Revealing heritability missed by single locus analyses by examining epistasis in mouse chromosome substitution strains

Anna K. Miller[*], Anlu Chen[†1], Jacquelaine Bartlett[‡], Li Wang[*], Scott M. Williams[*‡§] and David A. Buchner[*†§]

[*]Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH 44106

[†]Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland, OH 44106

[‡]Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH 44106

§ These authors contributed equally

[1]Division of Endocrinology, Diabetes, and Metabolism, Beth Israel Deaconess Medical Center, and Harvard Medical School, Boston, MA 02115

Data availability: accession number GSE145607, token “odehouushpaflqz”

Running title: Epistasis in complex traits

Key words: QTL, complex trait, missing heritability, genetic architecture, consomic strains

Corresponding author's contact information:

David A. Buchner, Case Western Reserve University School of Medicine, 10900 Euclid Ave., Cleveland, OH, 44106-4935, 216-368-1816, dab22@case.edu

Abstract

The genetic contribution of additive versus non-additive (epistasis) effects in the regulation of hematologic and other complex traits is unclear. While genome-wide association studies typically ignore gene-gene interactions, in part because of the lack of statistical power for detecting them, mouse chromosome substitution strains (CSSs) represent an alternate and powerful model for detecting epistasis given their limited allelic variation. Therefore, we utilized CSSs to identify and map both additive and epistatic loci that regulate a range of hematologic- and metabolism-related traits, as well as hepatic gene expression. Quantitative trait loci (QTLs) were identified using a modified backcross strategy involving the segregation of variants on the A/J-derived substituted chromosomes 4 and 6 on an otherwise C57BL/6J genetic background. By analyzing the transcriptomes of offspring from this cross, we identified and mapped additive QTLs regulating the expression of 768 genes, and epistatic QTL pairs for 519 genes. Similarly, we identified additive QTLs for fat pad weight, platelets, and the percentage of granulocytes in blood, as well as epistatic QTL pairs controlling the percentage of lymphocytes in blood and red cell distribution width. The variance attributed to the epistatic QTLs was approximately equal to that of the additive QTLs, demonstrating the importance of identifying genetic interactions to understand the genetic architecture of complex traits. Of particular note, even the epistatic QTLs that accounted for the largest variances were undetected in our single loci GWAS-like association analyses, highlighting the need to account for epistasis in association studies.

Introduction

The severity, presentation, and prevalence of complex traits and diseases are influenced by many genetic variants (Fu et al. 2013; Boyle et al. 2017). However, it remains unclear whether the variants work together in an additive manner or have non-linear effects on the phenotype, referred to as epistasis (Bloom et al. 2013; Jasnos and Korona 2007; Cordell 2002; Carlborg and Haley 2004; Manolio et al. 2009). Epistasis has been extensively observed in model organisms including *S. cerevisiae* (Bloom et al. 2013; Jasnos and Korona 2007; Forsberg et al. 2017), *C. elegans* (Lehner et al. 2006), *D. melanogaster* (Huang et al. 2012), and *M. musculus* (Shao et al. 2008; Tyler et al. 2017). In humans, epistasis has been more difficult to detect with standard genetic analyses, potentially due to low allele frequencies, limited sample sizes, complexity of interactions, insufficient effect sizes, diversity of genetic backgrounds, and methodological limitations due to the fact that in humans evidence for genetic effects is almost always statistical (Mackay 2014; Huang and Mackay 2016). However, despite these inherent difficulties, epistasis in humans has been detected in genome-wide interaction-based association studies and other methods for Crohn's disease, Glaucoma, Behçet's disease, multiple sclerosis, Hirschbrung disease, among others (Liu et al. 2011; Verma et al. 2016; Huang et al. 2015; Kirino et al. 2013; Hemani et al. 2014; Hu et al. 2013; Galarza-Munoz et al. 2017; Chatterjee and Chakravarti 2019; Tyler et al. 2020). Assuming these examples generalize, discovering and understanding epistatic interactions will be critical for improving the understanding of pathophysiology as well as the accuracy of predicting

phenotypes from genetic data, thereby better guiding precision medicine-based decisions (Moore and Williams 2009; Cole et al. 2017).

One particularly powerful method of detecting epistasis in a mammalian system is based on studies of mouse chromosome substitution strains (CSSs) (Nadeau et al. 2000). The genomes of CSSs are comprised of one chromosome derived from a donor strain, with the remainder derived from a host strain such that when two CSSs are crossed together, any genetic interactions between the two non-homologous donor chromosomes can be more readily detected. This system provides an efficient model to map interacting quantitative trait loci (QTLs) on a fixed genetic background. This is in contrast to populations with genome-wide allelic variation that segregate in the mapping population such as the Diversity Outbred mouse population or typical analyses in humans (Churchill et al. 2012). Specifically, as it relates to epistasis, what might otherwise be a rare allelic combination in a segregating cross can be fixed in the genome of a CSS, allowing for detailed and reproducible studies of specific allelic combinations. Initial studies of CSSs derived from strains C57BL/6J and A/J, which differ with respect to metabolic disease and cancer risk, indicated that epistatic interactions were a dominant feature of many complex traits (Shao et al. 2008).

More recent work on CSS combinations with two substituted chromosomes again highlighted the importance of epistasis in the regulation of complex traits and gene expression, and built upon previous studies by directly identifying pairs of QTLs (chromosomes) with non-additive phenotypic effects. For example, in mice carrying two

A/J-derived chromosomes on an otherwise C57BL/6J genetic background, the A/J chromosomes frequently showed evidence of non-additive interactions. The patterns of interactions were largely in the direction of negative (or suppressing) epistasis, such that traits that were altered due to the effects of one of the substituted chromosomes were returned to baseline levels due to the combined effects of the two interacting substituted chromosomes (Chen et al. 2017). These findings indicated that epistatic interactions can be a common feature of complex traits as well as important for maintaining homeostasis.

A major limitation to previous analyses of epistasis using CSSs is that the resolution of QTLs mapped was the size of the entire substituted chromosomes. However, to identify the molecular nature of such QTLs requires identification of the underlying genes, which necessitates higher resolution mapping studies. Therefore, to identify novel epistatic QTLs that regulate complex traits and gene expression levels with greater mapping resolution, we generated and analyzed the N2 generation mice from a modified backcross based on using some of the same CSSs described above. The resulting analyses demonstrate the ability of a CSS-based backcross to detect and fine-map epistatic QTLs for complex traits. The analyses discovered significant epistatic interactions that control lymphocyte quantity and the distribution of red cell width, and the expression of hundreds of genes in the liver. These studies are essential to make progress towards the molecular characterization of epistatic interactions that will provide insight into the biological pathways of disease pathophysiology as well as improve our understanding of trait heritability and genetic architecture.

