructose and mannose metabolism
hsa04260	5	1001	0.005	Cardiac muscle contraction
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
Custumized	34	1001	0.034	Manually curated cilia gene list

Table 2. Details about Cilia Gene set enrichment results using i-GSEA4GWAS. The gene IDs and their best associated variants are listed for the 178 significant and 97 non-significant genes from the Cilia Gene set. Indication of significance is provided as -log (P-value) and genes are ranked from the most to the least significant. Two genes (CP110 and PARD6A) had no variant tested in the GWAS and were not included in the analysis.

Pathway/Gene Set Name			Cilia Gene set		
Description			Manually curated set of genes inovlved in cilia biology		
P-value			0.009		
FDR			0.024		
Significant genes/Selected genes/All genes			178/275/277		
Significant Gene Names (N=178)	**Variant**	**-log(P-value)**	**Non-Significant Gene Names (N=97)**	**Variant**	**-log(P-value)**
GPR98	rs1995775	4.049781	SLC47A2	rs1989379	1.3541929
BBS9	rs75554572	3.70969	IFT27	rs11705490	1.3505613
RFX3	rs1333954	3.3721104	OFD1	rs5979958	1.3409394
RP1	rs369240	3.3407624	HTT	rs10012793	1.3364478
PCDH15	rs10763225	3.316575	DNAH2	rs113126700	1.3325487
CDH23	rs66812897	3.242835	HAP1	rs7216154	1.3306905
TTC29	rs12511630	3.2411432	CNGA4	rs11040813	1.3126054
SNX10	rs6950206	3.0700517	MKKS	rs6039921	1.306387
TRIP11	rs113732705	3.0549433	RP2	rs5905568	1.2997841
TTC30B	rs185156233	3.0468824	TULP1	rs2075611	1.298075
INVS	rs75959792	3.0293999	RSPH4A	rs784129	1.2884156
GLIS2	rs113679768	3.0171063	MCHR1	rs111950677	1.266455
BBS5	rs4667602	2.9528613	KIF3A	rs17691077	1.2632599
LCA5	rs72897736	2.8857617	DNAI2	rs3803792	1.2491696
ZNF423	rs188984258	2.8669262	DRD1	rs35853538	1.2451608
MDM1	rs1038163	2.82366	CEP41	rs10259022	1.2427897

DYNC2H1	rs1913462	2.8134785	KIF19	rs4789650	1.2416095
TCTN3	rs34808745	2.7765253	STOML3	rs9548560	1.2362481
TRAPPC10	rs9974857	2.7493951	ORC1	rs78572338	1.234462
NUP214	rs981759	2.7453704	RAB8A	rs78696309	1.2154264
NPHP4	rs709207	2.70061	NEK8	rs6505092	1.2153586
TTC30A	rs3821003	2.6782377	USH1C	rs72870313	1.2148414
DNAH6	rs35702893	2.673219	CCDC114	rs7258665	1.2075931
DYX1C1	rs8038410	2.667753	DNALI1	rs4653302	1.2042971
C8orf37	rs184837810	2.6035542	DNAI1	rs10972130	1.1905338
PKD1L1	rs78856648	2.5952675	TTLL6	rs7217007	1.1807418
MLF1	rs6784511	2.5745838	TUBA1C	rs10875943	1.1749096
WDPCP	rs6754628	2.5582101	DCDC2	rs10946691	1.1699665
B9D2	rs1982072	2.541348	TPPP2	rs17242881	1.164425
TTC8	rs61985281	2.537357	POC1A	rs12485855	1.1551374
TTK	rs239507	2.5270567	TCTN1	rs1106724	1.144031
EVC2	rs114248858	2.5157902	CBY1	rs117392630	1.1349158
TMEM237	rs7602598	2.489981	CTNNB1	rs2114286	1.1335202
DNAH10	rs116918834	2.483634	ALMS1	rs6546833	1.131468
TTC12	rs723079	2.4723976	PCM1	rs6989977	1.1190462
TMEM231	rs80329655	2.4712908	RABL5	rs35710113	1.1185158
TUBB3	rs4586435	2.4708242	WDR60	rs76778395	1.0704955
SEPT2	rs55821292	2.434298	TUBGCP3	rs1952109	1.0431881
PDZD7	rs11190788	2.4339464	TUBGCP6	rs5771238	1.0393614
AZI1	rs113931593	2.4140873	B9D1	rs9899649	1.0303645
SPEF2	rs10078369	2.4086246	EFHC1	rs9382109	0.982896
PKD1	rs1184357	2.4075742	RILPL1	rs13303364	0.9793965
DNAH11	rs11763291	2.3981025	IFT88	rs2149769	0.97247845
TTC21B	rs57608819	2.3849294	CEP72	rs76242473	0.96496993
CNGB1	rs16959471	2.377947	WDR35	rs13006748	0.96315026

