Detecting fitness epistasis in recent admixed populations with genome-wide data

CURRENT STATUS: POSTED

Xiaofeng Zhu
Case Western Reserve University School of Medicine

xxz10@case.edu
Corresponding Author
ORCID: 0000-0003-0037-411X

Xumin Ni
Case Western Reserve University School of Medicine

Mengshi Zhou
Case Western Reserve University School of Medicine

Heming Wang
Harvard Medical School

Karen He
Case Western Reserve University School of Medicine

Uli Broeckel
Medical College of Wisconsin

Craig Hanis
University of Texas Health Science Center at Houston

Sharon Kardia
University of Michigan

Susan Redline
Harvard Medical School

Richard Cooper
Loyola University Chicago

Hua Tang
Stanford University
SUBJECT AREAS
Epigenetics & Genomics

KEYWORDS
fitness epistasis, admixed population, admixture linkage disequilibrium, coevolution, diseases/traits
Abstract

Background: Fitness epistasis, the interaction effect of genes at different loci on fitness, has an important contribution for adaptive evolution. Although fitness interaction evidence has been observed in model organisms, it is less detectable and remains poorly understood in human populations owing to the limited statistical power and experimental constraints. Fitness epistasis is inferred from non-independence between unlinked loci. We previously observed ancestral block correlation between chromosomes 4 and 6 in African Americans. The same approach fails when examining ancestral blocks on the same chromosome due to strong confounding effect in a recently admixed population.

Results: We developed a novel approach to eliminate the bias caused by admixture linkage disequilibrium when searching for fitness epistasis on the same chromosome. We applied this approach in 16,252 unrelated African Americans and identified significant ancestral correlations in two pairs of genomic regions (P-value<8.11×10^-7) on chromosomes 1 and 10. The ancestral correlations were not explained by population admixture. Historical African-European crossover events are reduced between pair of epistatic regions. We observed multiple pairs of co-expressed genes between the two regions on each chromosome, including ADAR being co-expressed with IFI44 in almost all tissues and DARC being co-expressed with VCAM1, S1PR1 and ELTD1 in multiple tissues in GTEx. Moreover, the co-expressed gene pairs are associated with the same diseases/traits in the GWAS Catalog, such as white blood cell count, blood pressure, lung function, inflammatory bowel disease and educational attainment.

Conclusions: Our analyses revealed two instances of fitness epistasis on chromosomes 1 and 10, and the findings suggest a potential approach to better understand adaptive evolution.
Background

Epistasis, referring to interactions among genes, has been recognized to play an important role in the etiology of complex diseases[1-3]. Epistasis is an important factor in shaping genetic variance within and between populations, and consequently phenotypic variation[1, 4-6]; it is considered to be one potential reason for missing heritability in genome-wide association studies (GWAS)[7, 8]. Numerous statistical methods for detecting epistasis have been developed in recent years[9-11], including regression-based methods[12, 13], Bayesian statistical methods[14-16], linkage disequilibrium (LD)- and haplotype-based methods[17, 18] and machine-learning and data-mining methods[11, 19]. In general, the existing methods test for pairwise or higher-order interactions through either an exhaustive search of all marker combinations or a reduced marker set in the genome, which lead to a large number of tests and reduced statistical power.

Fitness epistasis refers to the interaction effects among genetic variants at different loci on fitness, and has important consequences for adaptive evolution[20]. The genotype-fitness map, or the fitness landscape as introduced by Sewall Wright[21], is a visualization of a high-dimensional map in which genotypes are organized in the x-y plane and fitness is plotted on the z axis[22]. The shape of the fitness landscape has been considered to have fundamental effects on the course of evolution[23]. Empirical information about the topography of real fitness landscapes have recently emerged from studies of mutations in the β-lactamase TEM1[24], HIV-1 protease and reverse transcriptase[25] and Drosophila melanogaster recombinant inbred lines[26]. However, direct investigation of fitness epistasis in human subjects has thus far been limited[27-29]. Based on the assumption that a functional interactive coevolution could be maintained through complementary mutations over evolutionary history[27, 30], findings from a protein-protein network that used polygenetic distance metrics of a large-scale high-throughput protein-protein
interaction dataset have suggested that Alzheimer’s disease (AD) associated genes *PICALM, BIN1, CD2AP,* and *EPHA1* demonstrate evidence of a pattern of coevolution [29]. A signature of coevolution has also been observe for the killer immunoglobulin receptor (*KIR*) and *HLA* loci, where strong negative correlation exists between the gene frequencies of *KIR* and the corresponding *HLA* ligand [28]. Combinations of *KIR* and *HLA* variants have different degrees of resistance to infectious diseases that affect human survival during epidemics [31].

Fitness epistasis has the potential to generate linkage disequilibrium [32, 33] and affect the efficiency of natural selection [34, 35]. Similarly, our previous study demonstrated that fitness epistasis can create LD among ancestry blocks in recently admixed populations such as African American and Hispanic populations, and this LD is detectable by testing the correlation of local ancestry between two unlinked loci [3]. Since ancestry blocks in recently admixed populations are often long and their frequencies are stable, testing the correlation between local ancestries is more powerful than testing the LD between single nucleotide polymorphisms (SNPs) in the genome by reducing the multiple comparison burden. Ancestry block LD can be generated as a result of population admixture, also termed admixture LD [36, 37]. It is then critical to separate the LD generated by fitness epistasis from admixture LD. To address this challenge, our previous study only searched for fitness epistasis occurring on different chromosomes.

In this study, we developed a statistical approach to eliminate the bias caused by admixture LD when searching for fitness epistasis on the same chromosome. We applied the method in African Americans first by estimating the local ancestral correlation distribution under the null hypothesis that there is no fitness epistasis. Next, we searched for local ancestral correlations departing from the null distribution between two loci within each chromosome. To verify the identified fitness epistasis, we searched for pairs of
tissue-specific co-expressed genes between the two identified regions on each chromosome by utilizing the GTEx V7 cis-eQTL expression dataset[38]. Finally, we examined whether there is an enrichment of diseases/traits associated with genes in the GWAS Catalog[39] within the fitness epistasis regions.