Materials and Methods

Mice

Strains C57BL/6J-Chr6$^{A/J/}$NaJ (B6.A6) (stock #004384), C57BL/6J-Chr4$^{A/J/}$NaJ (B6.A4) (stock #004382), and C57BL/6J (B6) (stock #000664) were purchased from The Jackson Laboratory. Mice were maintained by brother-sister matings with offspring weaned at 3 weeks of age. Mice were housed in ventilated racks with access to food and water *ad libitum* and maintained at 21ºC on a 12-hour light/12-hour dark cycle. All mice were cared for as described under the Guide for the Care and Use of Animals, eighth edition (2011) and all experiments were approved by IACUC (protocol #2016-0064) and carried out in an AAALAC approved facility. Mice were anesthetized with isoflurane prior to retro-orbital bleeding and subsequently euthanized under anesthesia by cervical dislocation for tissue collection.

Genotyping

Genomic DNA was isolated from mouse tail tissue using DNeasy Blood and Tissue Kits (Qiagen) following digestion with proteinase K. N2 offspring of the CSS backcross were genotyped with 57 SNP markers on Chr. 4 and 50 SNP markers on Chr. 6 (~1 marker per 1.5 cM) using the PlexSeq technology (Agriplex Genomics Inc.) (Kayima et al. 2017; Soong et al. 2018). Genotyping results are listed in Supplemental Table 1. Of the 150 samples, one sample did not pass the PlexSeq quality control (<66% of genotypes called) and was not used in these analyses.

Mouse Phenotyping

Five-week old mice were fasted overnight prior to phenotypic analysis. Mice were anesthetized with isoflurane and measurements were collected for total body weight and nose to anus length. Body mass index (BMI) was calculated as g/cm^2. Blood was collected retro-orbitally and glucose levels were measured using an OneTouch Ultra2 meter (LifeScan, Milpitas, CA, USA). Mice were subsequently euthanized by cervical dislocation, gonadal fat pads were removed and weighed, and the caudate lobe of the liver was collected and immediately placed in RNAlater Stabilization Solution (Thermo Fisher Scientific). Complete blood counts were performed on either a Heska HemaTrue or Drew Scientific Hemavet 950 blood analyzer and included measurements of white blood count (wbc), neutrophils (ne), lymphocytes (ly), monocytes (mo), red blood count (rbc), hemoglobin (hb), hematocrit (hct), mean corpuscular volume (mcv), mean corpuscular hemoglobin (mch), mean corpuscular hemoglobin concentration (mchc), red cell distribution width (rdw), platelet count (plt), mean platelet volume (mpv), granulocyte % (gran%), granulocyte KuL (gran KuL), and red cell distribution width (absolute) (rdwa).

Transcriptome analysis

Total RNA was isolated from liver using PureLink RNA mini kit (Thermo Fisher Scientific). RNA quality was analyzed using an Agilent TapeStation with all resulting RNA Integrity Numbers (RIN) greater than 8.5. Sequencing libraries were generated from total RNA (1 ug) using the TruSeq stranded mRNA kit (Illumina Corp.) and were sequenced using a NovaSeq6000 S4. Paired-end sequencing was performed at an average read depth of 51,404,587 ± 12,403,503 reads per sample with an average of 92.7% of bases reaching

quality scores of at least Q30. RNA quality control, library construction, and sequencing were performed by Macrogen Inc (now Psomagen) of Rockville, Maryland.

Prior to gene expression analysis, Fastqc (version 0.11.8) (Andrews 2010) was used to determine the sequence quality of the fastq files before and after they were trimmed with Trimmomatic (version 0.36) (Bolger et al. 2014). Sequencing reads were aligned to the mouse genome by Hisat2 (version 2.1.0) using the Ensembl GRCm38.p5 primary DNA assembly reference sequence (Kim et al. 2015; Cunningham et al. 2019). On average, the overall alignment rate was 98.1% +/- 0.3% [range: 96.9-98.6%]. Sequence alignment files were written into bam files, sorted, and indexed by Samtools (version 1.9.0) (Morgan et al. 2019) Genomic reads were counted by HTSeq version 0.11.1 (Anders et al. 2015). Genes with genomic read counts of 0 or a low yield were identified and removed from the analysis by EdgeR (version 3.10), followed by normalization for the remaining expressed genes (McCarthy et al. 2012).

QTL mapping

Trait data was analyzed from offspring of the CSS backcross (n=149) (Figure 1). All trait data for all samples can be found in Supplemental Table 2. Trait data was tested for missingness, and genotype data was tested for Hardy-Weinberg equilibrium and allele frequency in PLINK (Purcell et al. 2007). Summary statistics, Student t-tests, and Wilcoxon tests were performed in R (version 3.5.3). Gene expression was analyzed from a subset of the backcross offspring (n=49).

QTL linkage analysis was performed using R/qtl software (v.1.44-9) (Broman et al. 2003). One-dimensional and two-dimensional scans were conducted for both the trait data and the transcriptome data. The scans were performed with and without sex as a covariate using logistic regression with the HK algorithm (one-dimensional) or the EM algorithm (two-dimensional). Trait data was also performed with BMI as an additive covariate. LOD significance thresholds were obtained from 10,000 permutations; thus, accounting for concerns of multiple testing by using permutation p-values. QTL locations were updated in the context of a multiple QTL model based on maximum likelihood using the refineqtl function in R/qtl (Broman et al. 2019). Intra-chromosomal interacting QTLs within 30 MB of each other were removed as visual inspection of the data suggested a high frequency of suspected false positives, which is also consistent with the limited mapping resolution given the study's sample size. Significant interactions were visualized in three-dimensional space using the R packages "circularize" (Gu et al. 2014) and "rgl" (Adler 2019).

Data availability

All RNA sequencing files can be found at accession number GSE145607 at the NCBI gene expression omnibus

Results

Identification of additive QTLs for complex traits

B6.A6 and B6.A4 mice have previously been found to carry allelic variation on their respective substituted chromosomes that non-linearly interacted with each other to regulate plasma glucose levels (Chen et al. 2017). Additionally, mRNA expression of 83 genes were identified that were regulated by epistatic interactions between these two CSSs (Chen et al. 2017). To further map the genetic basis of these and other traits, a total of 17 metabolism- and blood-related traits were analyzed in N2 offspring from the modified backcross between B6.A6 and B6.A4 (Figure 1). These two CSSs were together backcrossed to B6 and trait data was measured and analyzed for 149 N2 offspring.

A single-QTL genome scan approach (R/qtl) was used to test for main (additive) effects in the 149 N2 mice (Broman et al. 2003). Unadjusted main effect QTL analysis identified significant QTLs for two traits: platelet count (plt) (LOD score > 2.04, p-value < 0.05) and granulocyte % (gran%) (LOD score > 2.02, p-value < 0.05). To test whether including sex as a covariate affected QTL detection, we tested for phenotypic differences in metabolism- and blood-related traits between male and female mice from the same crosses. There was a significant difference ($p < 0.05$) due to sex for 8 of the 17 factors tested: fasted body weight, fasting glucose, length, fat pad weight, mcv, mchc, plt, and rdwa (Table S3). Thus, sex was included as a covariate to better account for sex-specific trait differences. Sex-adjusted main effect QTLs were identified for three traits: plt and gran% as before, plus an additional QTL for fat pad weight (Table 1). The 1-LOD support QTL intervals for these traits were located between 38.9 - 104.8 Mb on Chr. 6 (fat pad weight), 83.2 - 128.8 Mb on Chr. 4 (plt), and 125.0 - 146.5 Mb on Chr. 6 (gran%) (Table 1).

