Integrative tissue-specific functional annotations in the human genome provide novel insights on many complex traits and improve signal prioritization in genome wide association studies

Qiongshi Lu[1,#], Ryan L Powles[2,#], Qian Wang[2], Beixin J He[3], Hongyu Zhao[1,2*]

[1] Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

[2] Program of Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA

[3] Division of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

# These authors contributed equally to this work.

* To whom correspondence should be addressed:

Dr. Hongyu Zhao
Department of Biostatistics
Yale School of Public Health
60 College Street,
New Haven, CT, 06520, USA
hongyu.zhao@yale.edu

Abstract

Extensive efforts have been made to understand genomic function through both experimental and computational approaches, yet proper annotation still remains challenging, especially in non-coding regions. In this manuscript, we introduce GenoSkyline, an unsupervised learning framework to predict tissue-specific functional regions through integrating high-throughput epigenetic annotations. GenoSkyline successfully identified a variety of non-coding regulatory machinery including enhancers, regulatory miRNA, and hypomethylated transposable elements in extensive case studies. Integrative analysis of GenoSkyline annotations and results from genome-wide association studies (GWAS) led to novel biological insights on the etiologies of a number of human complex traits. We also explored using tissue-specific functional annotations to prioritize GWAS signals and predict relevant tissue types for each risk locus. Brain and blood-specific annotations led to better prioritization performance for schizophrenia than standard GWAS p-values and non-tissue-specific annotations. As for coronary artery disease, heart-specific functional regions was highly enriched of GWAS signals, but previously identified risk loci were found to be most functional in other tissues, suggesting a substantial proportion of still undetected heart-related loci. In summary, GenoSkyline annotations can guide genetic studies at multiple resolutions and provide valuable insights in understanding complex diseases. GenoSkyline is available at http://genocanyon.med.yale.edu/GenoSkyline.

Introduction

Functionally annotating the human genome is a major goal in human genetics research. After years of community efforts, a variety of experimental and computational approaches have been developed and applied for genomic functional annotation. Comparative genomics studies have shown that approximately 4.5% of the human genome is conserved across mammals[1]. Furthermore, the rich collection of epigenomic data generated by large consortia (e.g. ENCODE[2] and Epigenomics Roadmap Project[3]) also provides great insight for understanding the functional effects of the genome, especially in terms of non-coding regulatory machinery. To best utilize these rich data, we recently developed GenoCanyon[4], a non-coding functional prediction approach based on integrative analysis of annotation data, whose performance was demonstrated through predicting well-studied regulatory DNA elements. GenoCanyon provides general predictions of non-coding functional regions in the human genome but does not fully utilize cell-type-specific information of epigenomic data. Incorporating cell-type-specific or tissue-specific information into annotation tools is essential not only for understanding the basic biology of the genome, but also for better characterizing genetic variation, as in the functional interpretation of risk loci identified from genome-wide association studies (GWAS).

GWAS has been a great success in the past decade, yet challenges still remain in both identifying additional risk variants and interpreting GWAS results. Current practice employs a significance threshold (i.e. 5×10^{-8}) that controls family-wise error rate. Yet this approach is known to be underpowered when effect sizes are weak or moderate at risk loci[5]. Moreover, nearly 90% of the genome-wide significant hits in published GWAS are located in non-coding regions whose functional impact to human complex traits is largely unknown[6]. Complex linkage disequilibrium (LD) patterns also hinder our ability to identify real functional sites among correlated SNPs. Several methods have been proposed to integrate annotation data for better prioritizing GWAS signals and their effectiveness has also been well demonstrated[7-10]. Tissue-specific functional annotations have the potential to bring even more biological insights to post-GWAS analysis and help understand complex disease etiology.

In this paper, we introduce GenoSkyline, a tissue-specific functional prediction tool based on integrated analysis of epigenomic annotation data. We demonstrate its ability to identify tissue-specific functionality from its performance to rediscover a number of experimentally validated non-coding elements. Next, we show valuable biological insights GenoSkyline can provide in post-GWAS analysis through integrative analysis of 15 human complex traits. We believe that GenoSkyline will prove to be a powerful tool for human genetics research because of its abilities to assess tissue-specific enrichment of GWAS signals, better prioritize GWAS signals, and offer biological interpretations of risk loci.

Results

Predicting tissue-specific functional regions in the human genome

The posterior probability of being functional given the annotation data is used to measure tissue-specific functional potential of each nucleotide in the human genome (**Online Methods**). It will be referred to as GenoSkyline (GS) score in following sections. We calculated GS scores for 7 human tissue types; brain, gastrointestinal tract (GI), lung, heart, blood, muscle, and epithelium (**Supplementary Table 1**). With a GS score cutoff of 0.5, 22.2% of the human genome is predicted to be functional in at least one of these tissue types, while 1.7% is functional in all 7 tissues (**Figure 1a**). Since GS score has a bimodal pattern, these results are not sensitive to cutoff choice (**Supplementary Notes**).

Across tissue types, the percentage of predicted functional genome ranges from 5.4% (Lung) to 9.7% (GI) (**Figure 1b** and **Supplementary Table 2**). The overlap between heart-specific and muscle-specific functional regions is the largest among all pairs of tissues. Interestingly, although the percentage of functional genome in blood (8.4%) is similar to other tissue types, it overlaps less with the functional regions in other tissues (**Figure 1c**). This is consistent with the recent discovery that blood has the lowest levels of eQTL sharing with other tissues[11].

Investigating the performance of tissue-specific functional annotations

Beta-globin gene complex

We now demonstrate GenoSkyline's ability to predict tissue-specific functionality using a variety of experimentally validated functional machinery. Beta-globin (*HBB*) gene complex is an extensively studied genomic region on chromosome 11, containing 6 genes and 23 cis-regulatory modules (CRMs) that are known to control both the timing and the spatial pattern of gene expression[12,13]. We compared GS scores for different tissue types in this region. Not surprisingly, blood-specific functionality was observed (**Figure 2a**). Among the 6 genes in this region, adult globin genes *HBB* and *HBD*, as well as pseudogene *HBBP1* are captured well by blood-specific GS scores (**Supplementary Table 3**). However, embryonically expressed *HBE1*, fetally expressed *HBG1* and *HBG2*, and the CRMs that regulate these genes have lower GS scores. This is possibly because 18 of the 24 cell lines used for developing blood-specific GS scores were acquired from adult samples (**Supplementary Table 1**). The mean blood-specific GS score in these genes increases from 0.388 to 0.704 after removing *HBG1*, *HBG2*, and *HBE1*. Similarly, a substantial boost in mean GS score is observed after removing CRMs regulating the embryonic and fetal globin genes (**Figure 2b**, **Supplementary Figure 1**). Compared with GenoCanyon, GenoSkyline provides less sensitive but highly specific functional predictions. Its ability of identifying tissue-specific functional coding and non-coding DNA elements has the potential to benefit diverse types of biological studies.

Tissue-specific enhancers

In vivo enhancers with tissue-specific activity in central nervous system (CNS; n=585), heart (n=96), and blood vessel (n=9) were downloaded from VISTA enhancer browser[14] (**Online Methods**). Mean GS scores for brain, heart, and blood tissues were calculated for each enhancer. Brain-specific and heart-specific GS scores were substantially higher in their respective enhancer categories compared to GS scores of non-relevant tissue types. Additionally, the mean blood-specific GS score also stands out for enhancers with observed activity in blood vessel despite the limited sample size (**Figure 2c**). In a separate study, 11 human-accelerated elements near the brain developmental transcription factor *NPAS3* have been identified to act as tissue-specific enhancers within the nervous system[15]. Brain-specific GS scores for these enhancers are substantially higher than those for other tissue types (**Figure 2d**), concurrent with previous results.

Regulatory miRNAs

Next, we test if GenoSkyline could also capture miRNAs expressed exclusively in certain tissue types. Liang et al. studied the tissue specific expression pattern of eight groups of miRNAs[16]. We extracted and annotated four groups (groups I, II, III+IVa, and V from Liang et al.) that could be represented by the currently available tissue types in GenoSkyline annotations. These four groups of miRNAs were found to be expressed preferentially in skeletal/cardiac muscle, organs lined with epithelium, brain/peripheral blood mononuclear cell (PBMC), and heart, respectively through unsupervised clustering. The most relevant tissue types suggested by GenoSkyline for these four groups are muscle/heart, GI/epithelium, brain, and heart, respectively (**Figure 2e**). Our results based on integrative analysis of epigenomic data are consistent with the tissue-specific expression pattern reported by Liang et al.

Inter-genic regulation of myosin heavy chain

We applied GenoSkyline to a validated biologic switch in cardiac development and disease. Myosin heavy chain (MHC) is the major contractile protein in human striated muscle[17]. Cardiac muscle cells primarily express two isoforms, alpha-MHC (*MYH6*) and beta-MHC (*MYH7*)[18]. The ratio of alpha-to-beta isoforms determines cardiac contractility and allows for effective response to a wide range of physiologic and pathologic stimuli[19]. Alpha-to-beta ratio decreases in cardiac diseased states[20,21], and reversal of this shift is associated with better clinical outcomes[22]. miRNAs can regulate alpha-to-beta isoform shift, and prior studies in rodents have outlined a network of crosstalk between intronically expressed miRNAs and their host muscle genes[23,24]. For instance, mir-208a, on an intron of *MYH6*, is a positive regulator of beta-MHC by targeting transcription factors that repress its expression[24]. GS scores for *MYH6* and mir-208a accurately reflect their cardiac-specific expression, whereas *MYH7* and mir-208b exhibit strong signals in both skeletal and cardiac tissue (**Figure 3a** and **Supplementary Table 4**). This

corresponds to known expression pattern of *MYH7* and mir-208b in slow twitch skeletal muscle fibers[17] as well as heart. We also explored tissue-specific functionality of two known distal enhancers of mir-208b identified on VISTA Enhancer Browser, hs2330 and hs1670. GS scores for hs2330 mirror *MYH7*/mir-208b signals. Interestingly, GS scores for hs1670, a distal enhancer flanking mir-208b, are also strong in nervous and GI tissue, a finding that agrees with its observed expression pattern in other tissues (based on VISTA Enhancer Browser data). Collectively, these results show that GenoSkyline can replicate the tissue-specific expression pattern of a complex inter-gene regulatory network.

Zone of polarizing activity regulatory sequence

GenoSkyline can also be generalized to identify tissue specificity outside of the 7 core categories discussed here, based on available experimental data. For example, Zone of polarizing activity regulatory sequence (ZRS), a well-studied developmental enhancer, is located in the fifth intron of *LMBR1* gene. Acting as an enhancer of *SHH*, ZRS has been shown to play a crucial role in limb development[25]. However, none of the seven tissue types in GenoSkyline suggest ZRS's functionality (**Supplementary Figure 2**). In order to see if ZRS could be identified using epigenomic data of other cell types, we extended GenoSkyline to two new groups of cells that are potentially important for development, embryonic stem cells (ESC) and fetal cells (**Supplementary Table 5**). Both ESC and fetal-cell-specific GS scores successfully identified ZRS with high resolution (**Figures 3b and 3c**). This example shows that GenoSkyline is a flexible framework. Researchers could develop their own cell-group-specific functional annotations if ChIP-seq data are available for the cells of interest.

Hypomethylated transposable elements

A recent study of genome-wide DNA methylation status identified tissue-specific hypomethylated transposable elements (TE) exhibiting enhancer activities[26]. We downloaded four groups of TEs that are hypomethylated in ESC H1, fetal brain/primary neural progenitor cells, adult breast epithelial cells, and PBMC/adult immune cells, respectively (**Online Methods**). Although DNA methylation data were not used for developing GenoSkyline, we were still able to provide highly consistent results, suggesting tissue-specific functionality of these TEs in ESC, brain, epithelium, and blood cell, respectively (**Figure 3d**).

Analyzing tissue-specific enrichment for 15 human complex traits

In the sections above, we demonstrated GenoSkyline's ability of identifying tissue-specific functional regions in the human genome. Next, we focus on how GenoSkyline could help us understand human complex traits. Finucane et al. recently proposed using LD score regression to partition heritability of complex traits by functional categories[27].

We applied LD score regression on 15 human complex diseases and traits (**Supplementary Table 6**), and calculated the tissue-specific enrichments using GenoSkyline annotations (**Online Methods**).

Our analysis successfully replicated some well-known findings and also provided novel insights to these complex traits (**Figure 4**; **Supplementary Figure 3**). For schizophrenia, enrichment in brain is much stronger than in any other tissue type ($p = 6.52\times10^{-26}$), while highly significant enrichment could be observed in heart ($p = 2.30\times10^{-7}$) and blood ($p = 1.65\times10^{-5}$) as well. Brain is also the most enriched tissue for anorexia nervosa ($p = 4.86\times10^{-2}$) despite the substantially weaker signal. For three autoimmune diseases (Crohn's disease, ulcerative colitis, and rheumatoid arthritis), the strongest enrichment was in blood. However, solid enrichment in GI could also be observed for both Crohn's disease and ulcerative colitis, but not rheumatoid arthritis. Sex-stratified summary statistics were available for two anthropometric traits - body mass index (BMI) and waist-hip ratio (WHR) adjusted for BMI[28,29]. Therefore, we performed gender-specific analyses for these two traits. Consistent with recently published results[27], brain possesses the strongest enrichment for BMI. Interestingly, the enrichment in brain is stronger in female samples ($p = 1.21\times10^{-8}$) than in male samples ($p = 2.39\times10^{-6}$), while epithelial tissue may play a more important functional role in male samples ($p = 1.34\times10^{-3}$ in males and 2.83×10^{-2} in females). Some patterns of gender-specific enrichment were also observed for WHR. GI is the dominant tissue for females ($p = 5.26\times10^{-5}$) but seems less important in male samples ($p = 2.95\times10^{-2}$), while enrichment in muscle is consistent between males and females.

It is worth noting that extra caution is needed when interpreting these enrichment results. For example, Finucane et al. reported connective/bone as the most enriched tissue type for human height[27], but GenoSkyline annotations for this tissue is not available at this moment due to incomplete epigenomic data (**Online Methods**). Similarly, we are not yet able to investigate the relationship between lipid traits and liver tissue because of the lack of tissue-relevant functionality data.

GWAS signal prioritization using tissue-specific functional annotations

We recently developed Genome Wide Association Prioritizer (GenoWAP), and showed that GWAS signals could be better prioritized through integrating GWAS summary statistics with GenoCanyon annotation[10]. From the results of tissue-specific enrichment analysis, it could be seen that some complex traits are strongly related to a few tissue types. In this section, we show that the performance of GWAS signal prioritization could be further improved through integrating GenoSkyline annotations of relevant tissue types.

Using both tissue-specific GS scores and GenoCanyon scores that quantify the overall functionality, we calculate the posterior probability $P(Z_D = 1, Z_T = 1|p)$ to measure the importance of each SNP. In this calculation, Z_D is the indicator of disease/trait-specific functionality, Z_T is the indicator of tissue-specific functionality, and p is the p-value

acquired from standard GWAS analysis (**Online Methods**). Psychiatric Genomics Consortium (PGC) has published two large GWAS meta-analyses for schizophrenia, a major psychiatric disorder. We applied our method to the smaller study[30] and attempted to replicate the findings of the larger study[31]. This analysis demonstrates GenoSkyline's ability to prioritize association signals that are more likely to be replicated in a larger sample. These two studies will be referred to as PGC2011 and PGC2014 studies in the following discussion.

Enrichment analysis suggests that brain is the most enriched of schizophrenia GWAS signals compared with other tissue types, and strong enrichment could also be observed in heart and blood (**Figure 4**). For each SNP in the PGC2011 study, mean GenoCanyon score of its surrounding region and mean GS scores of brain, blood, and heart tissues were calculated (**Online Methods**). SNPs in these tissue-specific functional regions and the SNPs in general functional regions are all enriched for associations with schizophrenia (**Figure 5a**; **Supplementary Figure 4**). Notably, tissue-specific functional regions are more enriched for associations with schizophrenia relative to general functional regions, with blood showing the strongest enrichment. It is also worth noting that non-functional regions are enriched of GWAS associations as well, most likely due to the LD between functional and non-functional SNPs[10].