Results

Testing fitness epistasis on the same chromosome

We developed a novel statistical method to detect fitness epistasis on the same chromosome (see MATERIALS AND METHODS). Our basic idea is that the ancestral correlations between two loci after eliminating the effect induced by population admixture suggests fitness epistasis[3]. We applied this method to the African Americans samples in the Candidate gene Association Resource (CARe), Family Blood Pressure Program (FBPP) and Women’s Health Initiative (WHI) cohorts. Our downstream analysis was based on 16,252 unrelated African Americans after removing related individuals and conducting quality controls (Table S1). The distributions of the departure of local ancestral correlations from the expected admixture LD on the same chromosomes are presented in Figure 1(a)-(c) for the three datasets. The distributions of the departure of local ancestral correlations from the null follow an approximately normal distribution for all three datasets. The standard deviation of local ancestral correlations calculated between the pairwise loci located on different chromosomes in FBPP was larger than that of CARe and WHI, which can be attributed to the relatively small sample size of FBPP (Table S2). The QQ-plot of P-values for testing fitness epistasis for CARe, FBPP and WHI are presented in Figure S1. The genomic control parameter were all less than 1, suggesting our approach is conservative.

We conducted meta-analysis to combine the results from these three cohorts. The
genomic control parameter in the meta-analysis was 0.947 (Figure 1(d)). The pairwise correlations of Z-score among these three cohorts ranged from 0.241 to 0.411(Table S3), suggesting our results were relatively consistent among the three cohorts.

We estimated a total of 61,616 independent tests among 1,440,130 pairwise tests performed within a chromosome using the approach by Li and Jin[40]. We used a significance threshold of $P$-value $= 8.11 \times 10^{-7}$ to define potential fitness epistasis in our analysis. After excluding pairwise loci with a genetic distance less than 50 cM, we observed two pairs of genomic regions with significant evidence of fitness epistasis (Table 1). One region pair is chr1:77.32–102.43Mb and chr1:153.22–165.73 Mb and the other is chr10:10.26–24.59Mb and chr10:55.20–73.20Mb. The heatmaps of $-\log_{10}(P$-value) for pairwise loci on chromosomes 1 and 10 are presented in Figures 2–3, respectively. On the heatmap of chromosome 1 (Figure 2 (d)), we observed two significant regions (red regions in Figure 2). But the genetic distance between the pairwise loci in the region in the lower right quadrant was less than 50cM; therefore, we excluded this signal due to the concern that admixture LD was not eliminated entirely. On the heatmap of chromosome 10, we also observed two significant regions in the meta-analysis (Figure 3(d)). However, one of the red regions was near the telomere, which may reflect local ancestry inference errors[41]. Therefore, this region was also excluded from further analyses. In the heatmaps of CARe, FBPP and WHI (Figure 2 (a)-(c) and Figure 3 (a)-(c)), similar heatmap patterns were observed, suggesting that the fitness landscapes in CARe, FBPP and WHI were consistent.

We observed the largest proportion of African ancestry on chr1:153.22–165.73 Mb and the largest proportion of European ancestry on chr10:10.26–24.59Mb (Figure S2). These two regions demonstrate substantial excess of local ancestry and may suggest natural
selection. We calculated the integrated haplotype score (iHS) statistic\cite{42} using \textit{selscan} \cite{43} in the four genomic regions using CARe samples (Figure 4). We observed multiple loci with positive selection evidence (|iHS| > 2) in these four genomic regions. Similar signals could also be observed in ARIC, CARDIA, CFS, JHS and MESA cohorts separately (Figure S3 and Figure S4).

If there were fitness epistasis between two loci on the same chromosome, then we would expect less recombination crossover events (or switch) between African and European chromosomes occurring between these two loci. We calculated the average number of crossovers between African and European chromosomes (ANCAEC) per centiMorgan in the region defined from the right boundary of region 1 and left boundary of region 2 (Table 1) on chromosomes 1 and 10 and then compared with the ANCAEC per centiMorgan in the rest of genome (Table S4). If there is no fitness epistasis between two genomic regions, then we would expect the ANCAEC per centiMorgan between the two regions to follow an approximately normal distribution with the mean and variance estimated from the whole genome data after excluding the two regions. The ANCAEC per centiMorgan between the two detected regions on chromosome 1 is significantly less than that of the rest of the genome (P-value = 7.51×10^{-35}), and similar results were observed on chromosome 10 (P-value = 2.53×10^{-7}), which is consistent with our findings of fitness epistasis in these two chromosomal regions.

\textbf{Co-expression of genes in the two epistatic regions on chromosome 1 and 10}

We hypothesized that the regions demonstrating fitness epistasis will likely harbor co-expressed genes in multiple tissues, attributing to similar functionary of genes. We identified genes residing within the four regions on chromosomes 1 and 10 using the
GENCODE dataset[44]. There are 400, 492, 217 and 211 protein-coding genes in these four regions (chr1:77.32–102.43Mb; chr1:153.22–165.73Mb; chr10:10.26–24.59Mb and chr10:55.20–73.20Mb), respectively. GTEx V7 tissue-specific normalized gene expression matrices and covariates were downloaded from the GTEx Portal (https://www.gtexportal.org/home/datasets). We calculated residuals of gene expression after adjusting for sex, platform, the first three principal components and tissue-specific latent factors inferred by the GTEx consortium using the PEER method[45]. We performed pairwise gene expression correlation analysis using the residuals of gene expression between genes in regions 1 and 2 of chromosome 1. Similar analysis was performed for the gene pairs between genes in regions 1 and 2 of chromosome 10. We applied Bonferroni correction to adjust for the number of tests, which was calculated by the number of independent genes in region 1 multiplied by the number of independent genes in region 2 for a pair of epistatic regions. We calculated the number of independent genes in a region using the approach by Li and Jin[40]. For each tissue, the number of genes expressed in each region varies, but we used the maximum number of independent genes when adjusting for multiple comparisons. Our calculations established the significance levels of $1.689 \times 10^{-6}$ and $5.261 \times 10^{-6}$ for chromosomes 1 and 10, respectively.