Main effect QTL analyses were also performed with BMI in the model to account for variation due to adiposity. QTL analyses that included sex and BMI resulted in the same significant main effect QTLs as sex alone. However, there was an increase in significance for the plt QTL when BMI was added to the model (sex p value = 0.018 vs. sex and BMI p value=0.010), which aligns with the previously seen association between platelet count and BMI (Table 1) (Samocha-Bonet et al. 2008; Marginean et al. 2019).

Identification of epistatic QTLs for complex traits

To test for epistatic QTL interactions, a two-dimensional genome scan with sex in the model was performed for the metabolism- and blood-related trait data from 149 N2 mice. Epistatic QTLs were identified for two of the traits: lymphocytes (ly%) and red cell distribution width (rdw). An inter-chromosomal interaction was discovered between loci on chromosomes 4 and 6 that regulated ly%. An intra-chromosomal interaction between two distinct loci on chromosome 6 regulated rdw (Table 2). The 1-LOD support QTL intervals for ly% were between 5.09 – 104.51 Mb on Chr. 4 and between 38.85 – 93.25 Mb on Chr. 6. The 1-LOD support QTL intervals for rdw were between 3.49 – 16.06 Mb and 14.36 Mb - 17.72 Mb on Chr. 6 (Table 2).

For ly%, the most significant interacting SNPs within the QTL intervals were rs13477644 (Chr. 4) and rs13478739 (Chr. 6). Testing for effects of these SNP genotypes individually on ly% revealed no marginal effect on lymphocyte count (Figure 2A, B). Accordingly, a main effect QTL analysis of these regions individually found no evidence for a QTL at

either locus (Figure 2C, S4). However, when accounting for both SNP genotypes together, association with lymphocyte count became readily apparent (Figure 2D), with the resulting QTL analysis for interacting loci revealing a strong epistatic QTL (Figure 2E). The context dependent effects of this interaction are graphically illustrated in Figure 2F, as shown by the crossing lines connecting the mean values for each genotype combination indicating suppression epistasis. Remarkably, this epistatic interaction alone accounted for 12.1% of the variance in this trait.

Identification of QTLs that regulate hepatic gene expression

In addition to the metabolism and blood traits, we measured the liver transcriptomes in a subset (n=49) of the offspring from the B6.A4 and B6.A6 backcross to identify additive and epistatic QTLs regulating hepatic gene expression. Linkage analyses of the RNA-Seq data revealed 768 main effect QTLs (single locus). Among the main effect QTLs, there were 159 genes within the QTL interval on either chromosome 4 or 6, and are thus cis-QTLs, whereas 609 were located elsewhere in the genome and are therefore trans-QTLs. There were 519 epistatic QTLs regulating hepatic gene expression (LOD Score > 2.44, p-value < 0.05) (Tables S5 & S6), demonstrating the pervasive effects of epistasis in the regulation of gene expression.

The relationship between main effect QTL LOD scores and interaction QTL LOD scores shows that a large interaction QTL LOD score does not predict a large main effect QTL LOD score, and *vice versa* (Figure 3). Of the 748 unique genes with significant main effects and 510 unique genes with significant epistatic effects, 727 genes had only main

effects, 489 had only epistatic effects, respectively, and 21 genes had both significant main and epistatic effects. Thus, there appears to be little overlap between marginal and epistatic effector loci (Figure 3). The percent variance explained by interaction QTLs (mean 24.5%; range 20.5% - 85.6%) was similar to that explained by single locus QTLs (mean 28.5%; range 20.5% - 44.4%). Of 519 epistatic QTLs (the 510 unique genes and genes regulated by multiple eQTLs), 243 (46.8%) exhibited positive epistasis whereas 276 (53.2%) exhibited negative epistasis. The detection of so many epistatic QTLs accounting for such a high proportion of the variance was not unexpected, as our previous data suggested the widespread importance of epistasis in regulating gene expression (Chen et al. 2017).

The most statistically significant epistatic QTL regulating gene expression was for *Arhgap25*, with LOD peaks at the SNPs rs13478003 (Chr. 4) and rs13478976 (Chr. 6). As seen in the analysis of ly%, analysis of *Arhgap25* expression associated with each individual SNP failed to reveal an effect of genotype (Figure 4A, B), and neither resulted in a significant QTL association (Figure 4C). However, when *Arhgap25* expression was analyzed with SNP-SNP interactions, a clear genetic effect of the combination of rs13478003 and rs13478976 was detected (Figure 4D-F). This interaction accounted for a remarkable 41.7% of the variance for this trait, but was completely absent in a standard single locus association analysis (Table S6). Beyond *Arhgap25*, 518 additional genes were identified with significant epistatic QTLs. Interaction plots of the next 5 most significant regulated genes (after *Arhgap25*) are shown in Figure 5A-E and include *Irf2bpl*, *Gnmt*, *Rasa3*, *Flnb*, and *Capn10*, all demonstrating interactions that act to maintain

homeostatic levels of gene expression, rather than exacerbate strain differences in expression levels (Figure 5). None of the interacting loci that associated with the regulation of these transcripts was also marginally associated with controlling gene expression levels.

Interacting loci on chromosomes 4 and 6 were widely distributed along each chromosome (Figure 6A). Of the 186 intra-chromosomal interactions, 37 were located on chromosome 4 (red) and 149 on chromosome 6 (blue) (Figure 6B). The remaining 333 interactions were inter-chromosomal (gray) (Figure 6C). Of the genes whose expression was regulated by an epistatic QTL pair, 57 were located within the QTL interval on either chromosome 4 or 6, and are thus likely controlled by cis-QTLs. The remaining 462 interacting eQTLs were located elsewhere in the genome and are therefore trans-QTLs. This represents a roughly 2-fold enrichment of cis-QTLs relative to the size of the QTL intervals.

Discussion

Since CSSs have a greatly simplified genetic background, N2 CSS mice with allelic variation from two substituted chromosomes were used to study epistatic interactions for complex traits and gene expression. The resulting analyses identified widespread epistatic interactions. Among the traits where an epistatic QTL was detected, those QTLs accounted for a considerable proportion of the variance for that trait (10.2% for blood-related traits, 24.5% for gene expression). Marginal analyses of gene expression QTLs revealed a similar number of loci with comparable variance explained (interaction: mean

24.5%; range 20.5% - 85.6%, marginal: mean 28.5%; range 20.5% - 44.4%). This demonstrates that epistatic interactions can have large effect sizes. The high variance accounted for by each interacting QTL pair is likely attributable in part to the limited allelic variation within the cross as well as the limited power for detecting smaller effects. Nonetheless, these same factors apply equally to main effects and interaction effects, and the fact that the variance explained was similar between these two classes of QTLs contributes to the growing body of evidence that epistatic interactions play an important role in the genetic architecture of complex traits.