Next, we define a new SNP-level metric for the tissue-specific GenoSkyline posterior (GSP) scores (i.e. $P(Z_D = 1, Z_T = 1|p)$) of brain, blood, and heart, as well as the non-specific functionality posterior (NSFP) scores (i.e. $P(Z_D = 1|p)$; see **Online Methods**) for each SNP in PGC2011 study. Enrichment analysis using GTEx whole-blood eQTLs[11] found that the top SNPs based on tissue-specific GSP scores are substantially more enriched of eQTLs than NSFP scores and p-values. As expected, blood GSP scores showed the strongest enrichment of whole-blood eQTLs (**Figure 5b**). When using a set of quantitative trait loci in human brain[32], tissue-specific GSP scores also showed superior performance, with the brain-specific scores dominating others as the number of top SNPs increase (**Figure 5c** and **Supplementary Figure 5**).

A total of 108 schizophrenia-associated loci were identified in the PGC2014 study. We removed three loci on chromosome X due to the absence of SNPs on sex chromosomes in the PGC2011 dataset. All the SNPs in the PGC2011 study were ranked based on their p-values, NSFP scores, and tissue-specific GSP scores, respectively (**Supplementary Table 7**). The maximum ranks at each of the 105 schizophrenia-associated loci based on these different criteria were then compared (**Supplementary Table 8**). Brain GSP score showed better performance in prioritizing these loci when compared with p-value. Sixty-seven out of 105 loci had an increased rank (p-value=0.003, one-sided binomial test). The performance of heart GSP score was slightly worse than brain-specific score, but still better than p-value ranking. Blood GSP score showed comparable performance with p-value ranking. Notably, the performance of brain and heart GSP scores was still significantly better than NSFP score, although NSFP score outperforms ranking based on p-value.

Tissue-specific functional annotations could provide even deeper insight when prioritizing SNPs locally at risk loci. The schizophrenia-associated locus on chromosome 8q21 is located in the intergenic region upstream of *MMP16* gene (**Figure 5d**). The p-values in the PGC2014 study clearly suggested two signal peaks. One is located near the transcription start site of *MMP16*, while the other resides nearly 200,000 bases upstream and shows slightly stronger signal. However, the two-peak pattern was not clear in the PGC2011 study. Instead, two SNPs close to the end of the LD block near 89.8M showed the strongest signal. We compared the local predictions based on brain, heart, and blood-specific GSP scores at this locus (**Figure 5e**). Brain GSP scores successfully revealed the multi-peak nature at this locus, suggested the importance of the peak near 89.6Mb, and diminished the signal strength at the two SNPs near 89.8Mb, concurrent with the PGC2014 results. Although the method was applied on PGC2011 p-values, the results after prioritization matched the signal pattern in the PGC2014 study very well. Heart GSP scores also suggested the existence of the signal peak near 89.6M. However, the posterior scores have lower values, and the overall signal pattern does not match the PGC2014 study very well. The signal peak near 89.6M was completely lost in the blood-specific results. The two SNPs near 89.8M, however, had large GSP scores. The differences across tissue types are concurrent with GS scores at this locus (**Figure 5d**). Upstream of *MMP16*, near 89.6M, several functional segments can be seen in brain, only one remains in heart, and none exists in blood. Through comparing the tissue-specific prioritization results with the p-values in PGC2014 study, we see that brain-specific GSP scores had the strongest signal strength, which can be quantified using the local maximum GSP score (**Online Methods**). The highly matched signal pattern also suggested that brain might be the tissue type in which this locus plays a functional role.

Further insight on risk loci associated with coronary artery disease

Next, we applied our method to another GWAS to further illustrate the biological insight that GenoSkyline can provide for understanding complex diseases. The CARDIoGRAM consortium published a large-scale GWAS meta-analysis of coronary artery disease (CAD) comprising 22,333 cases and 64,762 controls[33], in which they replicated 10 out of 12 previously reported risk loci and identified 13 new loci associated with CAD. We applied our method on the summary statistics and used the local maximum GSP score to measure the relatedness between each risk locus and different tissue types (**Online Methods**). We removed the locus on chromosome 1q41 (*MIA3*) and the locus on chromosome 6q25.3 (*LPA*) due to incomplete data in the meta-analysis stage of CARDIoGRAM study. The remaining 21 CAD-associated loci are summarized in **Table 1** and **Supplementary Table 9**.

The first impression of these results is that despite the strong overall enrichment of GWAS signals (**Figure 4**), heart is the most relevant tissue type for only two loci. On the contrary, a substantial proportion of risk loci (9 out of 21) seem to be functional in the GI tissue. Interestingly, GI was the most enriched tissue type for several known risk factors for CAD including LDL and total cholesterol (**Figure 4**). These results suggest not only

the larger effect sizes of CAD-associated loci in the gastrointestinal system, but also a substantial amount of undetected heart-related loci. Furthermore, brain was the least enriched tissue type for CAD GWAS signals, but the risk locus on chromosome 14q32.2 near *HHIPL1* and *CYP46A1* was predicted to be functional in brain. In fact, the *CYP46A1* gene encodes for Cholesterol 24-hydroxylase that is present mainly in brain, where it converts cholesterol from degraded neurons into 24S-hydroxychoelesterol[33,34]. This process is crucial for eliminating cholesterol from the brain since cholesterol is usually unable to pass the blood-brain barrier[35].

A larger GWAS for CAD was published during the preparation of this manuscript[36]. This large study may be used to validate the performance of our approach when its summary statistics become publicly available in the future.

Discussion

In this paper, we introduced GenoSkyline, an integrative framework for predicting tissue-specific functional regions in the human genome. Through integrating GenoSkyline annotations with GWAS summary statistics, we illustrated a variety of ways that GenoSkyline could help researchers understand human complex diseases and traits. We also showed that the GenoSkyline framework is customizable so that researchers can develop their own functional annotations for a selected group of cells. As epigenomic ChIP-seq data become available for an increasing number of cell types in the future, GenoSkyline's ability to facilitate studies of complex disease will be further enhanced.

Our approach is not without limitation. First, the annotation results are incomplete due to currently unavailable tissue types, and as a result, the GWAS enrichment results may not be comprehensive (e.g. liver may also be highly related to CAD, but there is no complete annotation data from liver yet). Second, some risk loci (or independent functional segments at the same locus) may play active roles in multiple tissue types. For example, in our PGC GWAS analysis, although local maximum GSP scores suggest that brain may be more relevant with the risk locus upstream of *MMP16*, two SNPs near 89.8MB are located near several functional segments in blood. Whether these SNPs can be functionally linked to schizophrenia remains to be investigated. Moreover, we emphasize that our method identifies regions of likely functionality, but does not provide conclusive proof of functionality for any individual SNP or locus. That said, our method still provides a simple and intuitive summary statistic that measures the relatedness between risk loci and sets of functionally related tissues. It has great potential to become a standard step in downstream GWAS analysis to help researchers generate new hypotheses regarding the etiology behind each risk locus.

The increasing accessibility of GWAS summary statistic datasets, coupled with the method's independence from requiring individual-level genotype and phenotype, make Genoskyline tissue-specific prioritization useful and easy to implement. Moreover, GWAS signal integration is just one way to utilize GenoSkyline annotations. Its nucleotide-level functional prediction based on unsupervised learning and the good predictive performance in non-coding regions promise a potential role in many fields of genomics, such as next-generation sequencing studies and understanding somatic mutations. GenoSkyline scores of seven tissue types and two additional cell types have been pre-calculated for the entire human genome and can be readily downloaded. Source code is available for all major OSes and can be accessed at (http://genocanyon.med.yale.edu/GenoSkyline). We believe that GenoSkyline and its applications can guide genetics research at multiple resolutions and greatly benefit the broader scientific community.

Online Methods

Consolidated epigenomes

Epigenetic data were selected from the Epigenomics Roadmap Project's 111 consolidated reference epigenomes database[3] (http://egg2.wustl.edu/roadmap/) based on anatomy type and mark availability. Each tissue type is a clustering of relevant samples in order to contain at least one of each of the following: H3k4me1, H3k4me3, H3k36me3, H3k27me3, H3k9me3, H3k27ac, H3k9ac, and DNase I Hypersensitivity. Samples are reduced to a per-nucleotide binary encoding of presence or absence of narrow contiguous regions of ChIP-seq signal enrichment compared to input (Poisson p-value threshold of 0.01), and a union of all tissue-specific samples for that mark is taken. The set of 8 marks was chosen due to the well-understood, localized regulatory interactions of histone marks[37] and DNase I[38]. We created nine unique tissue and cell type clusters based on these annotations (**Supplementary Table 1**) to represent common, physiologically-related organ systems. To reflect actual tissue-specific epigenetic behavior, a majority of samples chosen are primary tissues and cultures, and inclusion of immortalized cell lines has been kept to a minimum.

GenoSkyline model and estimation

Lu et al. previously proposed a method that applies unsupervised-learning techniques on genomic annotations to predict the functional potential of a genomic region[4]. Given a set of annotations $\boldsymbol{A}$, we assume the joint distribution of $\boldsymbol{A}$ along the genome to be a mixture of annotations at locations with no functionality, i.e. $f(\boldsymbol{A} \mid Z = 0)$, and annotations at locations that are functional, i.e. $f(\boldsymbol{A} \mid Z = 1)$. We assume that each annotation in $\boldsymbol{A}$ is conditionally independent given Z, allowing the conditional joint density of $\boldsymbol{A}$ given Z to be factorized as

$$f(\boldsymbol{A}|Z = c) = \prod_{i=1}^{8} f_i(A_i|Z = c), \qquad c = 0, 1 \tag{1}$$

Since all annotations used are binary classifiers, the Bernoulli distribution was used to model the marginal functional likelihood given each individual annotation.

$$f_i(A_i|Z = c) = p_{ic}^{A_i}(1 - p_{ic})^{1-A_i}, \qquad i = 1, \ldots, 8;\ c = 0, 1 \tag{2}$$

Assuming a prior probability π of being functional ($\pi = P(Z = 1)$), we can estimate the parameter p_{ic} of each annotation with the Expectation-Maximization (EM) algorithm, and calculate the posterior probability at a given genomic coordinate, referred to as the GS score.

$$P(Z=1|\boldsymbol{A}) = \frac{\pi f(\boldsymbol{A}|Z=1)}{\pi f(\boldsymbol{A}|Z=1) + (1-\pi) f(\boldsymbol{A}|Z=0)} = \frac{\pi \prod_{i=1}^{8} f_i(A_i|Z=1)}{\pi \prod_{i=1}^{8} f_i(A_i|Z=1) + (1-\pi) \prod_{i=1}^{8} f_i(A_i|Z=0)} \quad (3)$$

We must estimate 17 parameters for each tissue tract.

$$\boldsymbol{\Theta} = (\pi, \boldsymbol{P_1}, \boldsymbol{P_0}) \quad (4)$$

Where

$$\boldsymbol{P_c} = (p_{1c}, p_{2c}, \dots, p_{8c}), \quad c = 0,1 \quad (5)$$

Parameters were estimated using the GWAS Catalog[6], downloaded from the NHGRI website (http://www.genome.gov/gwastudies/), which at the time of download, contained 13,070 unique SNPs found to be significant in at least one published GWAS. These SNPs were expanded into 1k bp intervals, and formed a genome sampling covering 12,801,840 bp of the genome. While significant SNP associations are likely to tag the effects of nearby functional elements, the size and distance of these functional elements varies for each individual SNP. As a result, the total sampling serves as an effective and robust representation of functional and non-functional regions along the genome (**Supplemental Notes**).

Case studies of experimentally validated functional machinery

VISTA enhancers[14] were downloaded from the VISTA Enhancer Browser (http://enhancer.lbl.gov/), where enhancers with E11.5 reporter staining experimental data were selected. Brain enhancers were selected based on staining results identifying any CNS-related tissues (neural tube, cranial nerve, hindbrain, mesenchyme derived from neural crest, trigeminal V, forebrain, and midbrain). Heart enhancers were enhancers identified for positive reporter results in the heart region of E11.5 mouse reporter assays. Blood vessels enhancers were identified by selecting for "blood vessels" expression pattern. Hypomethylated TE loci in H1ES, brain, breast, and blood were downloaded from http://epigenome.wustl.edu/TE_Methylation/. All genomic coordinates were converted to genome build hg19.

SNP prioritization using tissue-specific functional annotation

We identify three disjoint cases for a given GWAS SNP:

I. The SNP is in a genomic region that is functional for the given phenotype and tissue ($Z_D = 1$, $Z_T = 1$).

II. The SNP is in a genomic region that is functional for the given tissue, but that tissue has no functionality in the phenotype ($Z_D = 0$, $Z_T = 1$).

III. The SNP is in a genomic region that is not functional in the given tissue ($Z_T = 0$).

A useful metric for prioritizing SNPs is the conditional probability that the SNP is classified under case-I given its p-value in a given GWAS study, i.e. $P(Z_D = 1, Z_T = 1 \mid p)$. We can calculate this probability by employing Bayes formula and considering all three cases as follows:

$$P(Z_D = 1, Z_T = 1 \mid p) = \frac{f(p|Z_D = 1, Z_T = 1) \times P(Z_D = 1, Z_T = 1)}{f(p|Z_D = 1, Z_T = 1) \times P(Z_D = 1, Z_T = 1) + f(p|Z_D = 0, Z_T = 1) \times P(Z_D = 0, Z_T = 1) + f(p|Z_T = 0) \times P(Z_T = 0)} \quad (6)$$

First, the case in which $Z_T = 0$ can be directly identified by assigning each SNP a prior probability of tissue-specific functionality (i.e. $P(Z_T = 1)$) defined as the average GS score of its surrounding 10,000 base pairs for that tissue (**Supplementary Notes**). We partition all the SNPs into two subgroups based on a mean GS score threshold of 0.1, although these probabilities take on a bimodal distribution and are not sensitive to changing threshold[10]. In this way, we can use these partitions to directly estimate $f(p|Z_T = 0)$ by applying density estimation techniques on the SNP subgroup with low GS scores. More specifically, we apply histogram for density estimation and use cross validation to choose the optimal number of bins.

Second, we estimate the p-value density of our second case, where $Z_D = 0$ and $Z_T = 1$. We can intuitively assume that SNPs that are functional in a tissue but not relevant to the phenotype will have similar p-value behavior to all other SNPs that are not relevant to the phenotype, which in turn behave similarly to SNPs that are not functional at all (**Supplementary Notes**). More formally, we can describe this relationship as follows:

$$f(p|Z_D = 0, Z_T = 1) = f(p|Z_D = 0) = f(p|Z = 0) \quad (7)$$

We can effectively estimate $f(p|Z = 0)$ by using a similar approach to estimating $f(p|Z_T = 0)$, but partitioning SNPs using the general functionality GenoCanyon score instead of tissue-specific GS score.

Next, we consider the following formulas.

$$P(Z_D = 1, Z_T = 1) = P(Z_D = 1|Z_T = 1) \times P(Z_T = 1) \quad (8)$$

$$P(Z_D = 0, Z_T = 1) = P(Z_D = 0|Z_T = 1) \times P(Z_T = 1) \quad (9)$$

The prior probability $P(Z_T = 1)$ can be calculated directly from GS scores as stated above, but the conditional probabilities of disease-specific functionality given tissue-specific functionality remains to be estimated.

Finally, we estimate all the remaining terms in formula 4 using the EM algorithm. In the first step of the estimation procedure, we acquired the subset of SNPs located in tissue-specific functional regions. The p-value distribution of these SNPs is the following mixture.

$$f(p|Z_T = 1) = P(Z_D = 1|Z_T = 1)\times f(p|Z_D = 1, Z_T = 1) + P(Z_D = 0|Z_T = 1)\times f(p|Z_D = 0, Z_T = 1) \quad (10)$$

Density $f(p|Z_D = 0, Z_T = 1)$ has been estimated in earlier steps. Applying the findings of Chung et al., we assume a beta distribution of the p-values of functional SNPs (i.e. $f(p|Z_D = 1, Z_T = 1)$) as a reasonable approximation under general assumptions of SNP effect size[9].

$$(p|Z_D = 1, Z_T = 1) \sim Beta(\alpha, 1), \qquad 0 < \alpha < 1 \quad (11)$$

The EM algorithm is then applied to the SNP subset located in tissue-specific functional regions. The beta assumption guarantees a closed-form expression in each iteration and all the remaining parameters can be subsequently estimated. We now have all the necessary terms for equation 4, and define this as our posterior probability score of tissue-specific disease functionality (GSP score). The feature of integrating tissue-specific functional annotations to prioritize GWAS signals has been added to the GenoWAP software available on our server (http://genocanyon.med.yale.edu/GenoSkyline).