We observed 599 pairs of genes that are significantly co-expressed in the epistatic regions on chromosome 1, and 161 pairs of genes that are co-expressed in the epistatic regions on chromosome 10, for at least 1 tissue. We performed a tissue-specific enrichment analysis for these co-expressed genes with the GENE2FUNC option implemented in FUMA[46]. Across 53 tissue types, enrichment test of differentially expressed genes (DEG) showed significantly higher expression of these co-expressed genes in lung (Bonferroni corrected $P < 0.05$) (Figure S5). The heatmaps of the $-\log_{10}(P\text{-value})$ for these co-
expressed gene pairs on chromosomes 1 and 10 are shown in Figures S6-S7, respectively. We observed multiple significantly co-expressed gene pairs in multiple tissues (Figure 5). For example, IFI44 and ADAR are co-expressed in almost all tissues in the GTEx data. We also observed the DARC gene, which encodes the Duffy antigen receptor for human malaria[47], was significantly co-expressed with VCAM1, S1PR1 and ELTD1 in multiple tissues. The proportion of significant co-expressed gene pairs in epistatic regions was substantially higher than the other regions not overlapped with the epistatic regions on chromosome 1 and chromosome 10 (Table S5).

Enrichment of diseases/traits-associated genes from the GWAS Catalog in epistatic regions

GWAS have identified genetic variants that are significantly associated with phenotypes, typically in large sample cohorts. We hypothesized that GWAS hits for the co-expressed gene pairs may have the same disease/phenotype. We examined the GWAS hits using the GWAS Catalog[39] and observed an approximate 2-fold enrichment in region 2 of chromosome 1 compared with the average number of GWAS hits on chromosome 1 (Table S6). We observed 15 pairs of genes associated with the same diseases/traits in chromosomes 1 and 10 (Table 2). Among them, 5 pairs of genes have GWAS hits for multiple traits.

Discussion

In this study, we developed a novel statistical method to detect fitness epistasis by testing the correlation between local ancestries on the same chromosome in a recently admixed population through eliminating bias caused by admixture LD. By applying our method to three large African American cohorts, CARe, FBPP and WHI, we identified two significant epistatic genomic region pairs on chromosomes 1 and 10. These genomic
regions also demonstrated high iHS scores, suggesting signatures of natural selection. We observed that historical recombination events are less likely to occur between a pair of epistatic genomic regions. A large number of gene pairs on the chromosomes 1 and 10 epistatic regions co-express in multiple tissues in the GTEx data. Furthermore, multiple co-expressed gene pairs in these epistatic regions are associated with same diseases/traits in GWAS Catalog.

Several statistical methods for detecting epistasis have been developed, either by exhaustively testing all possible pairwise interactions between SNPs or in a reduced SNP set. The pairwise searching methods that use genotyping array data would require billions of pairwise tests, which are computationally inefficient and result in a high statistical penalty because of the huge multiple testing burden[9]. In our method, we tested pairwise interactions between the ancestral blocks on the same chromosome in a recently admixed population. The current approach can be viewed as an extension of our previous study[3], which focused on pairs of ancestries on different chromosomes. This approach is more powerful because the ancestral blocks are long and often extend to beyond 50 cM[36, 48]. We divided the genome into 400kb bins and used the middle marker of each bin to represent the local ancestries of the corresponding bins[3]. This is reasonable because of the long admixture LD. It is well known that the local ancestries in neighboring bins are highly correlated. Therefore, we applied the widely used method by Li and Ji[40] to calculate the number of independent tests to determine the significance level. Our method could still be conservative because the genomic control values in CARe, FBPP, WHI as well as the meta-analysis were all less than 1. We observed two significant epistatic regions on chromosomes 1 and 10 in the meta-analysis. The general correlation patterns were similar across the stratified analysis in CARe, FBPP and WHI cohorts (Figures 2–3), suggesting our detected fitness epistasis regions are less likely due to chance. The gene pairs that likely
contribute to the detected fitness epistasis are co-expressed in multiple tissues and associated with the same traits.

In the chromosome 1 epistatic regions, *ELTD1* and *DARC* are co-expressed in multiple tissues (Table 2 and Figure 5) and also associated with white blood cell count[49, 50]. Both *ELTD1* and *DARC* have been reported to be under selection pressure[51, 52]. *DARC* encodes the Duffy antigen receptor for human malarial parasites and *ELTD1* plays an essential role in heart development and the prevention of cardiac hypertrophy. The genes *DPH5* and *DAP3* are co-expressed and associated with inflammatory bowel disease (IBD) [53]. IBD is a chronic inflammatory and autoimmune disease that plays an important role in pathogen defense and other functions that are under strong natural selection in humans; thus, the associated genes will exert a negative influence on reproductive fitness[54]. Gene pairs *ELTD1-DCST2* and *MTF2-NDUFS2* are associated with eosinophil counts[49]. Gene pairs *PKN2-ASH1L* and *FRMD4A-REEP3* are associated with red blood cell count[49]. Variation in red and white blood cell count are associated with allergic diseases and certain infections[50, 55, 56], which play important roles in natural selection. We also observed several gene pairs are associated with educational attainment, such as gene pairs *AK5-CADM3, MIR137HG-CADM3*, and *CAMK1D-JMJD1C*[57, 58]. These gene pairs are all co-expressed in brain tissues (see Table 2) and involved in brain-development processes and neuron-to-neuron communication[57]. Other interesting gene pairs associated with the same diseases/traits are showed in Table 2. We observed *IFI44* and *ADAR* were co-expressed in almost all the tissues in the GTEx data (Figure 5). It has been reported that *IFI44* is associated with psychiatric disorders[59], febrile seizures[60], immune response to measles vaccine (measles-specific neutralising antibody titre)[61] and asthma[62], and *ADAR* is associated with Aicardi-Goutières syndrome[63], cerebrospinal fluid levels of Alzheimer’s disease-related proteins[64], lung cancer[65] and
prostate cancer[66]. The above mentioned traits have different degrees of association to human fitness. However, it is not clear how these traits interactively affected human fitness and future research in warrant. It is worth noting that our approach is only applicable to recently admixed populations such as African Americans or Hispanics. One of our future directions is to extend this method to more complex admixed populations, such as the Uygur and Tibetan populations. In addition, the efficiency of our method is influenced by the accuracy of the local ancestry inference. With additional whole genome sequencing data becoming readily available, inference of local ancestry can be further improved. We expect more genomic regions with fitness epistasis will be identified in the near future.