The analysis of N2 offspring derived from mammalian CSSs discovered specific interacting loci that contributed to the heritable variation in lymphocyte percentage (ly%), red cell distribution width (rdw), and hepatic gene expression of over 500 genes. That this number of traits and genes were identified with detectable regulation by epistasis is remarkable considering that only a single time point was examined, gene expression was only measured in the liver, and only one pairwise strain combination of CSSs was examined. The high prevalence of epistatic interactions provides evidence for complex molecular models underlying the genetics of many quantitative traits (Sackton and Hartl 2016; Chow 2016; Gerke et al. 2009; Gerke et al. 2010). The identification of epistatic loci that control lymphocyte number and red cell width may improve our ability to predict these traits based on genetic data or modulate these traits based on improved knowledge of the biological pathways that underlie these interactions. Both of these blood traits have clinical relevance. Specifically, knowing a more precise basis for the regulation of ly% has clinical relevance as it is associated with overall survival in cancer patients (Iseki et al.

2017; Zhao et al. 2017). This is seen explicitly in colorectal cancer where blood count levels, including ly%, is informative for cancer diagnosis and prognosis (Kostakis et al. 2018). Additionally, ly% can be used as a screening method for large granular lymphocyte leukemia to improve patient outcome (Tanahashi et al. 2017). In addition, a high red blood cell distribution width (rdw) predicts incident coronary artery disease (CAD) and is associated with heart failure, peripheral vascular disease, atrial fibrillation, stroke, and cancer (Pilling et al. 2018). Rdw is a key biomarker of atrial fibrillation (AF) that may be helpful in predicting the risk of developing adverse complications of AF (Lippi et al. 2019; Cha et al. 2017; Wan et al. 2015). Rdw is also a predictor of poor cancer survival, specifically rectal (Ide et al. 2019) and gallbladder cancer (Xie et al. 2019). Hence, identifying a more complete genetic model that incorporates interaction will improve our ability to predict those at greatest risk.

In addition to variation controlling lymphocyte number and red cell width, over 500 genes were identified that were regulated by epistasis. This is significantly more than were first identified in studies of the B6.A4 and B6.A6 CSS strains carrying the entire nonrecombinant A/J-derived substituted chromosomes (Chen et al. 2017). This may be, in part, because the recombination events within chromosomes 4 and 6 decoupled intra-chromosomal interactions in order to reveal previously undetected QTLs. This phenomenon has been frequently observed in previous studies of CSS for a wide range of traits, with numerous previously hidden QTLs revealed upon deconstruction of the substituted chromosome (Buchner and Nadeau 2015).

Among the most significant genes whose expression is regulated by epistatic interactions, *Capn10* is particularly interesting given its controversial association with T2D risk. It was the first T2D susceptibility gene identified in linkage studies (Horikawa et al. 2000), with multiple subsequent GWAS or candidate gene studies having also identified an association between variants near *CAPN10* and T2D risk and other diabetes-related phenotypes (Castro-Martinez et al. 2018; Bayramci et al. 2017; Hou et al. 2017; Zhao et al. 2016; Cui et al. 2016; Ibrahim et al. 2015; Orozco et al. 2014; Baier et al. 2000; Evans et al. 2001). However, many other studies have failed to replicate these findings (Zhang et al. 2019; Plengvidhya et al. 2015; Al-Sinani et al. 2015; Khan et al. 2014; Hegele et al. 2001; Tsai et al. 2001). It is tempting to speculate that one potential reason for the inconsistent results could be an incomplete understanding of epistatic interactions between *CAPN10* and other loci. To illustrate how this could happen, previous simulation studies of interacting loci revealed that small differences in allelic frequencies between populations can lead to a failure to replicate GWAS results (Greene et al. 2009). Therefore, it is of note that one study of ~1,400 T2D and control individuals found no evidence of main effects in *CAPN10*, but did identify evidence for gene-gene interactions involving *CAPN10* influencing T2D risk (Uma Jyothi and Reddy 2015). This is remarkably similar to what we have shown with ly% (Figure 3) and *Arhgap25* expression (Figure 4), where a failure to account for interactions can easily mask an otherwise strong association.

The CSS interaction-based mapping paradigm enabled the detection of otherwise hidden QTLs, revealing numerous novel loci that interact to control a number of complex traits

and the expression of hundreds of genes that were not found using only a marginal analysis strategy. These findings point to the importance of searching for epistasis in as simplified a genetic context as possible, knowing that even within the limited allelic variation of a single substituted chromosome there is a remarkable genetic complexity contributing to phenotypic variation. As we attempt to move towards more personalized medicine, it will no doubt require a more comprehensive understanding of the many interacting loci throughout the genome.

Figure Legends

Figure 1. **Diagram of the modified CSS backcross strategy to map epistatic QTLs.** (A) (B6.A4 x B6)F1 male mice were crossed with (B6 x B6.A6)F1 female mice to generate 149 “N2” offspring. (B) Seven hypothetical offspring are shown. Chromosomes 4 and 6 are derived from either A/J or C57BL/6J as indicated. All other chromosomes are B6-derived in all mice and therefore lack any allelic variation in this cross.

Figure 2. Epistatic interaction between loci on chromosomes 4 and 6 regulate the percentage of lymphocyte cells in blood. Lymphocyte % based on genotype at SNP markers (A) rs13477644 on chromosome 4 and (B) rs13478739 on chromosome 6. Each dot represents one mouse. The blue and red lines indicate the mean and standard error. (C) QTL mapping results for main effects on chromosomes 4 and 6. The LOD threshold for significance (5%) was calculated using R/qtl and is indicated by a horizontal line. The small lines under the LOD peaks indicate the location of SNP genotyping markers. No significant QTLs were detected. (D) Lymphocyte % based on the combined genotypes at rs13477644 on chromosome 4 and rs13478739 on chromosome 6. (E) QTL mapping results for interaction effects on chromosomes 4 and 6. The LOD threshold for significance (5%) was calculated using R/qtl and is indicated by a horizontal line. The most significant interaction QTLs were detected with peaks at rs13477644 and rs13478739, with a potentially second peak on chromosome 4 at rs13477796. (F) Context-dependent effects of the AA and AB genotypes at rs13477644 and rs13478739.

Mean and standard error are shown for each genotype combination. An “A” genotype indicates A/J-derived. A “B” genotype indicates C57BL/6J-derived.

Figure 3. Little overlap between main effect QTLs and epistatic QTL pairs. (A,B) LOD scores for peak SNPs defining significant epistatic QTL pairs were plotted against the corresponding main effect LOD scores for each SNP within the epistatic pair. (C) LOD scores for peak SNPs defining significant main effect QTLs were plotted against the most significant interaction LOD score for that SNP. The solid red line indicates the level of significance for epistatic QTLs and the dashed red line indicates the level of significance for main effect QTLs.