NPHP3	rs73220112	2.369589	PIBF1	rs7990365	0.95301425
FAM161A	rs193292249	2.3501654	NME7	rs12060959	0.9527547
CCDC103	rs17628574	2.3436372	TTLL3	rs145299497	0.9492566
IFT80	rs7614701	2.328654	TBC1D7	rs1899262	0.94398576
DFNB31	rs17807115	2.323294	GLI1	rs10876987	0.932976
RSPH1	rs28635728	2.3124645	DVL1	rs150789461	0.9082146
NEK2	rs12749444	2.2968633	IFT122	rs2260840	0.9051855
VHL	rs79743995	2.2955236	CLDN2	rs6523903	0.9047715
PTCH1	rs111896098	2.2825189	BBS7	rs1507994	0.8992867
RTTN	rs3809953	2.2793326	TTC26	rs2883569	0.889214
TRIM32	rs2281627	2.2396812	TMEM138	rs10897156	0.88810146
EVC	rs61162715	2.233459	SSTR3	rs5756557	0.88596445
RAN	rs872396	2.23293	NUP37	rs113200478	0.8758896
PARD3	rs4315011	2.2260487	RANBP1	rs72490649	0.8743578
DNAAF1	rs11643042	2.2148767	FOPNL	rs215037	0.8740307
DISC1	rs9431730	2.1863098	BBS12	rs56338261	0.8542423
CEP97	rs62284187	2.1775026	USH1G	rs690330	0.84583205
LZTFL1	rs1985463	2.1711826	CEP89	rs4570989	0.84547806
HYLS1	rs11220256	2.1664858	KIF7	rs72750756	0.82271844
RAB3IP	rs112609629	2.1537206	DYNLT1	rs3799195	0.8217913
FLNA	rs191469451	2.1343029	ARL3	rs7072001	0.81812966
USH2A	rs4462116	2.1340833	LRRC6	rs8180952	0.8046078
TRAPPC9	rs4736126	2.133094	FUZ	rs62129186	0.78709966
RPGR	rs5963396	2.131396	TUBB2B	rs4129373	0.77714086
RAB23	rs2064245	2.126413	EXOC5	rs8010007	0.76761526
PHF17	rs77865637	2.1212392	SSNA1	rs66855886	0.75595665
HYDIN	rs3094452	2.115014	RILPL2	rs112097626	0.74851614
IFT172	rs78590292	2.0963209	TRAPPC3	rs6425965	0.7479577
HSPA8	rs74759868	2.090998	FBF1	rs11867538	0.732363

GSK3B	rs62264713	2.0868633	FOXJ1	rs1868824	0.73167104
KIF17	rs875015	2.082579	BBS4	rs3850988	0.7105325
SUFU	rs186502091	2.075622	IFT46	rs193264023	0.7009387
GLI2	rs55922034	2.0660746	BBS1	rs1791683	0.6990858
RSPH3	rs5015630	2.0444338	ARL6	rs9854857	0.6648377
CEP104	rs10752721	2.0436347	RPGRIP1L	rs113885317	0.66292
RAB11A	rs7181649	2.0315852	INPP5E	rs73566945	0.65943
TTBK2	rs9635326	2.0228605	SASS6	rs17121872	0.6456149
ARF4	rs62258960	2.02263	CEP290	rs2468240	0.6090315
CCDC41	rs77626725	2.013073	IFT57	rs139232028	0.5354753
KIF3B	rs6141286	2.0051498	CCDC28B	rs2295116	0.50569487
CNGA2	rs5925036	1.9974034	HSPB11	rs77840912	0.50068825
IQCB1	rs12492861	1.9944699	TMEM216	rs72929133	0.500389
TMEM67	rs55680024	1.9739156	NOTO	rs74454024	0.49999064
IFT43	rs10146195	1.9508908	PLK1	rs12149199	0.48594922
DNAH5	rs78628444	1.9341656	HTR6	rs12691495	0.4771013
MAL	rs146162203	1.9002173	VDAC3	rs75465246	0.44739956
HNF1B	rs112055772	1.8919181	TUBE1	rs17073310	0.38248608
SHH	rs74319244	1.8896863	ODF2	rs10117399	0.3673041
OCRL	rs3118108	1.8826948	TUBA1A	rs75588464	0.3395469
KIF24	rs17350261	1.8700527	CRB3	rs348359	0.30559033
STK38L	rs2881297	1.8685547	DRD5	rs6283	0.15395424
TUBB2A	rs9378771	1.8662341	IFT20	rs16964224	0.067641765
MYO7A	rs6592708	1.8478224	**Non mapped Genes (N=2)**		
SNAP25	rs3025873	1.8466988	CP110		
NPHP1	rs72826897	1.8455333	PARD6A		
CCDC37	rs79828270	1.8388326			
BBS10	rs4271425	1.8361943			
PDE6D	rs4973479	1.8239392			