SNP prioritization using GenoCanyon annotation

Non-tissue specific GenoCanyon scores are assigned to GWAS signals using GenoWAP[10]. Briefly, GenoWAP calculates the posterior score $P(Z_D = 1|p)$ using a simpler model for functionality.

$$P(Z_D = 1|p) = \frac{f(p|Z_D = 1)\times P(Z_D = 1)}{f(p|Z_D = 1)\times P(Z_D = 1) + f(p|Z_D = 0)\times P(Z_D = 0)} \quad (12)$$

This conditional probability can be calculated similarly to GS scores, making use of equation (5) to empirically estimate $f(p|Z_D = 0)$, a beta distribution on partitioned Genocanyon scores (calculated with 22 tissue non-specific ENCODE functionality annotations[4]) to estimate $f(p|Z_D = 1)$, and the EM algorithm on the functional marker p-value density to calculate $P(Z_D = 1)$ as described in Lu *et al*. These are referred to in the results as the NSFP scores to which GenoSkyline SNP prioritization is compared.

Calculating tissue-specific enrichment using LD score regression

Enrichment of GenoSkyline-derived tissue-specific annotations in GWAS summary statistics was calculated using stratified LD score regression[27]. First, tissue-specific

annotations were computed using GenoSkyline scores, 1000 Genomes data of European ancestry[39] and a 1-centiMorgan (cM) window. Then the annotations were analyzed by adding each one of them to the full baseline model to control for 53 categories of general annotations. For each tissue-specific annotation, partitioned heritability was estimated using stratified LD score regression[27] and enrichment was then calculated as the ratio of proportion of SNP heritability explained by the annotation and proportion of SNPs in the annotation.

Measuring relevant tissue types for GWAS risk loci

A large GSP score is obtained if the p-value for the SNP is small and the SNP is located in a highly functional region for the tissue type under investigation. Therefore, the maximal GSP score at a risk locus effectively measures how well the p-values match the pattern of GenoSkyline annotations, thereby measuring the relatedness between the GWAS locus and different tissue types. For each tissue, the maximal GSP score is acquired at the risk locus of interest. These scores are then compared across tissue types. The largest score is referred to as local maximum GSP score, and the corresponding tissue type is predicted to be the most relevant tissue.

Bioinformatics tools

Locus plots were generated using LocusZoom[40]. The "ggbio" R package[41] was used to plot genes. The "bigmemory" R package[42] was used to access and manipulate massive dataset.

Acknowledgements

This study was supported in part by the National Institutes of Health grants R01 GM59507, the VA Cooperative Studies Program of the Department of Veterans Affairs, Office of Research and Development, and the Yale World Scholars Program sponsored by the China Scholarship Council.

Author Contributions

Q.L. and R.L.P. conceived the project, wrote the initial draft, and performed the analyses. Q.W. performed tissue-specific enrichment analysis. B.J.H. performed one heart-related case study. H.Z. advised on statistical and genetic issues.

Figures and Tables

Figure 1. General characteristics of GenoSkyline annotations. (a) Number of tissues in which nucleotides are functional. (b) Proportion of functional genome for each tissue type. (c) Overlap of functional regions across seven tissue types. The scale is log odds ratio.

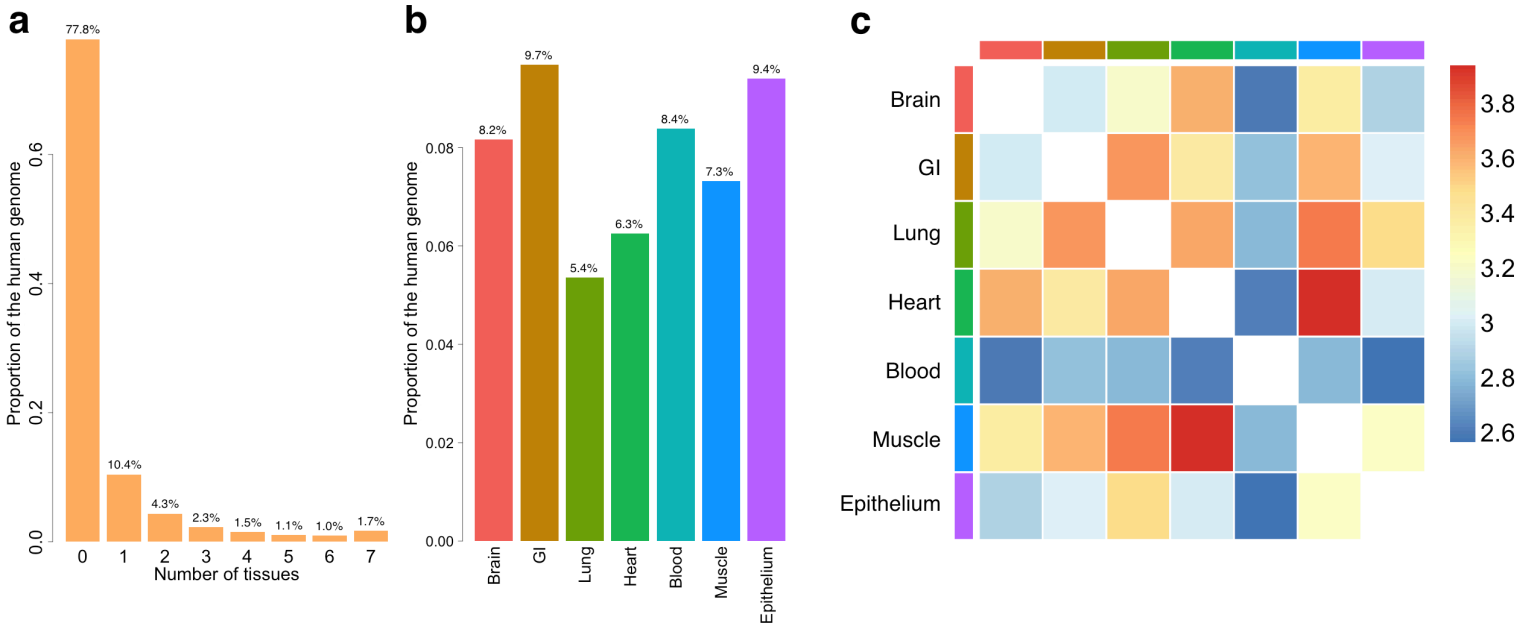


Figure 2. Case studies of *HBB* gene complex, *in vivo* enhancers, and regulatory miRNAs. (a) Comparison of GenoCanyon prediction and GenoSkyline scores for seven tissues in *HBB* gene complex region. Red boxes mark the locations of CRMs. The number of red boxes is less than 23 because some CRMs are next to each other. (b) Mean blood-specific GS score for different region categories. (c) Boxplot of mean GS scores for enhancers in CNS, heart, and blood vessel. (d) Boxplot of mean GS scores for 11 human-accelerated elements near *NPAS3*. (e) Boxplot of mean GS scores for tissue-specific regulatory miRNAs.

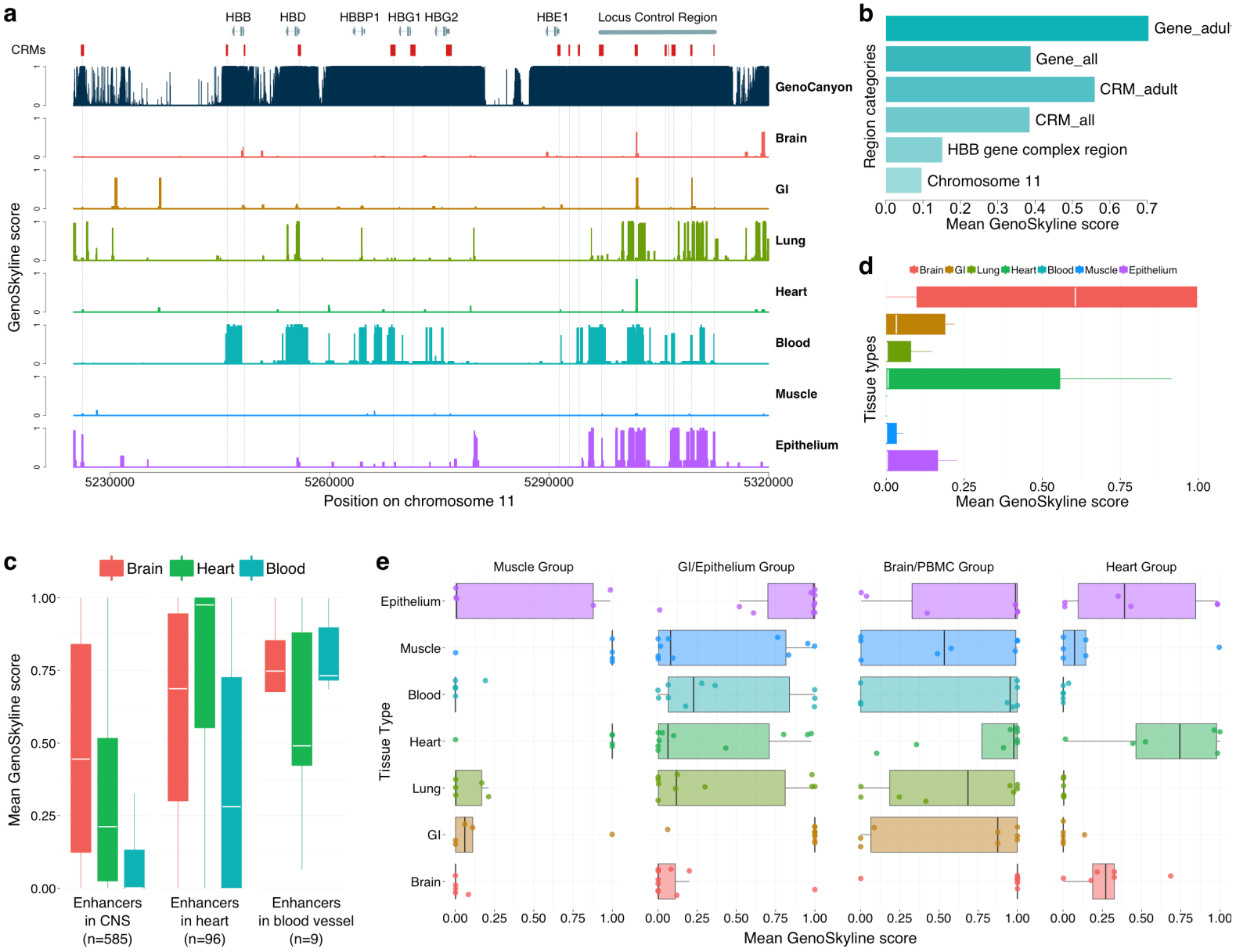


Figure 3. Case studies of MHC, ZRS, and hypomethylated TEs. (a) GenoSkyline scores for seven tissues in the genomic region surrounding *MYH6* and *MYH7*. (b) ESC-specific and fetal-cell-specific GS scores for the 5th intron of *LMBR1*. The red box marks the location of ZRS. (c) Bar plot of the mean GS scores for the 5th intron of *LMBR1* and ZRS across nine tissue and cell types. (d) Boxplot of mean GS scores for four groups of hypomethylated TEs.

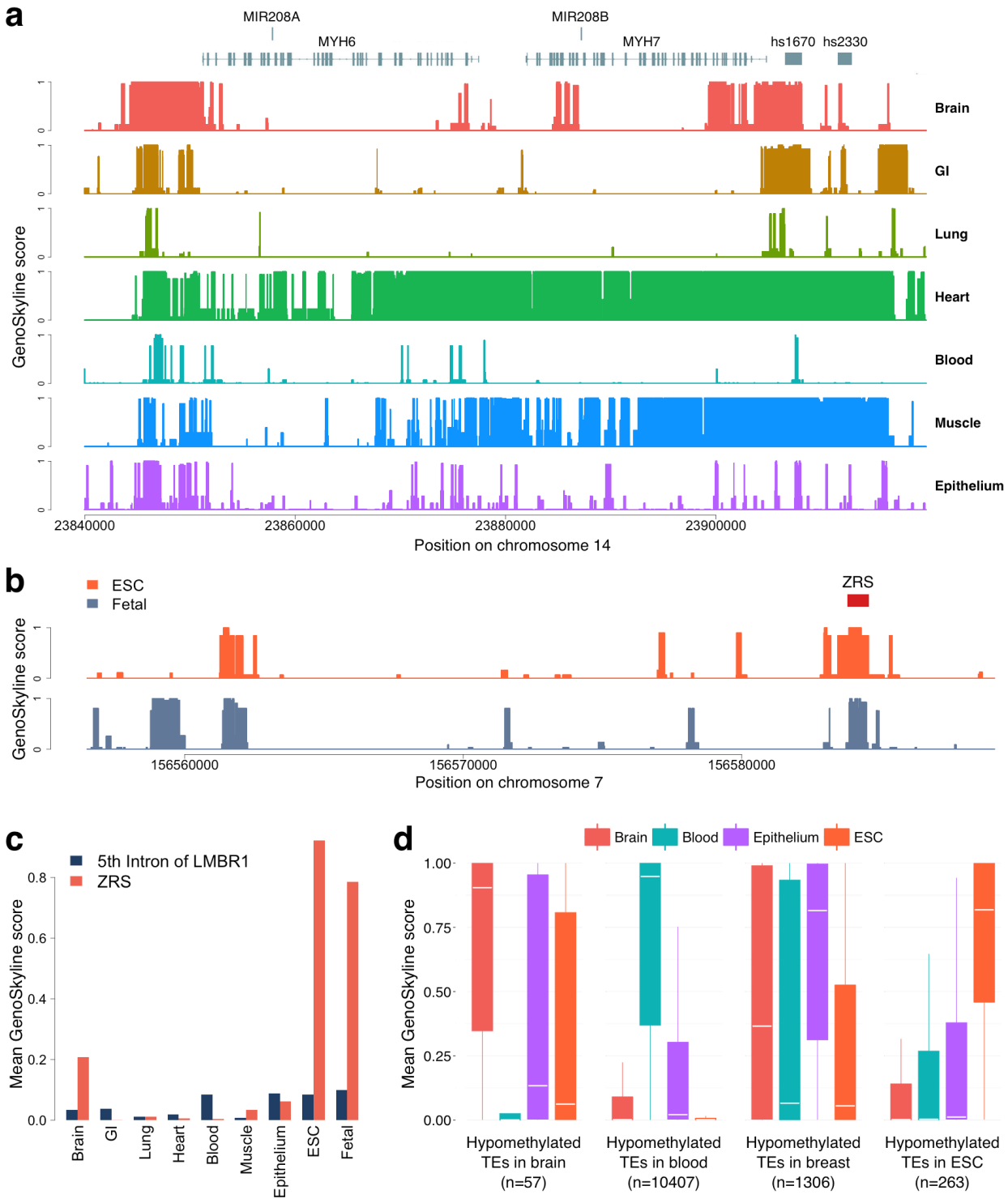


Figure 4. Tissue-specific enrichment of GWAS signals. Enrichment p-values were calculated using LD score regression. The grey line is the 0.05 cutoff for p-value.