Figure 4. Epistatic interaction between loci on chromosomes 4 and 6 regulates *Arhgap25* mRNA expression. *Arhgap25* mRNA expression based on genotype at SNP markers (A) rs13478003 on chromosome 4 and (B) rs13478976 on chromosome 6. Each dot represents one mouse. The blue and red lines indicate the mean and standard error. (C) QTL mapping results for main effects on chromosomes 4 and 6. The LOD threshold for significance (5%) was calculated using R/qtl [R/qtl or R/qtl2?] and is indicated by a horizontal line. The small lines under the LOD peaks indicate the location of SNP genotyping markers. No significant QTLs were detected. (D) *Arhgap25* mRNA expression based on the combined genotypes at rs13478003 on chromosome 4 and rs13478976 on chromosome 6. (E) QTL mapping results for interaction effects on chromosomes 4 and 6. The LOD threshold for significance (5%) was calculated by R/qtl [R/qtl or R/qtl2?] using 10,000 permutations and is indicated by a horizontal line. The most significant interaction

QTLs were detected with peaks at rs13478003 and rs13478976 (F) Context- dependent effects of the AA and AB genotypes at rs13478003 and rs13478976. Mean and standard error are shown for each genotype combination. An “A” genotype indicates A/J-derived. A “B” genotype indicates C57BL/6J-derived.

Figure 5. **Epistatic interactions control gene expression.** Context dependent effects on gene expression for (A) *Irf2bpl*, (B) *Gnmt*, (C) *Rasa3*, (D) *Flnb*, and (E) *Capn10*. The y-axis represents the number of unique sequencing reads per gene after normalizing read depth across samples. Mean and standard error are shown for each genotype combination. An “A” genotype indicates A/J-derived. A “B” genotype indicates C57BL/6J-derived.

Figure 6. Epistatic interaction of gene expression across the genome. Numbering around the circle plots indicates the Mb position on each chromosome. The circos plots illustrate the peak SNP-SNP interacting loci for all gene expression epistatic interactions on and between chromosomes 4 and 6. (A) There are 519 total epistatic interactions, (B) 186 intra-chromosomal interactions including on 149 on chromosome 4 (blue) and 37 on chromosome 6 (red), (C) and 333 inter-chromosomal interactions (gray).

References

Adler, D.M., D., 2019 rgl: 3D Visualization Using OpenGL. R package version 0.100.30. https://CRAN.R-project.org/package=rgl.

Al-Sinani, S., N. Woodhouse, A. Al-Mamari, O. Al-Shafie, M. Al-Shafaee *et al.*, 2015 Association of gene variants with susceptibility to type 2 diabetes among Omanis. *World J Diabetes* 6 (2):358-366.

Anders, S., P.T. Pyl, and W. Huber, 2015 HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31 (2):166-169.

Andrews, S., 2010 FastQC: a quality control tool for high throughput sequence data. *Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc*.

Baier, L.J., P.A. Permana, X. Yang, R.E. Pratley, R.L. Hanson *et al.*, 2000 A calpain-10 gene polymorphism is associated with reduced muscle mRNA levels and insulin resistance. *J Clin Invest* 106 (7):R69-73.

Bayramci, N.S., L. Acik, C. Kalkan, and I. Yetkin, 2017 Investigation of glucocorticoid receptor and calpain-10 gene polymorphisms in Turkish patients with type 2 diabetes mellitus. *Turk J Med Sci* 47 (5):1568-1575.

Bloom, J.S., I.M. Ehrenreich, W.T. Loo, T.L. Lite, and L. Kruglyak, 2013 Finding the sources of missing heritability in a yeast cross. *Nature* 494 (7436):234-237.

Bolger, A.M., M. Lohse, and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15):2114-2120.

Boyle, E.A., Y.I. Li, and J.K. Pritchard, 2017 An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169 (7):1177-1186.

Broman, K.W., D.M. Gatti, P. Simecek, N.A. Furlotte, P. Prins *et al.*, 2019 R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multiparent Populations. *Genetics* 211 (2):495-502.

Broman, K.W., H. Wu, S. Sen, and G.A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19 (7):889-890.

Buchner, D.A., and J.H. Nadeau, 2015 Contrasting genetic architectures in different mouse reference populations used for studying complex traits. *Genome Res* 25 (6):775-791.

Carlborg, O., and C.S. Haley, 2004 Epistasis: too often neglected in complex trait studies? *Nat Rev Genet* 5 (8):618-625.

Castro-Martinez, A.G., J. Sanchez-Corona, A.P. Vazquez-Vargas, A.G. Garcia-Zapien, A. Lopez-Quintero *et al.*, 2018 Association analysis of calpain 10 gene variants/haplotypes with gestational diabetes mellitus among Mexican women. *Cell Mol Biol (Noisy-le-grand)* 64 (3):81-86.

Cha, M.J., H.S. Lee, H.M. Kim, J.H. Jung, E.K. Choi *et al.*, 2017 Association between red cell distribution width and thromboembolic events in patients with atrial fibrillation. *Eur J Intern Med* 46:41-46.

Chatterjee, S., and A. Chakravarti, 2019 A gene regulatory network explains RET-EDNRB epistasis in Hirschsprung disease. *Hum Mol Genet* 28 (18):3137-3147.

Chen, A., Y. Liu, S.M. Williams, N. Morris, and D.A. Buchner, 2017 Widespread epistasis regulates glucose homeostasis and gene expression. *PLoS Genet* 13 (9):e1007025.

Chow, C.Y., 2016 Bringing genetic background into focus. *Nat Rev Genet* 17 (2):63-64.

Churchill, G.A., D.M. Gatti, S.C. Munger, and K.L. Svenson, 2012 The Diversity Outbred mouse population. *Mamm Genome* 23 (9-10):713-718.

Cole, B.S., M.A. Hall, R.J. Urbanowicz, D. Gilbert-Diamond, and J.H. Moore, 2017 Analysis of Gene-Gene Interactions. *Curr Protoc Hum Genet* 95:1 14 11-11 14 10.

Cordell, H.J., 2002 Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet* 11 (20):2463-2468.

Cui, J., X. Xu, S. Yin, F. Chen, P. Li *et al.*, 2016 Meta-analysis of the association between four CAPN10 gene variants and gestational diabetes mellitus. *Arch Gynecol Obstet* 294 (3):447-453.

Cunningham, F., P. Achuthan, W. Akanni, J. Allen, M.R. Amode *et al.*, 2019 Ensembl 2019. *Nucleic Acids Res* 47 (D1):D745-D751.

Evans, J.C., T.M. Frayling, P.G. Cassell, P.J. Saker, G.A. Hitman *et al.*, 2001 Studies of association between the gene for calpain-10 and type 2 diabetes mellitus in the United Kingdom. *Am J Hum Genet* 69 (3):544-552.