Conclusions

In summary, detecting fitness epistasis is extremely challenging, especially in human populations. Our method takes advantage of a recently admixed population and reliable local ancestry inference using genetic variants from genotyping array data. It is a powerful approach as demonstrated by the real data analysis. Our analyses revealed two fitness epistasis on chromosomes 1 and 10, and the findings provide novel insight for understanding adaptive evolution.

Materials And Methods

Admixture LD in an Admixed Population

In the hybrid isolation model, the admixture LD ($D$) decay between two loci without epistasis can be approximated by an exponential function[67, 68],

[Due to technical limitations, please see supplementary files for formulas]

where $d$ is the genetic distance between the two loci and $t$ is the time elapsed since the initial admixture event (admixture time). Admixture LD decay is more complex in the
continuous gene flow model[36]. However, this exponential function can well mimic our data, as demonstrated in the Women’s Health Initiative (WHI) African American samples (Figure S8). We observed that this exponential function well fits the empirical admixture LD curve. We did observe that there were departures from the fitting line, especially with distance over 50cM, which may be attributed to statistical noise or fitness epistasis. Our goal is to separate fitness epistasis from the statistical noise.

**Estimate the departure from the admixture LD curve**

Let $X_i$ be the local ancestry at locus $i$ and $X_j$ be the local ancestry at locus $j$. We assumed the two loci are located on the same chromosome. We denoted $\beta_{ij}$ as the observed correlation of local ancestries between loci $i$ and $j$.

[See supp. files]

Let $f_d$ be an exponential function representing the admixture LD between two loci with genetic distance ($d$) under no fitness epistasis,

[See supp. files]

where $a = (a_0,a_1,a_2)$ is the vector of parameters in the exponential function. We added a parameter $a_0$, which represents a background LD when the two loci are unlinked.

For each locus $i$, we calculated the correlation of local ancestries $\beta_{ij}$ between loci $i$ and $j$ for all $j \neq i$ on the same chromosome using genotyping array data[3]. We fit a nonlinear regression model by optimizing the following function,

[See supp. files]

We predicted the admixture LD between loci $i$ and $j$ under the null of no fitness epistasis by

[See supp. files]

The departure of observed admixture LD from the expected admixture LD is calculated by

[See supp. files]
The above calculation can also be applied to estimate $\hat{\beta}_{ji}$, that is, given locus $j$, we can estimate $\hat{\alpha}_j$ and therefore $\hat{\beta}_{ji}$ and $\hat{\beta}_{resj}$. In theory, $\beta_{ij} = \beta_{ji}$. But slight variation can be observed because different pairwise loci are applied. Thus, we averaged $\hat{\beta}_{resij}$ and $\hat{\beta}_{resji}$ as the final departure of observed admixture LD from the expected admixture LD when no fitness is present.

[See supp. files]

Testing for fitness epistasis

When there is no fitness epistasis, the departure ( $\hat{\beta}_{res}$) of admixture LD from the null follows a normal distribution $\hat{\beta}_{res} \sim N(0, \sigma^2)$, where $\sigma^2$ is the unknown variance. This variance can be estimated by the local ancestral correlations between two loci on different chromosomes, as suggested by Wang et al.[3]. Since the genetic distance between two loci located on different chromosomes was expected to be infinite, the standard deviation of local ancestral correlations between these loci was therefore served as the population standard deviation of the local ancestral correlations. Thus, we estimated $\sigma$ by using the standard deviation of ancestral correlations among the loci located on the different chromosomes. To test fitness epistasis, we applied a Z-test $Z_{ij} = \frac{\hat{\beta}_{resij}}{\hat{\sigma}}$, with the P-value calculated by $P_{ij} = 2 \times \Phi(-|Z_{ij}|)$.

Dataset

We applied our method to three African American cohorts: (1) the Candidate Gene Association Resources (CARe) study initiated by National Heart, Lung, and Blood Institute (NHLBI), which included 8,367 individuals from five studies: the Atherosclerosis Risk in Communities study (ARIC), the Coronary Artery Risk Development in Young Adults study (CARDIA), the Cleveland Family Study (CFS), the Jackson Heart Study (JHS), and the Multi-Ethnic Study of Atherosclerosis (MESA)[69]. All the samples were genotyped using the
Affymetrix 6.0 platform. These genotype data were downloaded from the dbGaP repository: ARIC: dbGaP phs000280.v1.p1, CARDIA: dbGaP phs000285.v2.p2, CFS: dbGaP phs000284.v1.p1, JHS: dbGaP phs000286.v1.p1, MESA: dbGaP phs000283.v1.p1. (2) the NHLBI Family Blood Pressure Program (FBPP), which collected 3,636 African American subjects from three networks, GenNet, GENOA, and HyperGEN[70], who were genotyped using either Affymetrix 6.0 or Illumina 1M platform; (3) the Women’s Health Initiative (WHI), which includes 8,150 African American postmenopausal women who were genotyped with the Affymetrix 6.0 platform (dbGaP phs000386.v7.p3). QCs were described in Wang et al. (2017)[3]. We excluded related samples and samples with extremely low (≤ 5%) or high (≥ 98%) African proportions. Our downstream analysis was based on 16,252 unrelated African Americans after quality control (Table S1).