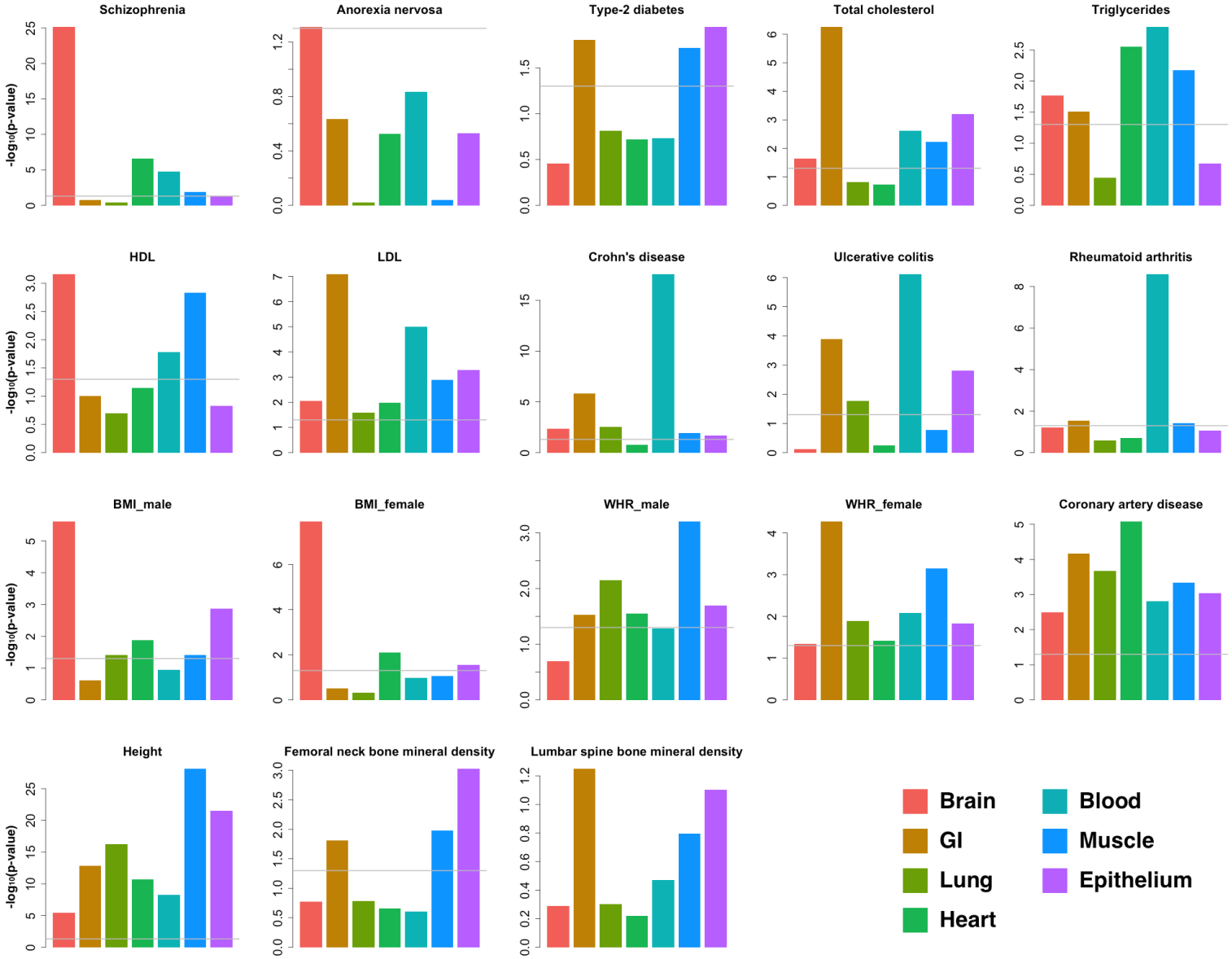


Figure 5. Prioritizing schizophrenia GWAS signals using GenoSkyline annotations. (a) Tissue-specific functional regions are more enriched of schizophrenia associations than generally functional regions and non-functional regions. (b) Enrichment of GTEx whole-blood eQTLs in top SNPs from PGC2011 study. (c) Enrichment of human brain quantitative trait loci in top SNPs from PGC2011 study. (d) Summary statistics at the schizophrenia-associated locus on chromosome 8q21 near *MMP16* gene. The top and middle panel show p-values from PGC2011 and PGC2-14 studies, respectively. The bottom panel shows GenoSkyline annotations at this locus. (e) Locus plots for tissue-specific posterior scores. From top to bottom, the three panels show posterior scores of brain, heart, and blood tissues, respectively.

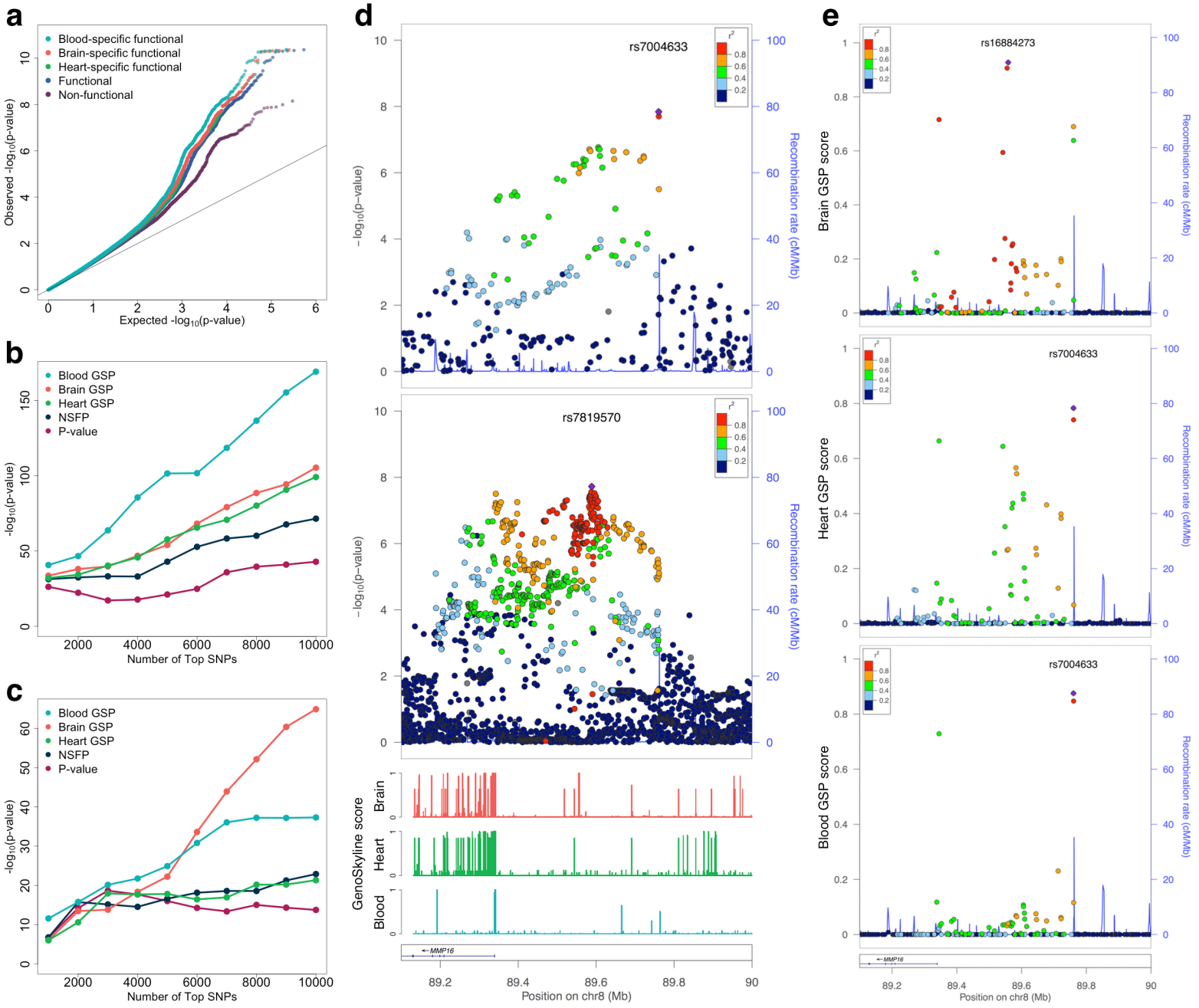


Table 1. Risk loci for coronary artery disease

Chr	Start[a]	Stop[a]	Genes in region	Tissue type[b]	Posterior score[c]
1p13.3	109,700,000	109,900,000	*SORT1*	GI	0.99991
2q33.1	203,600,000	204,000,000	*WDR12*	GI	0.99999
3q22.3	137,900,000	138,200,000	*MRAS*	Heart	0.99625
6p24.1	12,700,000	13,100,000	*PHACTR1*	GI	0.99997
9p21.3	21,900,000	22,200,000	*CDKN2A, CDKN2B*	NA[d]	NA[d]
10q11.21	44,400,000	44,900,000	*CXCL12*	GI	0.83314
19p13.2	11,000,000	11,400,000	*LDLR*	Blood	0.99967
21q22.11	35,500,000	35,700,000	*MRPS6*	Lung	0.9993
1p32.2	56,900,000	57,100,000	*PPAP2B*	GI	0.99699
6p21.31	34,600,000	35,300,000	*ANKS1A*	Heart	0.95156
6q23.2	134,000,000	134,300,000	*TCF21*	GI	0.99998
7q32.2	129,600,000	129,900,000	*ZC3HC1*	Muscle	0.9995
9q34.2	136,000,000	136,400,000	*ABO*	GI	0.99531
10q24.32	104,400,000	105,000,000	*CYP17A1, CNNM2, NT5C2*	GI	0.94966
11q23.3	116,500,000	116,700,000	*ZNF259, APOA5-A4-C3-A1*	Blood	0.9997
13q34	110,700,000	111,200,000	*COL4A1, COL4A2*	Muscle	0.99704
14q32.2	100,000,000	100,300,000	*HHIPL1, CYP46A1*	Brain	0.97293
15q25.1	78,900,000	79,200,000	*ADAMTS7*	Muscle	0.99934
17p13.3	2,000,000	2,300,000	*SMG6, SRR*	GI	0.98812
17p11.2	17,400,000	18,000,000	*RASD1, SMCR3, PEMT*	Blood	0.96864
17q21.32	46,800,000	47,200,000	*UBE2Z, GIP, ATP5G1, SNF8*	Blood	0.96206

[a]These coordinates are the roughly estimated boundaries of risk loci. The inference of relevant tissues is not sensitive to the boundary coordinates.
[b]The tissue type that provides the largest local maximum posterior score.
[c]Local maximum GSP score for the most relevant tissue type.
[d]Not applicable due to ties. See Supplementary Table 9.

Supplementary Notes

Bimodal pattern in GenoSkyline score

When estimating the proportion of functional genome for each tissue type, we adopted 0.5 as the cutoff for GS score. The GS score histograms for different tissues on chromosome 22 are plotted in **Supplementary Figure 6**. The GS score distributions have a clear and consistent bimodal pattern across different tissue types. The similar bimodal pattern can also be observed for other chromosomes. Therefore, the cutoff choice does not substantially affect the estimation of functional proportion.

Robustness of GenoSkyline parameter estimation

The parameters in the GenoSkyline framework were estimated from a set of 12,801,840 bases acquired from GWAS catalog. The reason of using this GWAS-based set is to guarantee the inclusion of a sufficient amount of functional bases. In our previous work, we showed that parameter estimation under this framework is robust[4]. Here, all the 17 parameters were re-estimated after adding 2,000,000 and 6,000,000 bases randomly selected from chromosome 1 to the initial set containing 12,801,840 bases. The parameter estimates remained highly stable (**Supplementary Table 10**). These results show that GenoSkyline parameter estimation is insensitive to the choice of the initial set.

Several remarks on GSP score

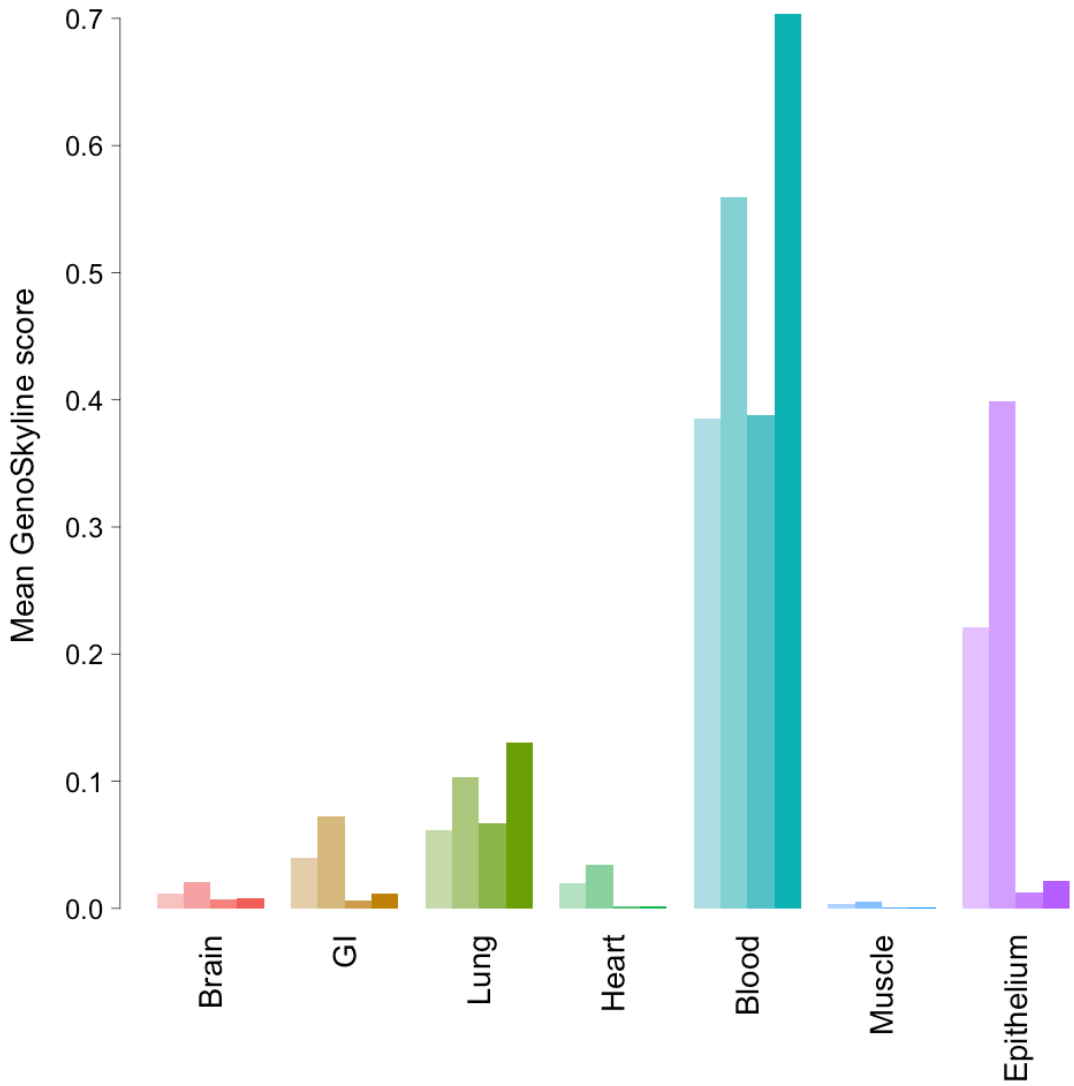
For GSP score calculation, we used the mean GS score of the surrounding 10,000 bases as the prior probability $P(Z_T = 1)$ for each SNP. This is because the nucleotide-level GS score may be insufficient for GWAS signal prioritization. In fact, each SNP in GWAS carries information of its nearby variants that are not genotyped or imputed. A better-informed metric needs to measure the functional potential for the surrounding region of each SNP. We chose 10,000 bases as the window size, but no substantial difference in empirical performance was observed when changing the window size to 5,000 or 20,000. In our implemented GenoWAP software for SNP prioritization (available at http://genocanyon.med.yale.edu/GenoSkyline), the users are allowed to use their own annotation data. Therefore, the window size can be changed when necessary. Since the mean GS score of surrounding regions was used as the prior, our SNP prioritization approach is in fact a region-based tool. We emphasize again that it identifies regions of likely functionality and substantially improves the resolution of GWAS, but does not provide conclusive proof of functionality for any individual SNP or locus.