Forsberg, S.K., J.S. Bloom, M.J. Sadhu, L. Kruglyak, and O. Carlborg, 2017 Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. *Nat Genet* 49 (4):497-503.

Fu, W., T.D. O'Connor, and J.M. Akey, 2013 Genetic architecture of quantitative traits and complex diseases. *Curr Opin Genet Dev* 23 (6):678-683.

Galarza-Munoz, G., F.B.S. Briggs, I. Evsyukova, G. Schott-Lerner, E.M. Kennedy *et al.*, 2017 Human Epistatic Interaction Controls IL7R Splicing and Increases Multiple Sclerosis Risk. *Cell* 169 (1):72-84 e13.

Gerke, J., K. Lorenz, and B. Cohen, 2009 Genetic interactions between transcription factors cause natural variation in yeast. *Science* 323 (5913):498-501.

Gerke, J., K. Lorenz, S. Ramnarine, and B. Cohen, 2010 Gene-environment interactions at nucleotide resolution. *PLoS Genet* 6 (9):e1001144.

Greene, C.S., N.M. Penrod, S.M. Williams, and J.H. Moore, 2009 Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS One* 4 (6):e5639.

Gu, Z., L. Gu, R. Eils, M. Schlesner, and B. Brors, 2014 circlize Implements and enhances circular visualization in R. *Bioinformatics* 30 (19):2811-2812.

Hegele, R.A., S.B. Harris, B. Zinman, A.J. Hanley, and H. Cao, 2001 Absence of association of type 2 diabetes with CAPN10 and PC-1 polymorphisms in Oji-Cree. *Diabetes Care* 24 (8):1498-1499.

Hemani, G., K. Shakhbazov, H.J. Westra, T. Esko, A.K. Henders *et al.*, 2014 Detection and replication of epistasis influencing transcription in humans. *Nature* 508 (7495):249-253.

Horikawa, Y., N. Oda, N.J. Cox, X. Li, M. Orho-Melander *et al.*, 2000 Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 26 (2):163-175.

Hou, Z., M. Li, and Y. Cao, 2017 TCF7L2, CAPN10 polymorphisms are associated with gestational diabetes mellitus (GDM) risks: a meta-analysis. *Gynecol Endocrinol* 33 (5):399-404.

Hu, T., Y. Chen, J.W. Kiralis, R.L. Collins, C. Wejse *et al.*, 2013 An information-gain approach to detecting three-way epistatic interactions in genetic association studies. *J Am Med Inform Assoc* 20 (4):630-636.

Huang, W., and T.F. Mackay, 2016 The Genetic Architecture of Quantitative Traits Cannot Be Inferred from Variance Component Analysis. *PLoS Genet* 12 (11):e1006421.

Huang, W., S. Richards, M.A. Carbone, D. Zhu, R.R. Anholt *et al.*, 2012 Epistasis dominates the genetic architecture of Drosophila quantitative traits. *Proc Natl Acad Sci U S A* 109 (39):15553-15559.

Huang, Y., C. Wang, Y. Yao, X. Zuo, S. Chen *et al.*, 2015 Molecular Basis of Gene-Gene Interaction: Cyclic Cross-Regulation of Gene Expression and Post-GWAS Gene-Gene Interaction Involved in Atrial Fibrillation. *PLoS Genet* 11 (8):e1005393.

Ibrahim, A.T., A. Hussain, M.A. Salih, O.A. Ibrahim, S.E. Jamieson *et al.*, 2015 Candidate gene analysis supports a role for polymorphisms at TCF7L2 as risk factors for type 2 diabetes in Sudan. *J Diabetes Metab Disord* 15:4.

Ide, S., Y. Toiyama, Y. Okugawa, Y. Omura, T. Kitajima *et al.*, 2019 Clinical significance of an increased red blood cell distribution width in patients with rectal cancer undergoing chemoradiotherapy followed by surgery. *Surg Today*.

Iseki, Y., M. Shibutani, K. Maeda, H. Nagahara, T. Tamura *et al.*, 2017 The impact of the preoperative peripheral lymphocyte count and lymphocyte percentage in patients with colorectal cancer. *Surg Today* 47 (6):743-754.

Jasnos, L., and R. Korona, 2007 Epistatic buffering of fitness loss in yeast double deletion strains. *Nat Genet* 39 (4):550-554.

Kayima, J., J. Liang, Y. Natanzon, J. Nankabirwa, I. Ssinabulya *et al.*, 2017 Association of genetic variation with blood pressure traits among East Africans. *Clin Genet* 92 (5):487-494.

Khan, I.A., S. Movva, N.A. Shaik, S. Chava, P. Jahan *et al.*, 2014 Investigation of Calpain 10 (rs2975760) gene polymorphism in Asian Indians with Gestational Diabetes Mellitus. *Meta Gene* 2:299-306.

Kim, D., B. Langmead, and S.L. Salzberg, 2015 HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12 (4):357-360.

Kirino, Y., G. Bertsias, Y. Ishigatsubo, N. Mizuki, I. Tugal-Tutkun *et al.*, 2013 Genome-wide association analysis identifies new susceptibility loci for Behcet's disease and epistasis between HLA-B*51 and ERAP1. *Nat Genet* 45 (2):202-207.

Kostakis, I.D., A.G. Vaiopoulos, Z. Garoufalia, A.G. Papavassiliou, S. Kykalos *et al.*, 2018 What can preoperative blood tests tell us about colorectal cancer? *J BUON* 23 (7):84-95.

Lehner, B., C. Crombie, J. Tischler, A. Fortunato, and A.G. Fraser, 2006 Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. *Nat Genet* 38 (8):896-903.

Lippi, G., G. Cervellin, and F. Sanchis-Gomar, 2019 Red blood cell distribution width: A marker of anisocytosis potentially associated with atrial fibrillation. *World J Cardiol* 11 (12):292-304.

Liu, Y., H. Xu, S. Chen, X. Chen, Z. Zhang *et al.*, 2011 Genome-wide interaction-based association analysis identified multiple new susceptibility Loci for common diseases. *PLoS Genet* 7 (3):e1001338.

Mackay, T.F., 2014 Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet* 15 (1):22-33.

Manolio, T.A., F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff *et al.*, 2009 Finding the missing heritability of complex diseases. *Nature* 461 (7265):747-753.

Marginean, C.O., L.E. Melit, D.V. Ghiga, and M.O. Marginean, 2019 Early Inflammatory Status Related to Pediatric Obesity. *Front Pediatr* 7:241.

McCarthy, D.J., Y. Chen, and G.K. Smyth, 2012 Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40 (10):4288-4297.

Moore, J.H., and S.M. Williams, 2009 Epistasis and its implications for personal genetics. *Am J Hum Genet* 85 (3):309-320.

Morgan, M., H. Pagès, V. Obenchain, and N. Hayden, 2019 Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. *R package version 2.0.1, http://bioconductor.org/packages/Rsamtools.*

Nadeau, J.H., J.B. Singer, A. Matin, and E.S. Lander, 2000 Analysing complex genetic traits with chromosome substitution strains. *Nat Genet* 24 (3):221-225.