We first inferred the local ancestries for the three cohorts with SABER+[71]. SABER+ was designed to reconstruct genetic ancestral blocks in admixed populations based on the Markov-hidden Markov model. Following the analysis procedure of our previous study[3], we divided the genome into 7,389 bins with an average length of 400 kb due to high correlations between adjacent local ancestries. The local ancestry at the middle marker was used to represent the local ancestry of each bin. Due to potentially high local ancestry inference errors on the telomeres and centromeres[41], we excluded bins located within 2Mb of these two types of regions from the analysis. We further performed meta-analysis to combine the results from the three datasets using the weighted Z-score method as described in the METAL software[72]. Finally, we used gene expression data from the GTEx dataset[38] and diseases/traits associations from the GWAS Catalog[39] to strengthen our findings of fitness epistasis.

Declarations
Ethics approval and consent to participate

The study was approved by Case Western Reserve University Institutional Review Board.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflicts of interest.

Availability of data and materials

All data are available on dbGaP: ncbi.nlm.nih.gov/gap

Acknowledgements

We would like to thank Kai Yuan for valuable discussions and suggestions.

Funding

The work was supported by the National Institutes of Health, grants HL086718 and HL053353 from the National Heart, Lung, Blood Institute, and HG003054 from the National Human Genome Research Institute. X. N. was supported by China Scholarship Council (CSC) and the National Natural Science Foundation of China (NSFC) (grants No. 11801027) during his visit to Case Western Reserve University.

CARe: The authors wish to acknowledge the support of the National Heart, Lung, and Blood Institute and the contributions of the research institutions, study investigators, field staff and study participants in creating this resource for biomedical research. The following nine parent studies have contributed parent study data, ancillary study data, and DNA samples through the Broad Institute (N01-HC-65226) to create this genotype/phenotype data base for wide dissemination to the biomedical research community.

Atherosclerotic Risk in Communities (ARIC): University of North Carolina at Chapel Hill (N01-HC-55015), Baylor Medical College (N01-HC-55016), University of Mississippi Medical
Center (N01-HC–55021), University of Minnesota (N01-HC–55019), Johns Hopkins University (N01-HC–55020), University of Texas, Houston (N01-HC–55017), University of North Carolina, Forsyth County (N01-HC–55018);

Cardiovascular Health Study (CHS): University of Washington (N01-HC–85079), Wake Forest University (N01-HC–85080), Johns Hopkins University (N01-HC–85081), University of Pittsburgh (N01-HC–85082), University of California, Davis (N01-HC–85083), University of California, Irvine (N01-HC–85084), New England Medical Center (N01-HC–85085), University of Vermont (N01-HC–85086), Georgetown University (N01-HC–35129), Johns Hopkins University (N01 HC–15103), University of Wisconsin (N01-HC–75150), Geisinger Clinic (N01-HC–45133), University of Washington (N01 HC–55222, U01 HL080295);

Cleveland Family Study (CFS): Case Western Reserve University (R01 HL46380–01–16) and Brigham and Women’s Hospital (HL114473; R35HL135818);

Coronary Artery Risk in Young Adults (CARDIA): University of Alabama at Birmingham (N01-HC–48047), University of Minnesota (N01-HC–48048), Northwestern University (N01-HC–48049), Kaiser Foundation Research Institute (N01-HC–48050), University of Alabama at Birmingham (N01-HC–95095), Tufts-New England Medical Center (N01-HC–45204), Wake Forest University (N01-HC–45205), Harbor-UCLA Research and Education Institute (N01-HC–05187), University of California, Irvine (N01-HC–45134, N01-HC–95100);

Multi-Ethnic Study of Atherosclerosis (MESA): MESA is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC–95159 through N01-HC–95169 and UL1-RR–024156. Funding for genotyping was provided by NHLBI Contract N02-HL–6–4278 and N01-HC–65226.

FBPP-Axiom study is supported by the National Institutes of Health, grant number HL086718 from National Heart, Lung, Blood Institute.
**GENOA**: Genetic Epidemiology Network of Arteriopathy (GENOA) study is supported by the National Institutes of Health, grant numbers HL087660 and HL100245 from the National Heart, Lung, Blood Institute.

**HyperGEN**: The hypertension network is funded by cooperative agreements (U10) with NHLBI: HL54471, HL54472, HL54473, HL54495, HL54496, HL54497, HL54509, HL54515, and 2 R01 HL55673-12. The study involves: University of Utah (Network Coordinating Center, Field Center, and Molecular Genetics Lab); Univ. of Alabama at Birmingham (Field Center and Echo Coordinating and Analysis Center); Medical College of Wisconsin (Echo Genotyping Lab); Boston University (Field Center); University of Minnesota (Field Center and Biochemistry Lab); University of North Carolina (Field Center); Washington University (Data Coordinating Center); Weil Cornell Medical College (Echo Reading Center); National Heart, Lung, & Blood Institute. For a complete list of HyperGEN Investigators please see: www.biostat.wustl.edu/hypergen/Acknowledge.html

**WHI**: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221.

**Authors’ contributions**

X. Z. conceived and designed the study. X. N. and X. Z. developed the statistical method. X. N. and M. Z. performed the data analysis. X. N. and X. Z. wrote the manuscript. All co-authors substantively revised the manuscript and approved the submitted version.

**References**

1. Wang X, Elston RC, Zhu X: *The meaning of interaction*. Hum Hered 2010, 70:269-277.

2. Zhang J: *Epistasis Analysis Goes Genome-Wide*. PLoS Genet 2017, 13:e1006558.
3. Wang HM, Choi Y, Tayo B, Wang X, Morris N, Zhang X, Broeckel U, Hanis C, Kardia S, Redline S, et al: *Genome-wide survey in African Americans demonstrates potential epistasis of fitness in the human genome. Genetic Epidemiology* 2017, 41:122-135.

4. Presgraves DC: *The molecular evolutionary basis of species formation. Nat Rev Genet* 2010, 11:175-180.

5. Carlbom O, Haley CS: *Epistasis: too often neglected in complex trait studies? Nature Reviews Genetics* 2004, 5:618-U614.