In order to calculate the GSP score, we assumed that SNPs that are functional in a tissue not relevant to the phenotype would have similar p-value behavior to all other SNPs that are not relevant to the phenotype, which in turn behave similarly to SNPs that are not functional at all (see equation 7 in **Online Methods**). This assumption may not hold exactly due to some intrinsic differences between SNPs located in non-functional regions

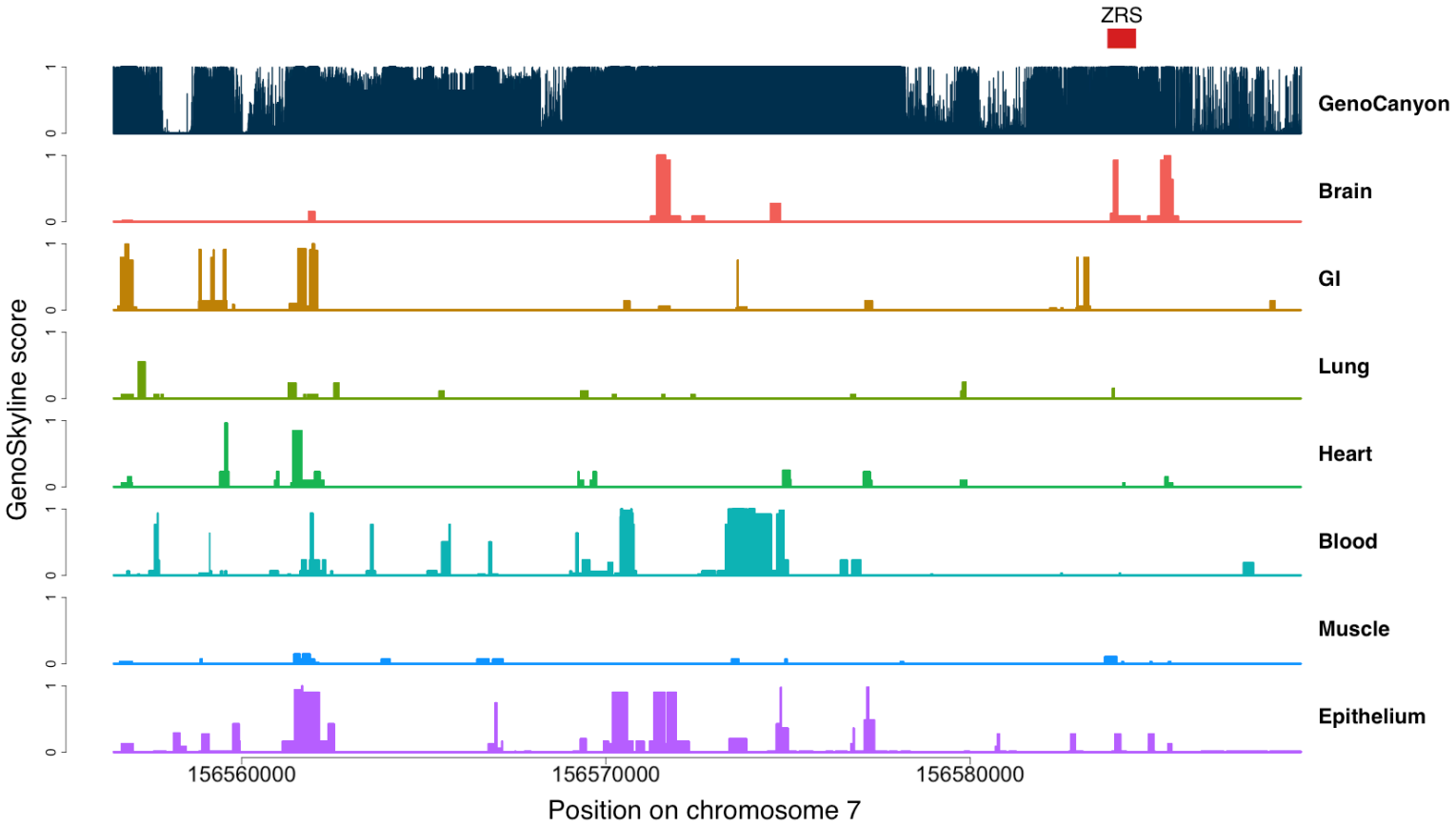
(Z=0) and those in non-specific functional (Z=1) and tissue-specific functional regions (Z_T=1). As far as we are aware, the main possible contributing factor may be different linkage disequilibrium (LD) patterns in those regions with Z = 0 and those with Z = 1 or Z_T=1. For example, if there is stronger LD in Z_T = 1 regions, then the markers with Z_D = 0 and Z_T = 1 may have a different p-value distribution from those with Z = 0.

In order to check if this is a serious issue, we compared the LD patterns in regions with Z = 0, Z =1, and Z_T = 1 for multiple tissue types on chromosome 22. We downloaded the pre-calculated LD scores[43] for the 1000 Genomes European population from the LD score GitHub page (https://github.com/bulik/ldsc/wiki/LD-Score-Estimation-Tutorial). Based on cutoff 0.1 for GenoCanyon and GenoSkyline scores, we divided all the SNPs on chromosome 22 into tissue-specific functional, non-specific functional and non-functional subcategories. The kernel density estimates of the two subgroups are plotted in **Supplementary Figure 7**. It can be seen that there is no substantial difference of LD score distributions between different categories. Therefore, it is reasonable to assume the LD patterns in regions with Z = 0, Z = 1, and Z_T = 1 to be similar. Moreover, as Z_D = 1 is a relatively small proportion of Z = 1, this assumption is likely to be a good approximation.

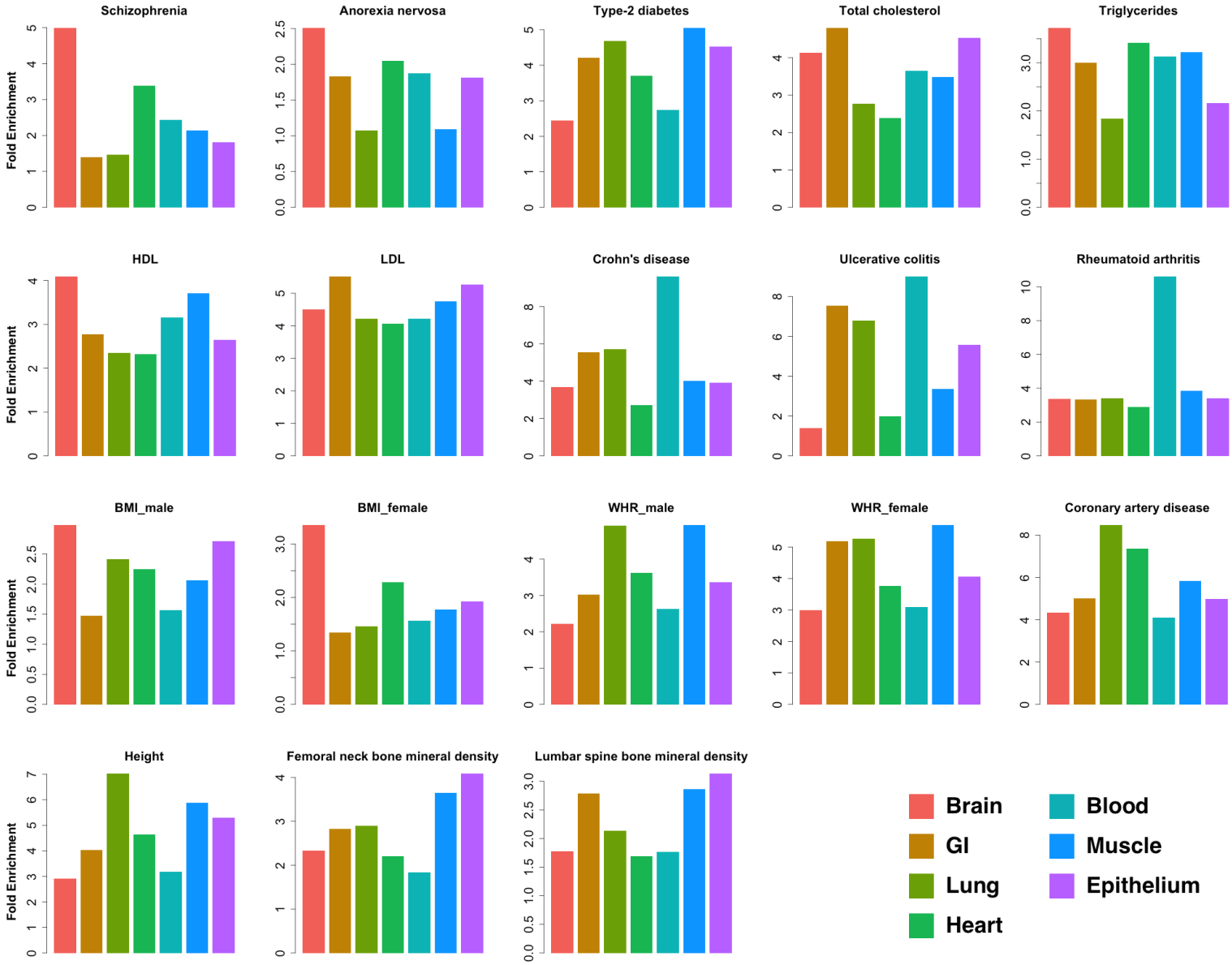
Supplementary Figure 1. Mean GS score for different region categories in *HBB* gene complex across seven tissue types. For each tissue, the four bars from left to right indicate all 23 CRMs, adult CRMs, all genes, and adult globins, respectively.



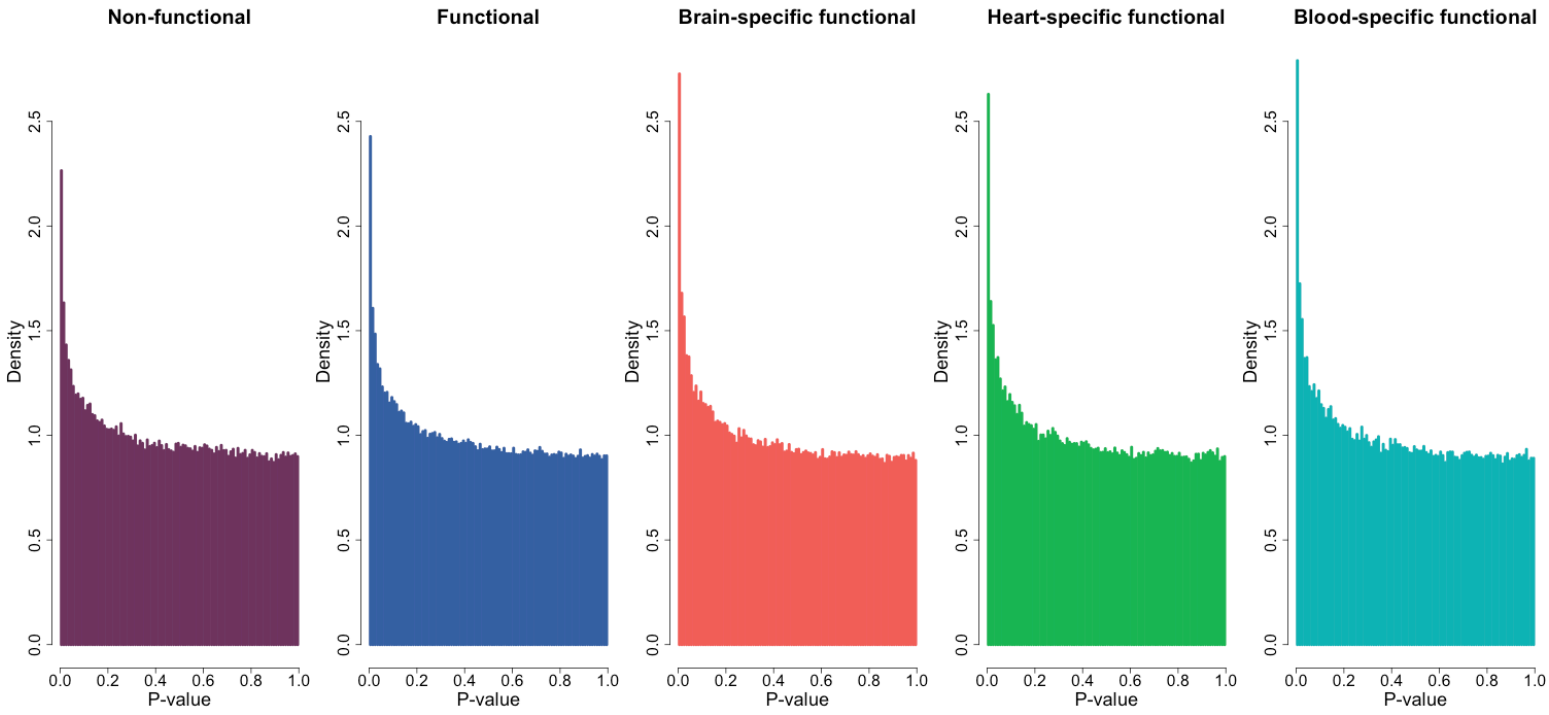
Supplementary Figure 2. GenoCanyon score and GenoSkyline scores for seven tissues in the 5th intron of *LMBR1*. The red box marks the location of ZRS.



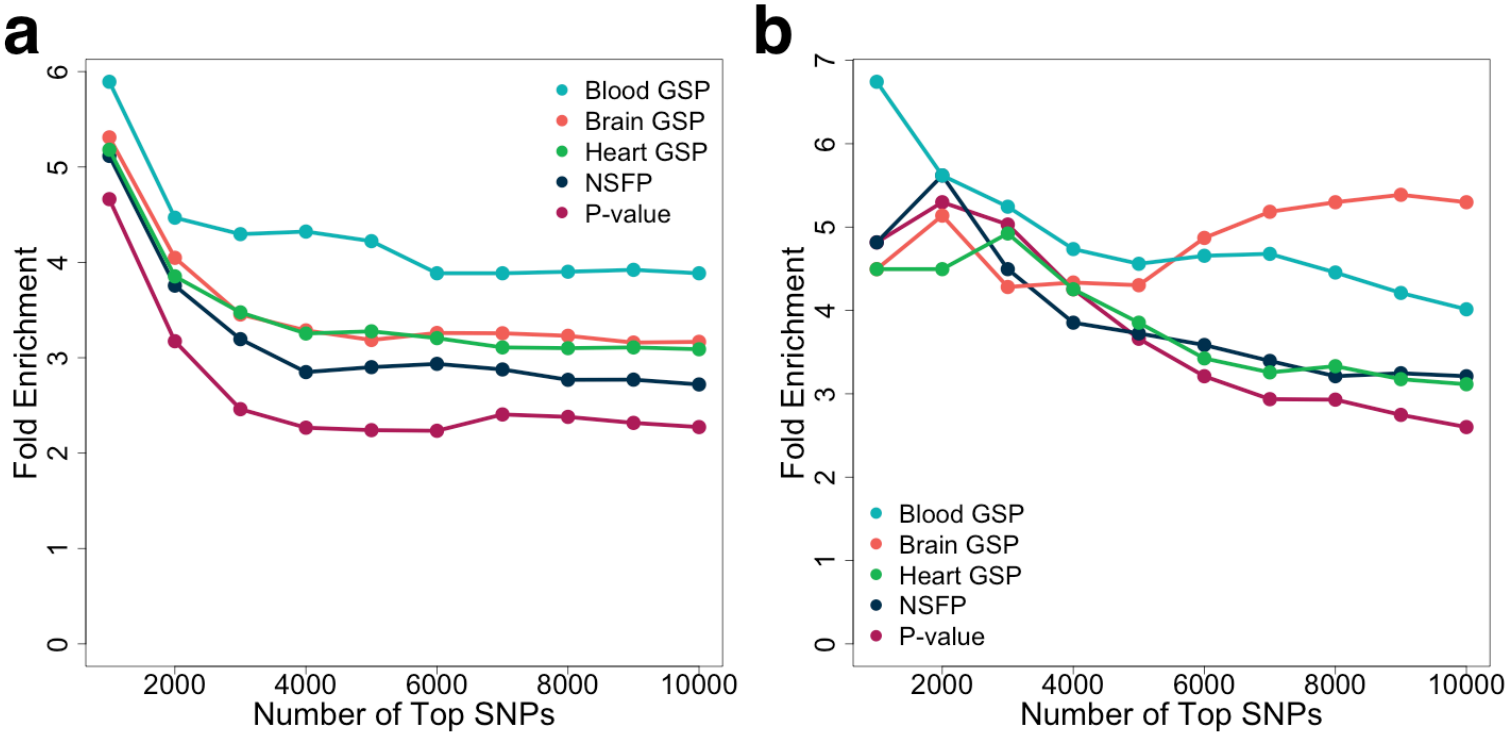
Supplementary Figure 3. Tissue-specific fold enrichment of GWAS signals.