Orozco, A.C., A.M. Munoz, C.M. Velasquez, R.M. Uscategui, M.V. Parra *et al.*, 2014 Variant in CAPN10 gene and environmental factors show evidence of association with excess weight among young people in a Colombian population. *Biomedica* 34 (4):546-555.

Pilling, L.C., J.L. Atkins, G.A. Kuchel, L. Ferrucci, and D. Melzer, 2018 Red cell distribution width and common disease onsets in 240,477 healthy volunteers followed for up to 9 years. *PLoS One* 13 (9):e0203504.

Plengvidhya, N., K. Chanprasert, W. Tangjittipokin, W. Thongnoppakhun, and P.T. Yenchitsomanus, 2015 Detection of CAPN10 copy number variation in Thai patients with type 2 diabetes by denaturing high performance liquid chromatography and real-time quantitative polymerase chain reaction. *J Diabetes Investig* 6 (6):632-639.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira *et al.*, 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81 (3):559-575.

Sackton, T.B., and D.L. Hartl, 2016 Genotypic Context and Epistasis in Individuals and Populations. *Cell* 166 (2):279-287.

Samocha-Bonet, D., D. Justo, O. Rogowski, N. Saar, S. Abu-Abeid *et al.*, 2008 Platelet counts and platelet activation markers in obese subjects. *Mediators Inflamm* 2008:834153.

Sawcer, S., R.J. Franklin, and M. Ban, 2014 Multiple sclerosis genetics. *Lancet Neurol* 13 (7):700-709.

Shao, H., L.C. Burrage, D.S. Sinasac, A.E. Hill, S.R. Ernest *et al.*, 2008 Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. *Proc Natl Acad Sci U S A* 105 (50):19910-19914.

Soong, T.R., B.E. Howitt, A. Miron, N.S. Horowitz, F. Campbell *et al.*, 2018 Evidence for lineage continuity between early serous proliferations (ESPs) in the Fallopian tube and disseminated high-grade serous carcinomas. *J Pathol* 246 (3):344-351.

Tanahashi, T., N. Sekiguchi, K. Matsuda, A. Matsumoto, T. Ito *et al.*, 2017 A screening method with lymphocyte percentage and proportion of granular lymphocytes in the peripheral blood for large granular lymphocyte (LGL) leukemia. *Int J Hematol* 105 (1):87-91.

Tsai, H.J., G. Sun, D.E. Weeks, R. Kaushal, M. Wolujewicz *et al.*, 2001 Type 2 diabetes and three calpain-10 gene polymorphisms in Samoans: no evidence of association. *Am J Hum Genet* 69 (6):1236-1244.

Tyler, A., J.M. Mahoney, and G.W. Carter, 2020 Genetic Interactions Affect Lung Function in Patients with Systemic Sclerosis. *G3 (Bethesda)* 10 (1):151-163.

Tyler, A.L., B. Ji, D.M. Gatti, S.C. Munger, G.A. Churchill *et al.*, 2017 Epistatic Networks Jointly Influence Phenotypes Related to Metabolic Disease and Gene Expression in Diversity Outbred Mice. *Genetics* 206 (2):621-639.

Uma Jyothi, K., and B.M. Reddy, 2015 Gene-gene and gene-environment interactions in the etiology of type 2 diabetes mellitus in the population of Hyderabad, India. *Meta Gene* 5:9-20.

Verma, S.S., J.N. Cooke Bailey, A. Lucas, Y. Bradford, J.G. Linneman *et al.*, 2016 Epistatic Gene-Based Interaction Analyses for Glaucoma in eMERGE and NEIGHBOR Consortium. *PLoS Genet* 12 (9):e1006186.

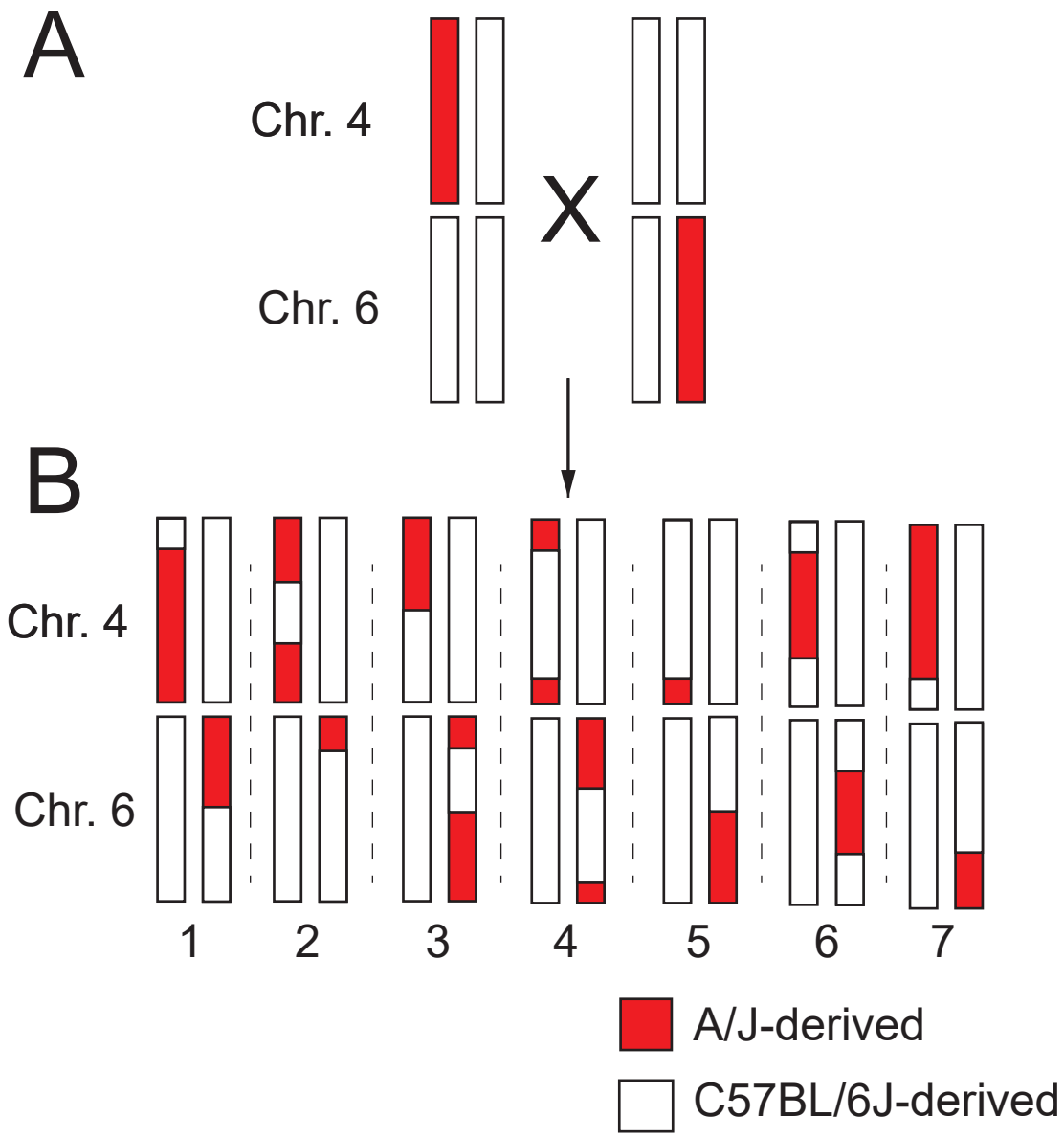
Wan, H., Y. Yang, J. Zhu, B. Huang, J. Wang *et al.*, 2015 The relationship between elevated red cell distribution width and long-term outcomes among patients with atrial fibrillation. *Clin Biochem* 48 (12):762-767.