EXOC6	rs7922067	1.8204021			
EXOC4	rs7800830	1.8171909			
SYNE2	rs2357158	1.8005353			
CCDC39	rs58980055	1.7903215			
NGFR	rs11466155	1.7893227			
MKS1	rs12051662	1.7860507			
NEK4	rs17477867	1.7841814			
TUBGCP2	rs2860672	1.7815745			
ADCY3	rs3198665	1.7696272			
GPR161	rs16828104	1.7688429			
GLI3	rs13438132	1.7663101			
PKD2	rs2728121	1.7512183			
WDR19	rs6852518	1.740615			
INTU	rs6854119	1.7373399			
MAK	rs145807088	1.7305803			
XPNPEP3	rs28417895	1.7252613			
CROCC	rs1755332	1.7010155			
EXOC3	rs1053247	1.7004834			
NINL	rs151302894	1.6641428			
RPGRIP1	rs56185443	1.6526235			
HEATR2	rs79603951	1.6522979			
TNPO1	rs3010278	1.6514411			
SMO	rs6962740	1.6505326			
STIL	rs10890472	1.6498666			
DPYSL2	rs13276720	1.6441295			
CLUAP1	rs115049371	1.643728			
ASAP1	rs6983862	1.6414963			
NEK1	rs10019795	1.6371351			
DPCD	rs11191054	1.6360129			

SPATA7	rs79410100	1.6296213			
C2CD3	rs10898958	1.627693			
BBS2	rs9935360	1.6163777			
IFT140	rs62012851	1.6144943			
TTLL9	rs6087789	1.6120064			
AHI1	rs7772778	1.6119034			
TUBGCP5	rs140870563	1.6094668			
SDCCAG8	rs12134747	1.5920515			
RAB17	rs75215977	1.5864794			
PTPDC1	rs7023944	1.5836576			
TBC1D30	rs612053	1.57017			
TRAF3IP1	rs186555991	1.5658567			
NUP62	rs12977375	1.5647361			
NIN	rs8011819	1.5609293			
RSPH9	rs111978957	1.5589633			
C2orf71	rs56408933	1.5570436			
NUP35	rs113083714	1.5548288			
PKHD1	rs13219549	1.5448289			
DNAL1	rs17782112	1.5274962			
DRD2	rs61902832	1.5239567			
IFT52	rs6017133	1.5182192			
DNAAF2	rs8952	1.5143787			
STX3	rs484083	1.5120811			
TUBGCP4	rs2012467	1.4972718			
SCLT1	rs112160201	1.4902138			
CCDC40	rs11654594	1.484137			
DNAAF3	rs61618136	1.4809886			
MYO15A	rs2056841	1.4785907			
CENPJ	rs7331734	1.4712069			

EXOC6B	rs13003935	1.4675277			
RAB11FIP3	rs7196963	1.4585929			
IFT74	rs12236850	1.4554797			
MAPRE1	rs6141824	1.4437467			
NUP93	rs4784727	1.4413549			
WDR78	rs1925407	1.4359096			
C21orf2	rs2070572	1.4348238			
TCTN2	rs34705855	1.4325618			
TUBA4A	rs72955459	1.4318526			
CC2D2A	rs7697314	1.4315405			
TULP3	rs2907606	1.4312627			
KIF3C	rs62129488	1.425333			
SEPT7	rs345373	1.4168583			
IFT81	rs7308785	1.4129051			
TOPORS	rs15014	1.4121599			
ARL13B	rs62265480	1.4108893			
AK7	rs186252163	1.3932661			