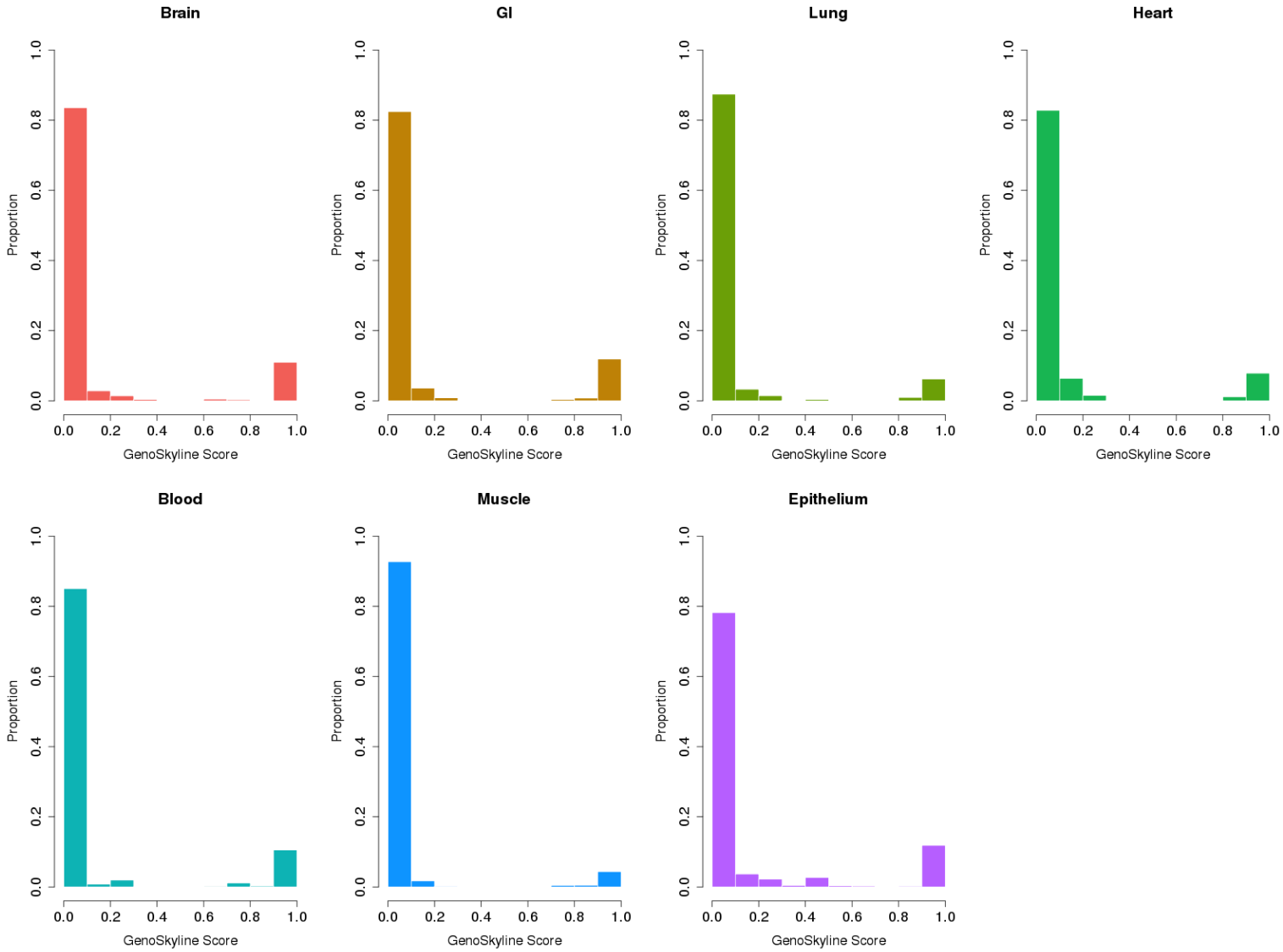
Supplementary Figure 4. Histograms of p-values for SNPs located in non-functional, functional, and tissue-specific functional regions.



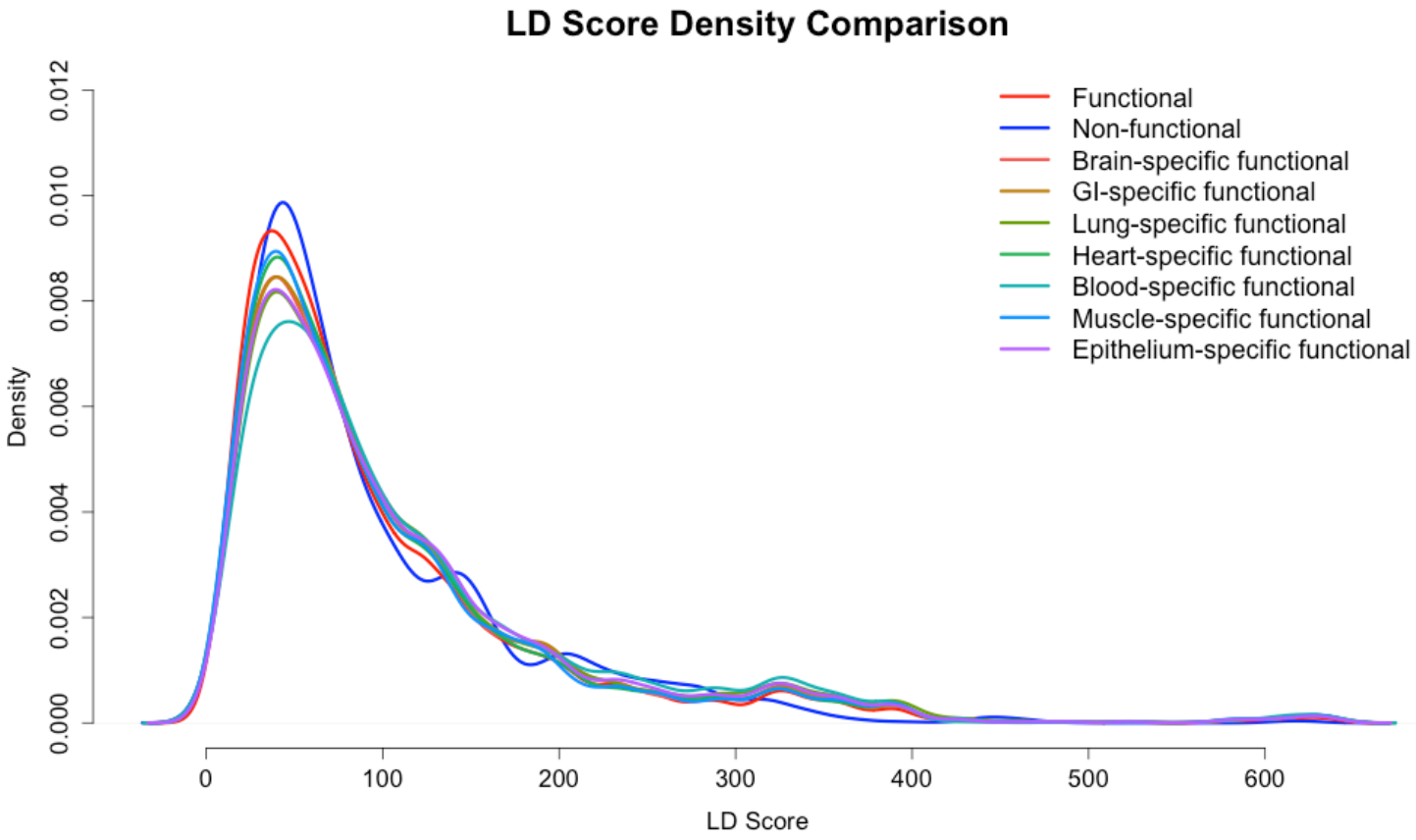
Supplementary Figure 5. Fold enrichment of eQTLs in top SNPs from PGC2011 study. (a) GTEx whole-blood eQTLs. (b) Human brain quantitative trait loci.



Supplementary Figure 6. Distribution of GenoSkyline scores on chromosome 22.



Supplementary Figure 7. Comparison of LD score densities on chromosome 22 across different SNP categories.



Supplementary Table 1. Cell types used for developing GenoSkyline annotations of seven tissue types.

Tissue	EID[a]	Anatomy	Standardized Epigenome Name	Type
Brain	E054	BRAIN	Ganglion Eminence derived primary cultured neurospheres	PrimaryCulture
	E053	BRAIN	Cortex derived primary cultured neurospheres	PrimaryCulture
	E071	BRAIN	Brain Hippocampus Middle	PrimaryTissue
	E074	BRAIN	Brain Substantia Nigra	PrimaryTissue
	E068	BRAIN	Brain Anterior Caudate	PrimaryTissue
	E069	BRAIN	Brain Cingulate Gyrus	PrimaryTissue
	E072	BRAIN	Brain Inferior Temporal Lobe	PrimaryTissue
	E067	BRAIN	Brain Angular Gyrus	PrimaryTissue
	E073	BRAIN	Brain Dorsolateral Prefrontal Cortex	PrimaryTissue
	E070	BRAIN	Brain Germinal Matrix	PrimaryTissue
	E082	BRAIN	Fetal Brain Female	PrimaryTissue
	E081	BRAIN	Fetal Brain Male	PrimaryTissue
GI	E078	GI_DUODENUM	Duodenum Smooth Muscle	PrimaryTissue
	E076	GI_COLON	Colon Smooth Muscle	PrimaryTissue
	E103	GI_RECTUM	Rectal Smooth Muscle	PrimaryTissue
	E111	GI_STOMACH	Stomach Smooth Muscle	PrimaryTissue
	E092	GI_STOMACH	Fetal Stomach	PrimaryTissue
	E085	GI_INTESTINE	Fetal Intestine Small	PrimaryTissue
	E084	GI_INTESTINE	Fetal Intestine Large	PrimaryTissue
	E109	GI_INTESTINE	Small Intestine	PrimaryTissue
	E106	GI_COLON	Sigmoid Colon	PrimaryTissue
	E075	GI_COLON	Colonic Mucosa	PrimaryTissue
	E101	GI_RECTUM	Rectal Mucosa Donor 29	PrimaryTissue
	E102	GI_RECTUM	Rectal Mucosa Donor 31	PrimaryTissue
	E110	GI_STOMACH	Stomach Mucosa	PrimaryTissue
	E077	GI_DUODENUM	Duodenum Mucosa	PrimaryTissue
	E079	GI_ESOPHAGUS	Esophagus	PrimaryTissue
	E094	GI_STOMACH	Gastric	PrimaryTissue
Lung	E017	LUNG	IMR90 fetal lung fibroblasts Cell Line	CellLine
	E088	LUNG	Fetal Lung	PrimaryTissue
	E096	LUNG	Lung	PrimaryTissue
Heart	E083	HEART	Fetal Heart	PrimaryTissue
	E104	HEART	Right Atrium	PrimaryTissue
	E095	HEART	Left Ventricle	PrimaryTissue
	E105	HEART	Right Ventricle	PrimaryTissue
	E065	VASCULAR	Aorta	PrimaryTissue
Blood	E062	BLOOD	Primary mononuclear cells from peripheral blood	PrimaryCell
	E034	BLOOD	Primary T cells from peripheral blood	PrimaryCell
	E045	BLOOD	Primary T cells effector/memory enriched from peripheral blood	PrimaryCell
	E033	BLOOD	Primary T cells from cord blood	PrimaryCell
	E044	BLOOD	Primary T regulatory cells from peripheral blood	PrimaryCell
	E043	BLOOD	Primary T helper cells from peripheral blood	PrimaryCell
	E039	BLOOD	Primary T helper naive cells from peripheral blood	PrimaryCell
	E041	BLOOD	Primary T helper cells PMA-I stimulated	PrimaryCell
	E042	BLOOD	Primary T helper 17 cells PMA-I stimulated	PrimaryCell
	E040	BLOOD	Primary T helper memory cells from peripheral blood 1	PrimaryCell
	E037	BLOOD	Primary T helper memory cells from peripheral blood 2	PrimaryCell
	E048	BLOOD	Primary T CD8+ memory cells from peripheral blood	PrimaryCell
	E038	BLOOD	Primary T helper naive cells from peripheral blood	PrimaryCell

	E047	BLOOD	Primary T CD8+ naive cells from peripheral blood	PrimaryCell
	E029	BLOOD	Primary monocytes from peripheral blood	PrimaryCell
	E031	BLOOD	Primary B cells from cord blood	PrimaryCell
	E035	BLOOD	Primary hematopoietic stem cells	PrimaryCell
	E051	BLOOD	Primary hematopoietic stem cells G-CSF-mobilized Male	PrimaryCell
	E050	BLOOD	Primary hematopoietic stem cells G-CSF-mobilized Female	PrimaryCell
	E036	BLOOD	Primary hematopoietic stem cells short term culture	PrimaryCell
	E032	BLOOD	Primary B cells from peripheral blood	PrimaryCell
	E046	BLOOD	Primary Natural Killer cells from peripheral blood	PrimaryCell
	E030	BLOOD	Primary neutrophils from peripheral blood	PrimaryCell
Muscle	E100	MUSCLE	Psoas Muscle	PrimaryTissue
	E108	MUSCLE	Skeletal Muscle Female	PrimaryTissue
	E107	MUSCLE	Skeletal Muscle Male	PrimaryTissue
	E089	MUSCLE	Fetal Muscle Trunk	PrimaryTissue
	E090	MUSCLE_LEG	Fetal Muscle Leg	PrimaryTissue
	E052	MUSCLE	Muscle Satellite Cultured Cells	PrimaryCulture
Epithelium	E055	SKIN	Foreskin Fibroblast Primary Cells skin01	PrimaryCulture
	E056	SKIN	Foreskin Fibroblast Primary Cells skin02	PrimaryCulture
	E059	SKIN	Foreskin Melanocyte Primary Cells skin01	PrimaryCulture
	E061	SKIN	Foreskin Melanocyte Primary Cells skin03	PrimaryCulture
	E057	SKIN	Foreskin Keratinocyte Primary Cells skin02	PrimaryCulture
	E058	SKIN	Foreskin Keratinocyte Primary Cells skin03	PrimaryCulture
	E028	BREAST	Breast variant Human Mammary Epithelial Cells (vHMEC)	PrimaryCulture
	E027	BREAST	Breast Myoepithelial Primary Cells	PrimaryCell

[a] Epigenome ID in Roadmap project

Supplementary Table 2. Proportion of functional genome across seven tissue types under GS score cutoff 0.5.

	Brain	GI	Lung	Heart	Blood	Muscle	Epithelium
Chr 1	0.097	0.116	0.059	0.073	0.099	0.084	0.109
Chr 2	0.078	0.099	0.054	0.060	0.083	0.072	0.092
Chr 3	0.078	0.096	0.050	0.063	0.080	0.072	0.090
Chr 4	0.054	0.068	0.034	0.042	0.051	0.047	0.065
Chr 5	0.073	0.082	0.047	0.057	0.069	0.063	0.082
Chr 6	0.079	0.099	0.053	0.062	0.090	0.075	0.095
Chr 7	0.079	0.090	0.051	0.061	0.080	0.070	0.090
Chr 8	0.074	0.086	0.051	0.062	0.073	0.070	0.088
Chr 9	0.085	0.099	0.056	0.061	0.075	0.074	0.094
Chr 10	0.098	0.116	0.064	0.079	0.093	0.091	0.103
Chr 11	0.111	0.121	0.069	0.079	0.098	0.090	0.120
Chr 12	0.091	0.109	0.061	0.073	0.104	0.083	0.114
Chr 13	0.054	0.061	0.038	0.044	0.052	0.048	0.062
Chr 14	0.083	0.095	0.052	0.059	0.088	0.071	0.086
Chr 15	0.088	0.102	0.060	0.071	0.085	0.078	0.099
Chr 16	0.102	0.120	0.069	0.072	0.107	0.090	0.119
Chr 17	0.158	0.182	0.105	0.112	0.168	0.144	0.183
Chr 18	0.071	0.081	0.049	0.058	0.063	0.059	0.079
Chr 19	0.152	0.175	0.100	0.110	0.189	0.138	0.200
Chr 20	0.117	0.136	0.074	0.084	0.120	0.101	0.131
Chr 21	0.061	0.083	0.047	0.045	0.074	0.062	0.077
Chr 22	0.118	0.131	0.073	0.090	0.122	0.104	0.128
Chr X	0.029	0.039	0.019	0.023	0.041	0.030	0.034
Chr Y	0.001	0.005	0.001	0.003	0.006	0.001	0.004
Overall	**0.082**	**0.097**	**0.054**	**0.063**	**0.084**	**0.073**	**0.094**

Supplementary Table 3. Mean GS scores for functional elements in the *HBB* gene complex.