6. Cutter AD: *The polymorphic prelude to Bateson-Dobzhansky-Muller incompatibilities. Trends Ecol Evol* 2012, 27:209-218.

7. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al: *Finding the missing heritability of complex diseases. Nature* 2009, 461:747-753.

8. Zuk O, Hechter E, Sunyaev SR, Lander ES: *The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci U S A* 2012, 109:1193-1198.

9. Wei WH, Hemani G, Haley CS: *Detecting epistasis in human complex traits. Nat Rev Genet* 2014, 15:722-733.

10. Upton A, Trelles O, Cornejo-Garcia JA, Perkins JR: *Review: High-performance computing to detect epistasis in genome scale data sets. Brief Bioinform* 2016, 17:368-379.

11. Niel C, Sinoquet C, Dina C, Rocheleau G: *A survey about methods dedicated to epistasis detection. Frontiers in Genetics* 2015, 6.

12. Cordell HJ: *Epistasis: what it means, what it doesn’t mean, and statistical methods to detect it in humans. Human Molecular Genetics* 2002, 11:2463-2468.

13. Cordell HJ: *Detecting gene-gene interactions that underlie human diseases. Nature Reviews Genetics* 2009, 10:392-404.

14. Zhang Y, Liu JS: *Bayesian inference of epistatic interactions in case-control studies.*
Nature Genetics 2007, 39:1167-1173.

15.Yi NJ, Kaklamani VG, Pasche B: Bayesian Analysis of Genetic Interactions in Case-control Studies, with Application to Adiponectin Genes and Colorectal Cancer Risk. Annals of Human Genetics 2011, 75:90-104.

16.Zhang Y: A Novel Bayesian Graphical Model for Genome-Wide Multi-SNP Association Mapping. Genetic Epidemiology 2012, 36:36-47.

17.Prabhu S, Pe’er I: Ultrafast genome-wide scan for SNP-SNP interactions in common complex disease. Genome Research 2012, 22:2230-2240.

18.Ueki M, Cordell HJ: Improved Statistics for Genome-Wide Interaction Analysis. Plos Genetics 2012, 8:141-159.

19.Wan X, Yang C, Yang Q, Xue H, Tang NLS, Yu WC: Predictive rule inference for epistatic interaction detection in genome-wide association studies. Bioinformatics 2010, 26:30-37.

20.da Silva J, Coetzer M, Nedellec R, Pastore C, Mosier DE: Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region. Genetics 2010, 185:293-303.

21.Wright S: The roles of mutation, inbreeding, crossbreeding, and selection in evolution. na; 1932.

22.de Visser JA, Krug J: Empirical fitness landscapes and the predictability of evolution. Nat Rev Genet 2014, 15:480-490.

23.Phillips PC: Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet 2008, 9:855-867.

24.Schenk MF, Szendro IG, Salverda ML, Krug J, de Visser JA: Patterns of Epistasis between beneficial mutations in an antibiotic resistance gene. Mol Biol Evol 2013, 30:1779-1787.

25.Hinkley T, Martins J, Chappey C, Haddad M, Stawiski E, Whitcomb JM, Petropoulos CJ, Bonhoeffer S: A systems analysis of mutational effects in HIV-1 protease and reverse
transcriptase. Nat Genet 2011, 43:487-489.

26. Corbett-Detig RB, Zhou J, Clark AG, Hartl DL, Ayroles JF: Genetic incompatibilities are widespread within species. Nature 2013, 504:135-137.

27. Rohlfs RV, Swanson WJ, Weir BS: Detecting coevolution through allelic association between physically unlinked loci. Am J Hum Genet 2010, 86:674-685.

28. Single RM, Martin MP, Gao X, Meyer D, Yeager M, Kidd JR, Kidd KK, Carrington M: Global diversity and evidence for coevolution of KIR and HLA. Nat Genet 2007, 39:1114-1119.

29. Raj T, Shulman JM, Keenan BT, Chibnik LB, Evans DA, Bennett DA, Stranger BE, De Jager PL: Alzheimer disease susceptibility loci: evidence for a protein network under natural selection. Am J Hum Genet 2012, 90:720-726.

30. Jothi R, Cherukuri PF, Tasneem A, Przytycka TM: Co-evolutionary analysis of domains in interacting proteins reveals insights into domain-domain interactions mediating protein-protein interactions. J Mol Biol 2006, 362:861-875.

31. Parham P: MHC class I molecules and KIRs in human history, health and survival. Nat Rev Immunol 2005, 5:201-214.

32. Kimura M: A model of a genetic system which leads to closer linkage by natural selection. Evolution 1956:278-287.

33. Lewontin R, Kojima Ki: The evolutionary dynamics of complex polymorphisms. Evolution 1960, 14:458-472.

34. Felsenstein J: Sex and the evolution of recombination. The evolution of sex 1988:74-86.

35. Kondrashov A: Classification of hypotheses on the advantage of amphimixis. Journal of Heredity 1993, 84:372-387.

36. Zhu XF, Cooper RS, Elston RC: Linkage analysis of a complex disease through use of admixed populations. American Journal of Human Genetics 2004, 74:1136-1153.

37. Montana G, Pritchard JK: Statistical tests for admixture mapping with case-control and
cases-only data. Am J Hum Genet 2004, 75:771–789.

38. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, et al: The Genotype-Tissue Expression (GTEx) project. Nature Genetics 2013, 45:580–585.

39. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, et al: The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 2019, 47:D1005-D1012.

40. Li J, Ji L: Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity 2005, 95:221–227.

41. Bhatia G, Tandon A, Patterson N, Aldrich MC, Ambrosone CB, Amos C, Bandera EV, Berndt SI, Bernstein L, Blot WJ: Genome-wide scan of 29,141 African Americans finds no evidence of directional selection since admixture. The American Journal of Human Genetics 2014, 95:437–444.

42. Voight BF, Kudaravalli S, Wen X, Pritchard JK: A map of recent positive selection in the human genome. PLoS Biol 2006, 4:e72.

43. Szpiech ZA, Hernandez RD: selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol 2014, 31:2824–2827.

44. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, et al: GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Research 2019, 47:D766-D773.

45. Stegle O, Parts L, Piipari M, Winn J, Durbin R: Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nature Protocols 2012, 7:500–507.

46. Watanabe K, Taskesen E, van Bochoven A, Posthuma D: Functional mapping and
annotation of genetic associations with FUMA. Nat Commun 2017, 8:1826.

47.Cole-Tobian J, King CL: Diversity and natural selection in Plasmodium vivax Duffy binding protein gene. Mol Biochem Parasitol 2003, 127:121-132.

48.Zhu XF, Zhang SL, Tang H, Cooper R: A classical likelihood based approach for admixture mapping using EM algorithm. Human Genetics 2006, 120:431-445.

49.Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, Schoech A, Pasaniuc B, Price AL: Leveraging Polygenic Functional Enrichment to Improve GWAS Power. Am J Hum Genet 2019, 104:65-75.

50.Astile WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, Mead D, Bouman H, Riveros-Mckay F, Kostadima MA, et al: The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. Cell 2016, 167:1415-1429 e1419.

51.McManus KF, Taravella AM, Henn BM, Bustamante CD, Sikora M, Cornejo OE: Population genetic analysis of the DARC locus (Duffy) reveals adaptation from standing variation associated with malaria resistance in humans. PLoS Genet 2017, 13:e1006560.

52.Eichstaedt CA, Antao T, Pagani L, Cardona A, Kivisild T, Mormina M: The Andean adaptive toolkit to counteract high altitude maladaptation: genome-wide and phenotypic analysis of the Collas. PLoS One 2014, 9:e93314.

53.Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, Ripke S, Lee JC, Jostins L, Shah T, et al: Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 2015, 47:979-986.

54.Brinkworth JF, Barreiro LB: The contribution of natural selection to present-day susceptibility to chronic inflammatory and autoimmune disease. Curr Opin Immunol 2014, 31:66-78.

55.Kay AJTjoid: The eosinophil in infectious diseases. 1974, 129:606-613.
56. Galli SJ, Tsai M, Piliponsky AMJN: *The development of allergic inflammation*. 2008, 454:445.

57. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, Sidorenko J, Karlsson Linner R, et al: *Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals*. Nat Genet 2018, 50:1112-1121.

58. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P, Chen GB, Emilsson V, Meddens SF, et al: *Genome-wide association study identifies 74 loci associated with educational attainment*. Nature 2016, 533:539-542.

59. Cross-Disorder Group of the Psychiatric Genomics C, Genetic Risk Outcome of Psychosis C: *Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis*. Lancet 2013, 381:1371-1379.

60. Feenstra B, Pasternak B, Geller F, Carstensen L, Wang T, Huang F, Eitson JL, Hollegaard MV, Svanstrom H, Vestergaard M, et al: *Common variants associated with general and MMR vaccine-related febrile seizures*. Nat Genet 2014, 46:1274-1282.

61. Haralambieva IH, Ovsyannikova IG, Kennedy RB, Larrabee BR, Zimmermann MT, Grill DE, Schaid DJ, Poland GA: *Genome-wide associations of CD46 and IFI44L genetic variants with neutralizing antibody response to measles vaccine*. Human Genetics 2017, 136:421-435.

62. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmuller J, Ang W, Barr RG, Beaty TH, Becker AB, Beilby J, et al: *Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks*. Nature Genetics 2018, 50:42-+.

63. Rice GI, Forte GMA, Szynkiewicz M, Chase DS, Aeby A, Abdel-Hamid MS, Ackroyd S, Allcock R, Bailey KM, Balottin U, et al: *Assessment of interferon-related biomarkers in*
Aicardi-Goutieres syndrome associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, and ADAR: a case-control study. Lancet Neurology 2013, 12:1159-1169.

64. Kauwe JSK, Bailey MH, Ridge PG, Perry R, Wadsworth ME, Hoyt KL, Staley LA, Karch CM, Harari O, Cruchaga C, et al: Genome-Wide Association Study of CSF Levels of 59 Alzheimer’s Disease Candidate Proteins: Significant Associations with Proteins Involved in Amyloid Processing and Inflammation. Plos Genetics 2014, 10.

65. McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, Caporaso NE, Johansson M, Xiao X, Li Y, et al: Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nat Genet 2017, 49:1126-1132.

66. Eeles RA, Al Olama AA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, Ghoussaini M, Luccarini C, Dennis J, Jugurnauth-Little S, et al: Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nature Genetics 2013, 45:385-391.

67. Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, Hutchinson RG, Ferrell RE, Boerwinkle E, Shriver MD: Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. Am J Hum Genet 2001, 68:198-207.

68. Chakraborty R, Weiss KM: Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc Natl Acad Sci U S A 1988, 85:9119-9123.

69. Zhu XF, Young JH, Fox E, Keating BJ, Franceschini N, Kang SJ, Tayo B, Adeyemo A, Sun YV, Li YL, et al: Combined admixture mapping and association analysis identifies a novel blood pressure genetic locus on 5p13: contributions from the CARe consortium. Human
Molecular Genetics 2011, 20:2285–2295.

70. Williams RR, Rao DC, Ellison RC, Arnett DK, Heiss G, Oberman A, Eckfeldt JH, Leppert MF, Province MA, Mockrin SC, Hunt SC: *NHLBI family blood pressure program: methodology and recruitment in the HyperGEN network*. Hypertension genetic epidemiology network. *Ann Epidemiol* 2000, 10:389-400.

71. Tang H, Coram M, Wang P, Zhu X, Risch N: *Reconstructing genetic ancestry blocks in admixed individuals*. *Am J Hum Genet* 2006, 79:1-12.

72. Willer CJ, Li Y, Abecasis GR: *METAL: fast and efficient meta-analysis of genomewide association scans*. *Bioinformatics* 2010, 26:2190-2191.