	Name	Start[a]	Stop[a]	Brain	GI	Lung	Heart	Blood	Muscle	Epithelium
Genes in *HBB* gene complex	*HBB*	5246696	5248301	0.0232	0.0100	0.0017	0.0013	0.7336	0.0008	0.0029
	HBD	5254059	5255858	0.0013	0.0138	0.3161	0.0014	0.8895	0.0010	0.0340
	HBBP1	5263184	5264822	0.0012	0.0097	0.0527	0.0020	0.4709	0.0008	0.0258
	HBG1	5269502	5271087	0.0010	0.0016	0.0017	0.0014	0.0668	0.0008	0.0030
	HBG2	5274421	5276011	0.0010	0.0015	0.0018	0.0013	0.1430	0.0008	0.0033
	HBE1	5289580	5291373	0.0135	0.0007	0.0017	0.0015	0.0003	0.0008	0.0034
CRMs in *HBB* gene complex	3'HS1	5226013	5226493	0.0063	0.0089	0.0346	0.0186	0.0090	0.0130	0.3177
	HBB_3'enh	5245876	5246140	0.0014	0.0005	0.0022	0.0013	0.9062	0.0008	0.0030
	HBB_prom	5248301	5248556	0.0018	0.0150	0.0016	0.0013	0.0003	0.0008	0.0023
	HBD_prom	5255713	5256160	0.0014	0.0005	0.3528	0.0013	0.9999	0.0008	0.0595
	HBG1_3'enh	5268365	5269114	0.0010	0.0005	0.0193	0.0013	0.6284	0.0008	0.0216
	HBG1_prom	5271086	5271290	0.0010	0.0005	0.0017	0.0013	0.0003	0.0008	0.0030
	HBG1_up	5271291	5271813	0.0010	0.0005	0.0201	0.0013	0.0494	0.0008	0.0029
	HBG2_prom	5276011	5276214	0.0010	0.0005	0.0017	0.0013	0.0040	0.0008	0.0030
	HBG2_up	5276215	5276745	0.0010	0.0005	0.0270	0.0013	0.0535	0.0043	0.0204
	HBE1_prom	5291175	5291343	0.0010	0.0011	0.0017	0.0016	0.0003	0.0008	0.0047
	HBE1_up	5291344	5291610	0.0041	0.0006	0.0017	0.0120	0.0008	0.0008	0.0031
	HBE1_PRB	5292690	5292886	0.0010	0.0005	0.0016	0.0021	0.0003	0.0008	0.0015
	HBE1_NRB	5292886	5292928	0.0010	0.0005	0.0017	0.0013	0.0003	0.0008	0.0030
	HBE1_PRA	5293982	5294081	0.0010	0.0005	0.0017	0.0013	0.3566	0.0008	0.0030
	HBE1_NRA	5294082	5294308	0.0010	0.0005	0.0017	0.0013	0.0535	0.0008	0.0030
	HS1	5296894	5297517	0.0010	0.0007	0.0403	0.0013	0.9625	0.0040	0.1062
	HS2_pos	5301795	5302089	0.2231	0.4767	0.0604	0.4046	0.6358	0.0336	0.7855
	HS2_neg	5302090	5302174	0.0833	0.4471	0.0635	0.0094	1.0000	0.0051	0.9195
	HS3	5305882	5306169	0.0010	0.0005	0.0017	0.0013	0.5671	0.0008	0.0044
	HS3.1	5306356	5306418	0.0010	0.0005	0.0022	0.0013	0.0233	0.0008	0.0030
	HS3.2	5306814	5307392	0.0010	0.0005	0.0071	0.0013	0.4680	0.0008	0.9783
	HS4	5309419	5309707	0.0010	0.2965	0.5436	0.0013	0.4414	0.0008	0.9378
	HS5	5312534	5312694	0.0010	0.0163	0.0543	0.0013	0.0890	0.0008	0.8171

[a] hg19 coordinates

Supplementary Table 4. Mean GS scores for functional elements near *MYH6* and *MYH7*.

Name	Location[a]	Brain	GI	Lung	Heart	Blood	Muscle	Epithelium
MYH6	chr14: 23851199-23877486	0.0562	0.0055	0.0052	0.6050	0.0356	0.1749	0.0833
MYH7	chr14: 23881947-23904870	0.2315	0.0199	0.0051	0.9788	0.0028	0.7498	0.0739
mir208a	chr14: 23857805-23857875	0.0010	0.0011	0.0026	0.2331	0.0004	0.0040	0.2066
mir208b	chr14: 23887196-23887272	0.0010	0.0011	0.0028	1.0000	0.0004	0.4221	0.0047
hs1670	chr14: 23906587-23908214	0.9203	0.9660	0.0306	1.0000	0.1639	1.0000	0.2983
hs2330	chr14: 23911613-23912923	0.2340	0.2818	0.0024	0.9999	0.0025	0.9999	0.1892

[a] Coordinates are based on hg19.

Supplementary Table 5. Cell types used for developing GenoSkyline annotations of ESC and fetal cells.

Group	EID[a]	Anatomy	Standardized Epigenome Name	Type
ESC	E002	ESC	ES-WA7 Cells	PrimaryCulture
	E008	ESC	H9 Cells	PrimaryCulture
	E001	ESC	ES-I3 Cells	PrimaryCulture
	E015	ESC	HUES6 Cells	PrimaryCulture
	E014	ESC	HUES48 Cells	PrimaryCulture
	E016	ESC	HUES64 Cells	PrimaryCulture
	E003	ESC	H1 Cells	PrimaryCulture
	E024	ESC	ES-UCSF4 Cells	PrimaryCulture
Fetal cells	E017	LUNG	IMR90 fetal lung fibroblasts Cell Line	CellLine
	E093	THYMUS	Fetal Thymus	PrimaryTissue
	E082	BRAIN	Fetal Brain Female	PrimaryTissue
	E081	BRAIN	Fetal Brain Male	PrimaryTissue
	E089	MUSCLE	Fetal Muscle Trunk	PrimaryTissue
	E090	MUSCLE_LEG	Fetal Muscle Leg	PrimaryTissue
	E083	HEART	Fetal Heart	PrimaryTissue
	E092	GI_STOMACH	Fetal Stomach	PrimaryTissue
	E085	GI_INTESTINE	Fetal Intestine Small	PrimaryTissue
	E084	GI_INTESTINE	Fetal Intestine Large	PrimaryTissue
	E086	KIDNEY	Fetal Kidney	PrimaryTissue
	E088	LUNG	Fetal Lung	PrimaryTissue
	E080	ADRENAL	Fetal Adrenal Gland	PrimaryTissue
	E099	PLACENTA	Placenta Amnion	PrimaryTissue
	E091	PLACENTA	Placenta	PrimaryTissue

[a] Epigenome ID in Roadmap project

Supplementary Table 6. List of 15 complex diseases and traits.

Trait/Disease	Category	Data source	Data Link	Ref.
Schizophrenia	Psychiatric disease	PGC	http://www.med.unc.edu/pgc/downloads	31
Anorexia nervosa	Psychiatric disease	GCAN	http://www.med.unc.edu/pgc/files/resultfiles/gcan_meta-out.gz	44
Coronary artery disease	Cardiovascular disease	CARDIoGRAM	http://www.cardiogramplusc4d.org/downloads/	33
Crohn's disease	IBD	IIBDGC	http://www.ibdgenetics.org/downloads.html	45
Ulcerative colitis	IBD	IIBDGC	http://www.ibdgenetics.org/downloads.html	46
Rheumatoid arthritis	Inflammatory disorder	Stahl et al	http://www.broadinstitute.org/ftp/pub/rheumatoid_arthritis/Stahl_etal_2010NG/	47
LDL cholesterol	Lipids	Global lipids consortium	http://csg.sph.umich.edu/abecasis/public/lipids2013/	48
HDL cholesterol	Lipids	Global lipids consortium	http://csg.sph.umich.edu/abecasis/public/lipids2013/	48
Triglycerides	Lipids	Global lipids consortium	http://csg.sph.umich.edu/abecasis/public/lipids2013/	48
Total cholesterol	Lipids	Global lipids consortium	http://csg.sph.umich.edu/abecasis/public/lipids2013/	48
Type 2 diabetes	Metabolic disorder	DIAGRAM	http://diagram-consortium.org/downloads.html	49
Bone mineral density	Osteoporosis	GEFOS	http://www.gefos.org/?q=content/data-release	50
BMI	Anthropometric traits	GIANT	http://www.broadinstitute.org/collaboration/giant/index.php	51
WHR adjusted for BMI	Anthropometric traits	GIANT	http://www.broadinstitute.org/collaboration/giant/index.php	29
Height	Anthropometric traits	GIANT	http://www.broadinstitute.org/collaboration/giant/index.php	52

Supplementary Table 7. Ranks of top signals under different criteria at 105 schizophrenia-associated loci.

			p-value		NSFP score		Brain GSP		Blood GSP		Heart GSP	
Chr.	Start	Stop	p-value[a]	Rank[b]	Posterior[c]	Rank[b]	Posterior[c]	Rank[b]	Posterior[c]	Rank[b]	Posterior[c]	Rank[b]
	2,372,401	2,402,501	2.84E-03	12084	0.1056	6249	0.1574	2596	0.0200	27259	0.0229	11615
	8,411,184	8,638,984	4.90E-06	714	0.9392	516	0.9416	296	0.9399	376	0.9271	266
	30,412,551	30,437,271	5.84E-06	746	0.8895	636	0.6823	662	0.0434	10109	0.3926	864
	44,029,384	44,128,084	4.84E-03	17544	0.0548	12702	0.0716	5581	0.0168	34368	0.0065	62012
	73,766,426	73,991,366	1.25E-03	7070	0.0043	412743	0.0094	79535	0.0091	76196	0.0137	22944
	97,792,625	98,559,084	5.72E-07	395	0.9874	242	0.9857	143	0.9696	294	0.9739	168
	149,998,890	150,242,490	1.05E-04	1981	0.5870	1211	0.5175	917	0.5468	926	0.3929	863
	177,247,821	177,300,821	2.20E-02	53656	0.0182	59742	0.0474	9045	0.0001	661561	0.0148	20601
	207,912,183	208,024,083	3.81E-04	3724	0.3703	2010	0.3362	1371	0.4604	1084	0.2496	1210
	243,503,719	244,002,945	3.26E-06	654	0.9308	534	0.9082	368	0.9117	442	0.7326	486
2	57,943,593	58,502,192	4.07E-06	679	0.8879	640	0.8684	438	0.7826	605	0.7055	510
2	72,357,335	72,368,185	2.76E-02	64075	0.0322	27702	0.0227	23864	0.0025	252518	0.0155	19397
2	146,416,922	146,441,832	3.66E-03	14379	0.0045	377737	0.0007	548697	0.0053	141404	0.0018	240571
2	149,390,778	149,520,178	9.92E-03	29507	0.0474	15941	0.0074	108959	0.0118	54607	0.0037	124999
2	162,798,555	162,910,255	7.99E-05	1738	0.5070	1434	0.3934	1184	0.4618	1080	0.2412	1253
2	185,601,420	185,785,420	1.84E-04	2629	0.0273	34285	0.0082	95586	0.0613	6660	0.0120	27282
2	193,848,340	194,028,340	3.09E-06	645	0.0073	203243	0.0285	17620	0.0124	51011	0.0324	7481
2	198,148,577	198,835,577	9.97E-04	6166	0.2216	3253	0.2422	1826	0.2971	1598	0.1832	1524
2	200,161,422	200,309,252	9.31E-02	166270	0.0093	146143	0.0178	33178	0.0002	579606	0.0040	114052
2	200,715,237	200,848,037	2.15E-03	9958	0.0865	7519	0.0837	4634	0.0856	4572	0.0400	5813
2	225,334,096	225,467,796	1.35E-02	37243	0.0436	17834	0.0296	16725	0.0579	7117	0.0098	35861
2	233,559,301	233,753,501	2.57E-06	621	0.9587	450	0.9643	244	0.8407	535	0.8133	411
3	2,532,786	2,561,686	7.29E-02	136349	0.0069	221634	0.0065	125958	0.0000	856766	0.0016	270704
3	17,221,366	17,888,266	3.88E-03	14986	0.0630	10203	0.0736	5373	0.0375	12106	0.0233	11371
3	36,843,183	36,945,783	6.26E-06	762	0.8385	698	0.5470	858	0.6562	770	0.4179	816
3	52,541,105	52,903,405	2.25E-06	610	0.9644	427	0.9176	351	0.8988	460	0.9144	282
3	63,792,650	64,004,050	4.04E-04	3818	0.2571	2831	0.1049	3693	0.3103	1525	0.0670	3434
3	135,807,405	136,615,405	1.23E-04	2154	0.4755	1548	0.2693	1660	0.2904	1634	0.2130	1384
3	180,588,843	181,205,585	8.86E-06	836	0.7621	816	0.8113	504	0.7801	608	0.6028	599
4	23,366,403	23,443,403	2.01E-03	9461	0.0203	51863	0.0045	185235	0.0002	583898	0.0203	13681
4	103,146,888	103,198,090	9.03E-03	27523	0.0239	41691	0.0031	252957	0.0038	187692	0.0011	338682
4	170,357,552	170,646,052	2.68E-04	3116	0.4339	1694	0.3201	1427	0.4673	1062	0.2960	1046
4	176,851,001	176,875,801	4.31E-04	3933	0.0645	9890	0.1732	2424	0.0730	5411	0.0455	5034
5	45,291,475	45,393,775	1.14E-02	32785	0.0094	143152	0.0023	310985	0.0002	602326	0.0147	20877
5	60,499,143	60,843,543	8.54E-05	1789	0.5445	1343	0.6016	752	0.4314	1150	0.2724	1106
5	88,581,331	88,854,331	2.90E-02	66539	0.0289	31739	0.0050	166949	0.0217	24568	0.0030	157458
5	109,030,036	109,209,066	4.34E-03	16236	0.0856	7598	0.0529	7906	0.0621	6543	0.0107	31835
5	137,598,121	137,948,092	2.38E-03	10684	0.1077	6135	0.1350	2960	0.0828	4739	0.0628	3661
5	140,023,664	140,222,664	4.43E-03	16465	0.0807	8014	0.0480	8912	0.0667	6031	0.0286	8731
5	151,941,104	152,797,656	7.23E-06	791	0.0932	7022	0.0214	25900	0.0098	69165	0.1669	1623
5	153,671,057	153,688,217	4.31E-02	90406	0.0184	58777	0.0110	64488	0.0054	140385	0.0069	57520
6	26,000,000	34,000,000	4.30E-11	1	1.0000	1	1.0000	1	1.0000	1	0.9999	1
6	73,132,701	73,171,901	1.01E-04	1936	0.5363	1367	0.3914	1187	0.0083	85203	0.1884	1491
6	84,279,922	84,407,274	7.86E-04	5388	0.0280	33156	0.0384	11880	0.0068	110376	0.0031	149664
6	96,300,000	96,500,000	2.35E-03	10575	0.0059	268799	0.0406	11045	0.0018	299188	0.0056	74562
7	1,896,096	2,190,096	5.28E-08	141	0.9967	125	0.9878	132	0.9914	151	0.9910	87
7	24,619,494	24,832,094	5.60E-04	4505	0.2238	3219	0.0886	4384	0.0854	4587	0.1044	2343
7	86,403,226	86,459,326	4.07E-03	15490	0.0242	40865	0.0888	4369	0.0003	520272	0.0050	87823
7	104,598,064	105,063,064	9.81E-05	1900	0.6200	1105	0.6050	745	0.6272	811	0.5428	654
7	110,034,393	110,106,693	1.45E-04	2345	0.0809	7999	0.0263	19576	0.0033	211501	0.1161	2155
7	110,843,815	111,205,915	2.82E-04	3202	0.4147	1784	0.1106	3500	0.2599	1771	0.0357	6676
7	131,539,263	131,567,263	6.65E-04	4941	0.0699	9175	0.0635	6426	0.0076	96657	0.0227	11786
7	137,039,644	137,085,244	2.51E-04	3023	0.1728	4030	0.2320	1885	0.0027	241954	0.0931	2610
8	4,177,794	4,192,544	3.84E-07	333	0.5168	1404	0.7306	599	0.0328	14441	0.1769	1559

3	27,412,627	27,453,627	1.39E-03	7514	0.0433	18028	0.0119	57853	0.0124	51024	0.0063	64220
3	60,475,469	60,954,469	9.82E-04	6114	0.0439	17687	0.0155	40016	0.0281	17459	0.0339	7094
3	89,340,626	89,753,626	1.70E-07	216	0.9642	428	0.9272	335	0.7287	669	0.6640	552
3	111,460,061	111,630,761	3.91E-03	15059	0.0036	487064	0.0035	233126	0.0018	305453	0.0023	203483
3	143,309,503	143,330,533	5.42E-06	736	0.7783	789	0.5864	781	0.1934	2227	0.3955	855
)	84,630,941	84,813,641	7.53E-05	1694	0.5693	1270	0.2540	1746	0.3620	1336	0.0578	3971
.0	18,681,005	18,770,105	8.08E-04	5469	0.0284	32514	0.0382	11985	0.0150	39742	0.1988	1448
.0	104,423,800	104,957,618	2.23E-08	104	0.9989	69	0.9981	42	0.9975	75	0.9971	44
.1	24,367,320	24,412,990	1.17E-04	2107	0.2311	3130	0.0132	50237	0.0951	4058	0.0939	2592
.1	46,342,943	46,751,213	8.95E-04	5801	0.1905	3717	0.2473	1791	0.2683	1734	0.1849	1512
.1	57,386,294	57,682,294	1.66E-05	974	0.8615	676	0.8028	518	0.8586	515	0.7216	495
.1	109,285,471	109,610,071	1.69E-03	8492	0.0109	119181	0.0401	11216	0.0150	39754	0.0163	18188
.1	113,317,794	113,423,994	8.57E-03	26495	0.0260	36705	0.0154	40330	0.0017	310756	0.0030	156299
.1	123,394,636	123,395,986	8.42E-05	1772	0.5832	1228	0.5683	810	0.0691	5783	0.2721	1110
.1	124,610,007	124,620,147	6.39E-02	122871	0.0230	44197	0.0253	20594	0.0235	22066	0.0182	15772
.1	130,714,610	130,749,330	2.17E-04	2853	0.4724	1555	0.0740	5349	0.0100	67071	0.0683	3368
.1	133,808,069	133,852,969	1.99E-04	2736	0.2672	2734	0.4294	1079	0.2092	2101	0.2959	1047
.2	2,321,860	2,523,731	8.92E-07	471	0.9807	321	0.9740	205	0.9322	399	0.9841	119
.2	29,905,265	29,940,365	1.91E-03	9158	0.1361	5014	0.0423	10488	0.0296	16391	0.0410	5651
.2	57,428,314	57,682,971	3.24E-04	3450	0.3985	1857	0.3858	1202	0.4911	1012	0.1840	1515
.2	92,243,186	92,258,286	3.81E-03	14776	0.0195	54405	0.0152	41371	0.0313	15304	0.0066	61029
.2	103,559,855	103,616,655	1.26E-02	35330	0.0044	394586	0.0145	44094	0.0018	303011	0.0020	225327
.2	110,723,245	110,723,245	1.93E-03	9249	0.1474	4640	0.1788	2354	0.2313	1948	0.1828	1525
.2	123,448,113	123,909,113	4.93E-04	4211	0.2575	2825	0.2840	1585	0.3743	1294	0.1796	1539
.4	30,189,985	30,190,316	2.54E-02	60026	0.0376	22470	0.0095	77670	0.0004	502684	0.0019	233466
.4	72,417,326	72,450,526	2.80E-04	3191	0.1055	6258	0.1715	2445	0.0048	155736	0.2152	1375
.4	99,707,919	99,719,219	3.42E-01	490951	0.0082	176098	0.0159	38805	0.0155	38176	0.0058	72338
.4	103,996,234	104,184,834	8.41E-04	5586	0.2379	3040	0.2980	1518	0.2307	1953	0.1838	1517
.5	40,566,759	40,602,237	3.49E-07	323	0.9911	194	0.9702	215	0.9905	160	0.9813	134
.5	61,831,663	61,909,663	4.01E-06	677	0.2979	2468	0.5435	864	0.0516	8187	0.4856	715
.5	70,573,672	70,628,872	3.27E-02	72789	0.0313	28593	0.0102	71003	0.0143	42332	0.0028	165433
.5	78,803,032	78,859,610	6.08E-04	4729	0.2858	2566	0.1636	2519	0.2923	1623	0.2095	1398
.5	84,661,161	85,153,461	4.59E-04	4061	0.2832	2588	0.2511	1766	0.3249	1468	0.1304	1964
.5	91,416,560	91,429,040	2.84E-03	12100	0.1144	5836	0.1359	2936	0.1604	2607	0.0792	2958
.6	9,875,519	9,970,219	6.22E-04	4776	0.2373	3046	0.1860	2263	0.0015	331487	0.0269	9411
.6	13,728,459	13,761,359	1.68E-03	8477	0.0004	892172	0.0007	529431	0.0003	546403	0.0020	228263
.6	29,924,377	30,144,877	2.20E-03	10117	0.1090	6075	0.1205	3246	0.1154	3411	0.0841	2826
.6	58,669,293	58,682,833	3.69E-02	80047	0.0101	131547	0.0001	791714	0.0029	230810	0.0004	547506
.6	67,709,340	68,311,340	3.28E-03	13303	0.1038	6354	0.0860	4515	0.1434	2843	0.0967	2528
.7	2,095,899	2,220,799	9.46E-04	5987	0.2285	3159	0.1995	2127	0.2611	1765	0.2229	1338
.7	17,722,402	18,030,202	5.23E-03	18490	0.0571	11854	0.0801	4875	0.0804	4879	0.0522	4388
.8	52,747,686	53,200,117	2.35E-08	106	0.9989	73	0.9948	77	0.9619	323	0.9885	102
.8	53,453,389	53,804,154	9.33E-04	5943	0.0851	7632	0.0197	29033	0.0082	87830	0.0320	7618
.9	19,374,022	19,658,022	1.57E-04	2439	0.5334	1373	0.5187	914	0.6273	810	0.4148	820
.9	30,981,643	31,039,023	2.56E-02	60492	0.0111	116888	0.0266	19320	0.0005	486139	0.0024	190514
.9	50,067,499	50,135,399	1.61E-04	2465	0.3551	2096	0.3299	1385	0.4655	1069	0.2772	1088
!0	37,361,494	37,485,994	3.69E-04	3671	0.3643	2046	0.4222	1102	0.4345	1144	0.1492	1756
!0	48,114,136	48,131,649	1.85E-03	8994	0.1357	5027	0.0644	6310	0.0110	59613	0.0384	6117
!2	39,975,317	40,016,817	3.67E-04	3661	0.2671	2735	0.1155	3381	0.2309	1950	0.0357	6677
!2	41,408,556	41,675,156	8.39E-06	826	0.6889	971	0.4121	1141	0.8327	544	0.2278	1312
!2	42,315,744	42,689,414	1.85E-04	2635	0.4690	1568	0.4215	1106	0.4540	1097	0.2495	1211

[a] Smallest p-value at risk loci.

[b] Rank of the SNP with the smallest p-value or largest NSFP or GSP score at risk loci.

[c] Largest NSFP score or tissue-specific GSP score at risk loci

Supplementary Table 8. Ranking performance comparison. The value in each cell is the p-value acquired from one-sided binomial test.

	p-value	**NSFP score**	**Brain GSP**	**Blood GSP**	**Heart GSP**
p-value		1.00E+00	9.99E-01	5.00E-01	9.75E-01
NSFP score	1.61E-03[a]		1.00E+00	6.52E-01	1.00E+00
Brain GSP	3.01E-03	8.30E-04		3.92E-02	5.77E-01
Blood GSP	6.52E-01	5.00E-01	9.84E-01		1.00E+00
Heart GSP	5.90E-02	1.61E-03	5.77E-01	1.95E-04	

[a] For example, comparing the ranks for all 105 loci based on NSFP score with the ranks based on p-values in the PGC2011 study, 68 out of 105 loci have an increased rank. The p-value from one-sided binomial test is 1.61E-03, suggesting the better performance of NSFP score.

Supplementary Table 9. Largest GSP scores for each tissue type at CAD-associated risk loci.

Chr	Start	Stop	Brain	GI	Lung	Heart	Blood	Muscle	Epithelium	Tissue[a]
1	109700000	109900000	0.99980	**0.99991**	0.99988	0.99966	0.99981	0.99988	0.99990	GI
2	203600000	204000000	0.99994	**0.99999**	0.99998	0.99996	0.99997	0.99998	0.99998	GI
3	137900000	138200000	0.99468	0.98753	0.98791	**0.99625**	0.97710	0.99517	0.99128	Heart
6	12700000	13100000	0.99989	**0.99997**	0.99959	0.99987	0.99965	0.99988	0.99988	GI
9	21900000	22200000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	NA[b]
10	44400000	44900000	0.61747	**0.83314**	0.65399	0.76752	0.44882	0.81753	0.77258	GI
19	11000000	11400000	0.99630	0.99958	0.99869	0.99074	**0.99967**	0.99937	0.99967	Blood
21	35500000	35700000	0.97074	0.99918	**0.99930**	0.99885	0.99899	0.99638	0.99902	Lung
1	56900000	57100000	0.99604	**0.99699**	0.99606	0.99664	0.98939	0.99418	0.99031	GI
6	34600000	35300000	0.89477	0.95134	0.90695	**0.95156**	0.87966	0.91155	0.94269	Heart
6	134000000	134300000	0.99982	**0.99998**	0.99997	0.99994	0.99988	0.99995	0.99995	GI
7	129600000	129900000	0.99291	0.96390	0.98226	0.99718	0.99872	**0.99950**	0.99832	Muscle
9	136000000	136400000	0.96826	**0.99531**	0.97820	0.96826	0.98821	0.98188	0.98024	GI
10	104400000	105000000	0.91358	**0.94966**	0.92530	0.90768	0.93213	0.91816	0.91192	GI
11	116500000	116700000	0.99672	0.99937	0.99845	0.99761	**0.99970**	0.99917	0.99949	Blood
13	110700000	111200000	0.99273	0.99698	0.99689	0.99597	0.98510	**0.99704**	0.99649	Muscle
14	100000000	100300000	**0.97293**	0.96188	0.91046	0.86511	0.89987	0.92592	0.96228	Brain
15	78900000	79200000	0.99655	0.99927	0.99768	0.99859	0.99795	**0.99934**	0.99883	Muscle
17	2000000	2300000	0.96108	**0.98812**	0.98356	0.97961	0.98167	0.98789	0.98272	GI
17	17400000	18000000	0.93179	0.95199	0.93160	0.94046	**0.96864**	0.95360	0.95436	Blood
17	46800000	47200000	0.94037	0.94009	0.93078	0.90609	**0.96206**	0.94738	0.95791	Blood

[a] The tissue type that provides the largest GSP score.

[b] Not applicable due to ties. A careful comparison between p-values and the GS score pattern is needed in order to infer the relatedness between this locus and tissue types.

Supplementary Table 10. GenoSkyline parameter estimates for brain tissue.

Annotation	Parameter	Raw Estimation[a]	Extra 2,000,000[b]	Extra 6,000,000[c]
-	π	0.11550	0.11413	0.11229
H3K4me1	p_{10}	0.85643	0.85737	0.85852
	p_{11}	0.05999	0.05913	0.05803
H3K4me3	p_{20}	0.40660	0.40582	0.40474
	p_{21}	0.00488	0.00473	0.00453
H3K36me3	p_{30}	0.22752	0.22749	0.22792
	p_{31}	0.09690	0.09540	0.09338
H3K27me3	p_{40}	0.16558	0.16505	0.16395
	p_{41}	0.09713	0.09412	0.08995
H3K9me3	p_{50}	0.02340	0.02309	0.02259
	p_{51}	0.07451	0.07304	0.07106
H3K27ac	p_{60}	0.76178	0.76365	0.76647
	p_{61}	0.02214	0.02175	0.02124
H3K9ac	p_{70}	0.49093	0.49215	0.49405
	p_{71}	0.00527	0.00519	0.00504
DNase I HS	p_{80}	0.22578	0.21962	0.21124
	p_{81}	0.01484	0.01387	0.01255

[a] Estimation results based on the 12,801,840 bases acquired from GWAS catalog.
[b] Estimation results based on the 12,801,840 bases and additional 2,000,000 random bases on chromosome 1.
[c] Estimation results based on the 12,801,840 bases and additional 6,000,000 random bases on chromosome 1.

References

1. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476-82 (2011).
2. Bernstein, B.E. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
3. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
4. Lu, Q. *et al.* A Statistical Framework to Predict Functional Non-Coding Regions in the Human Genome Through Integrated Analysis of Annotation Data. *Sci. Rep.* **5**(2015).
5. Efron, B. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, (Cambridge University Press, 2010).
6. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-7 (2009).
7. Kichaev, G. *et al.* Integrating functional data to prioritize causal variants in statistical fine-mapping studies. (2014).
8. Pickrell, J.K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *The American Journal of Human Genetics* **94**, 559-573 (2014).
9. Chung, D., Yang, C., Li, C., Gelernter, J. & Zhao, H. GPA: a statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation. (2014).
10. Lu, Q., Yao, X., Hu, Y. & Zhao, H. GenoWAP: Post-GWAS Prioritization Through Integrated Analysis of Genomic Functional Annotation. *bioRxiv*, 019539 (2015).
11. Ardlie, K.G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648-660 (2015).
12. Kellis, M. *et al.* Defining functional DNA elements in the human genome. *Proc Natl Acad Sci U S A* (2014).
13. King, D.C. *et al.* Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. *Genome Res* **15**, 1051-60 (2005).
14. Pennacchio, L.A. *et al.* In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499-502 (2006).
15. Kamm, G.B., Pisciottano, F., Kliger, R. & Franchini, L.F. The developmental brain gene NPAS3 contains the largest number of accelerated regulatory sequences in the human genome. *Mol Biol Evol* **30**, 1088-102 (2013).
16. Liang, Y., Ridzon, D., Wong, L. & Chen, C. Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics* **8**, 166 (2007).
17. Jandreski, M.A., Sole, M.J. & Liew, C.C. Two different forms of beta myosin heavy chain are expressed in human striated muscle. *Hum Genet* **77**, 127-31 (1987).

18. Gorza, L. *et al.* Myosin types in the human heart. An immunofluorescence study of normal and hypertrophied atrial and ventricular myocardium. *Circ Res* **54**, 694-702 (1984).
19. Baldwin, K.M. & Haddad, F. Effects of different activity and inactivity paradigms on myosin heavy chain gene expression in striated muscle. *J Appl Physiol (1985)* **90**, 345-57 (2001).
20. Gupta, M.P. Factors controlling cardiac myosin-isoform shift during hypertrophy and heart failure. *J Mol Cell Cardiol* **43**, 388-403 (2007).
21. Miyata, S., Minobe, W., Bristow, M.R. & Leinwand, L.A. Myosin heavy chain isoform expression in the failing and nonfailing human heart. *Circ Res* **86**, 386-90 (2000).
22. Lowes, B.D. *et al.* Myocardial gene expression in dilated cardiomyopathy treated with beta-blocking agents. *N Engl J Med* **346**, 1357-65 (2002).
23. van Rooij, E. *et al.* A family of microRNAs encoded by myosin genes governs myosin expression and muscle performance. *Dev Cell* **17**, 662-73 (2009).
24. Callis, T.E. *et al.* MicroRNA-208a is a regulator of cardiac hypertrophy and conduction in mice. *J Clin Invest* **119**, 2772-86 (2009).
25. VanderMeer, J.E. & Ahituv, N. cis-regulatory mutations are a genetic cause of human limb malformations. *Dev Dyn* **240**, 920-30 (2011).
26. Xie, M. *et al.* DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet* **45**, 836-41 (2013).
27. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics* (2015).
28. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
29. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187-196 (2015).
30. Consortium, S.P.G.-W.A.S. Genome-wide association study identifies five new schizophrenia loci. *Nature genetics* **43**, 969-976 (2011).
31. Consortium, S.W.G.o.t.P.G. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).
32. Gibbs, J.R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* **6**, e1000952 (2010).
33. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature genetics* **43**, 333-338 (2011).
34. Pikuleva, I.A. Cytochrome P450s and cholesterol homeostasis. *Pharmacol Ther* **112**, 761-73 (2006).
35. Russell, D.W. The enzymes, regulation, and genetics of bile acid synthesis. *Annu Rev Biochem* **72**, 137-74 (2003).
36. Consortium, C.D. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics* (2015).
37. Bannister, A.J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell research* **21**, 381-395 (2011).

38. Crawford, G.E. *et al.* Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 992-997 (2004).
39. Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
40. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-2337 (2010).
41. Yin, T., Cook, D. & Lawrence, M. ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biol* **13**, R77 (2012).
42. Kane, M.J., Emerson, J.W. & Weston, S. Scalable Strategies for Computing with Massive Data. *Journal of Statistical Software* **55**, 1-19 (2013).
43. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* **47**, 291-295 (2015).
44. Boraska, V. *et al.* A genome-wide association study of anorexia nervosa. *Molecular psychiatry* **19**, 1085-1094 (2014).
45. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* **42**, 1118-25 (2010).
46. Anderson, C.A. *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* **43**, 246-52 (2011).
47. Stahl, E.A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature genetics* **42**, 508-514 (2010).
48. Consortium, G.L.G. Discovery and refinement of loci associated with lipid levels. *Nature genetics* **45**, 1274-1283 (2013).
49. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics* **44**, 981 (2012).
50. Estrada, K. *et al.* Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nature genetics* **44**, 491-501 (2012).
51. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
52. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics* **46**, 1173-1186 (2014).