Tables

Table 1. Significantly epistatic region pairs on the same chromosome.

| Chromosome | Region 1 (Mb) | Region 2 (Mb) |
|------------|---------------|---------------|
| Chr 1      | 77.32-102.43  | 153.22-165.73 |
| Chr 10     | 10.26-24.59   | 55.20-73.20   |

Table 2. Co-expressed gene pairs and their common associated diseases/traits.
| Chr  | Gene in region 1 | Gene in region 2 | Disease/trait                                      | Significant tissues                                                                 |
|------|------------------|------------------|---------------------------------------------------|--------------------------------------------------------------------------------------|
| Chr 1| AK5              | CADM3            | Educational attainment (years of education)        | Brain Nucleus accumbens basal ganglia                                                |
|      | DPH5             | DAP3             | Inflammatory bowel disease                         | Uterus                                                                              |
|      | ELTD1            | DARC             | White blood cell count                             | Artery Aorta; Artery Tibial; Colon Transverse; Esophagus Gastroesophageal Junction; Esophagus Mucosa; Ovary; Skin Not Sun Exposed Suprapubic; Vagina Adipose Subcutaneous |
|      | ELTD1            | DCST2            | Eosinophil counts                                 | Brain Hypothalamus                                                                  |
|      | GFI1             | SLAMF7           | Multiple sclerosis                                 | Heart Atrial Appendage; Ovary; Uterus                                               |
|      | MIR137HG         | CADM3            | Educational attainment (years of education)        | Brain Hypothalamus                                                                  |
|      | MTF2             | NDUFS2           | Eosinophil counts                                 | Brain Hippocampus                                                                   |
|      | PKN2             | ASH1L            | Red blood cell count                               | Skin Not Sun Exposed Suprapubic                                                      |
|      | CAMK1D           | JMJD1C           | General cognitive ability                          | Brain Hypothalamus                                                                  |
|      | CAMK1D           | JMJD1C           | Educational attainment                             | Brain Hypothalamus                                                                  |
|      | CAMK1D           | JMJD1C           | Educational attainment (years of education)        | Brain Hypothalamus                                                                  |
|      | CAMK1D           | JMJD1C           | Highest math class taken                           | Brain Hypothalamus                                                                  |
|      | CAMK1D           | JMJD1C           | Lung function (FEV1/FVC)                          | Brain Hypothalamus                                                                  |
|      | CAMK1D           | JMJD1C           | Educational attainment                             | Brain Hypothalamus                                                                  |
|      | CELF2            | CCDC6            | Systolic blood pressure                            | Spleen                                                                              |
|      | CELF2            | CCDC6            | Pulse pressure                                     | Spleen                                                                              |
|      | FAM107B          | CTNNA3           | Night sleep phenotypes                             | Brain Caudate basal ganglia; Brain Cortex; Brain Nucleus accumbens basal ganglia; Brain Putamen basal ganglia Breast Mammary Tissue |
|      | FRMD4A           | REEP3            | Red blood cell count                               | Brain Caudate basal ganglia; Brain Cortex; Brain Nucleus accumbens basal ganglia; Brain Putamen basal ganglia Breast Mammary Tissue |
|      | NEBL             | JMJD1C           | Interleukin-10 levels                              | Pituitary                                                                           |
|      | NEBL             | JMJD1C           | Lung function (FEV1/FVC)                           | Pituitary                                                                           |
|      | PIP4K2A          | CTNNA3           | Breast cancer                                      | Brain Caudate basal ganglia; Brain Cortex; Brain Nucleus accumbens basal ganglia    |
|      | PIP4K2A          | CTNNA3           | Obesity-related traits                              | Brain Caudate basal ganglia; Brain Cortex; Brain Nucleus accumbens basal ganglia    |
|      | PLXDC2           | BICC1            | Heel bone mineral density                          | Stomach                                                                             |
|      | PLXDC2           | BICC1            | Pulse pressure                                     | Stomach                                                                             |

Figures
Figure 1

Distributions of the departure of local ancestral correlations from null and the corresponding statistical evidence. Distributions of the departure of local ancestral correlations from null in (a) CARe, (b) FBPP and (c) WHI. (d) QQ-plot of P-values in meta-analysis.
Heatmap of $-\log_{10}(P\text{-value})$ between pairwise loci located on chromosome 1 in (a) CARe, (b) FBPP, (c) WHI and (d) meta-analysis. Each point represents the $-\log_{10}(P\text{-value})$ between two loci. In (a), (b) and (c), if $-\log_{10}(P\text{-value})$ is larger than 6, we set the value as 6. In meta-analysis (d), if $-\log_{10}(P\text{-value})$ is larger than $-\log_{10}(\text{significant level})$, we set the value as 7, which reaches the significant level.
Figure 3

Heatmap of \(-\log_{10}(P\text{-value})\) between pairwise loci located on chromosome 10 in (a) CARe, (b) FBPP, (c) WHI and (d) meta-analysis. Each point represents the \(-\log_{10}(P\text{-value})\) between two loci. In (a), (b) and (c), if \(-\log_{10}(P\text{-value})\) is larger than 6, we set the value as 6. In meta-analysis (d), if \(-\log_{10}(P\text{-value})\) is larger than \(-\log_{10}(\text{significant level})\), we set the value as 7, which reaches the significant level.
The recent selection signal ($|\text{iHS}| > 2$) on the epistatic regions in CARe cohort. (a) and (b) are the selection signal on region 1 (chr1:77.32-102.43Mb) and region 2 (chr1:153.22-165.73 Mb) on chromosome 1, respectively. (c) and (d) are the selection signal on region 1 (chr10:10.26-24.59Mb) and region 2 (chr10:55.20-73.20Mb) on chromosome 10, respectively.
Figure 5

Heatmap of P-values of significantly co-expressed gene pairs located on (a) chromosome 1 and (b) chromosome 10 in different tissues. Y-axis represents the names of different tissues. X-axis represents the names of gene pairs. These gene pairs are significantly co-expressed in more than 2 tissues. Red block represents the significant signals.

Supplementary Files

This is a list of supplementary files associated with the primary manuscript. Click to download.

Methods - formulas.docx
SupplementaryFile_Oct2.docx