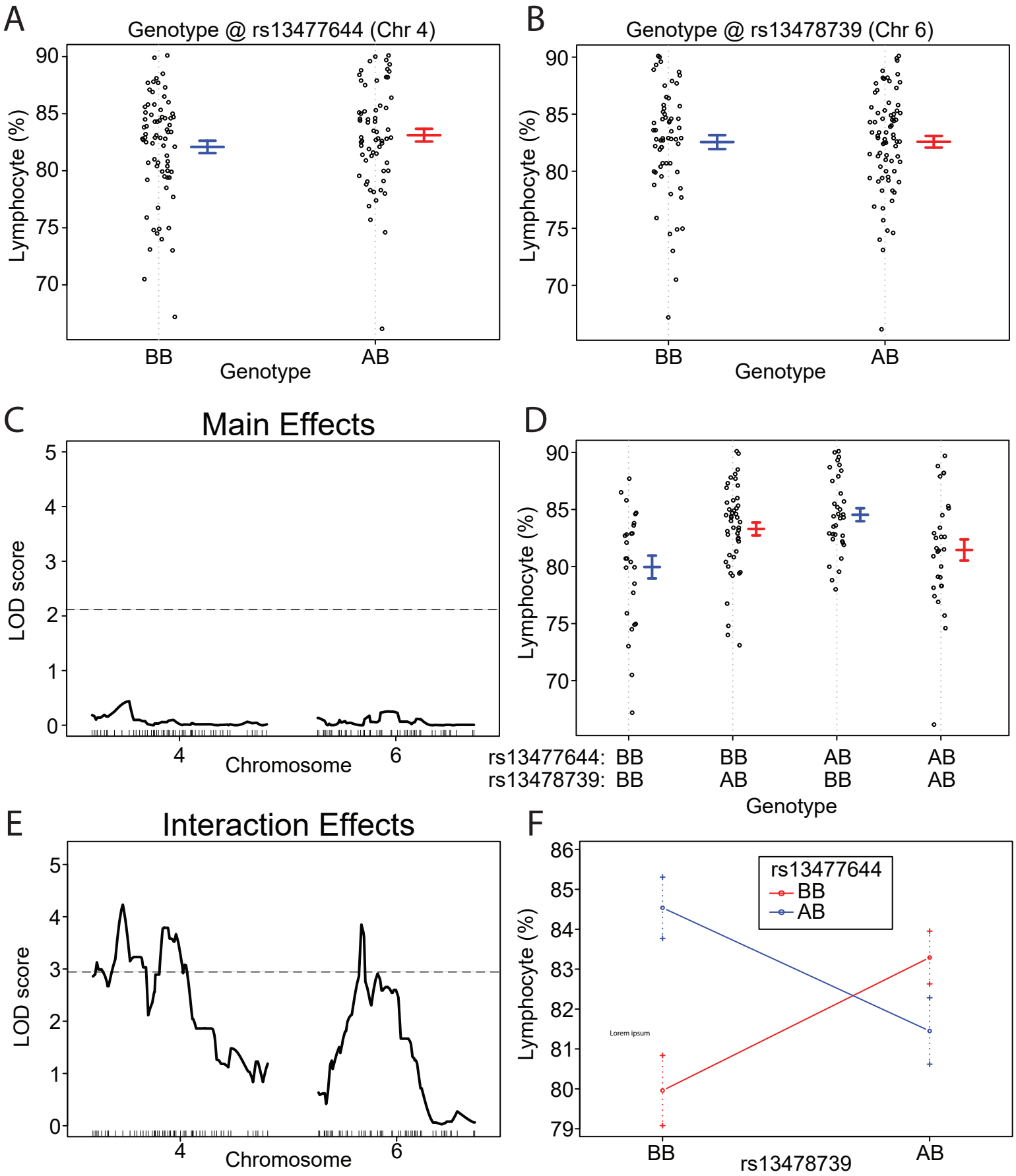
Xie, Y., L. Zhang, and L. Zhan, 2019 Can we assess the advancements of gallbladder cancer using red blood cell distribution width? *Medicine (Baltimore)* 98 (51):e18364.

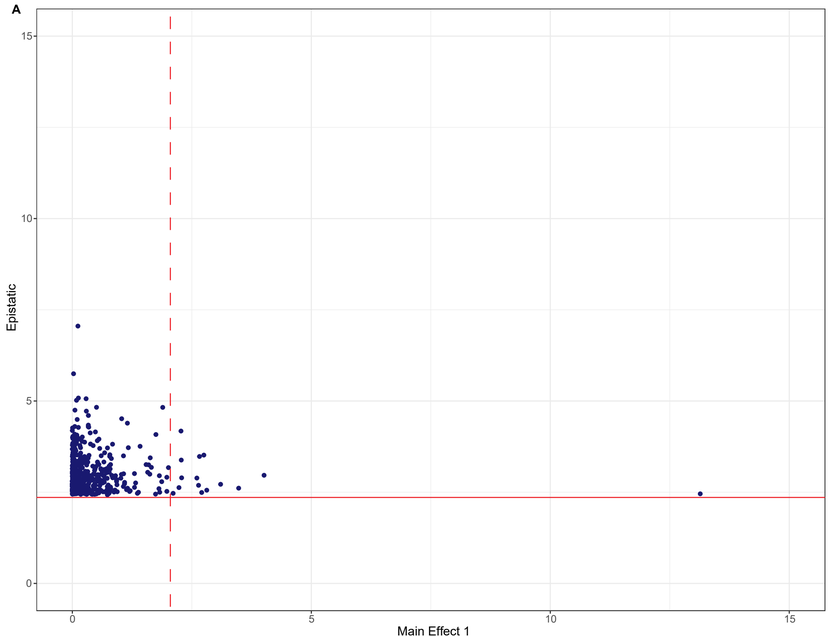
Zhang, X., C. Shi, L. Wei, F. Sun, and L. Ji, 2019 The Association Between the rs2975760 and rs3792267 Single Nucleotide Polymorphisms of Calpain 10 (CAPN10) and Gestational Diabetes Mellitus. *Med Sci Monit* 25:5137-5142.

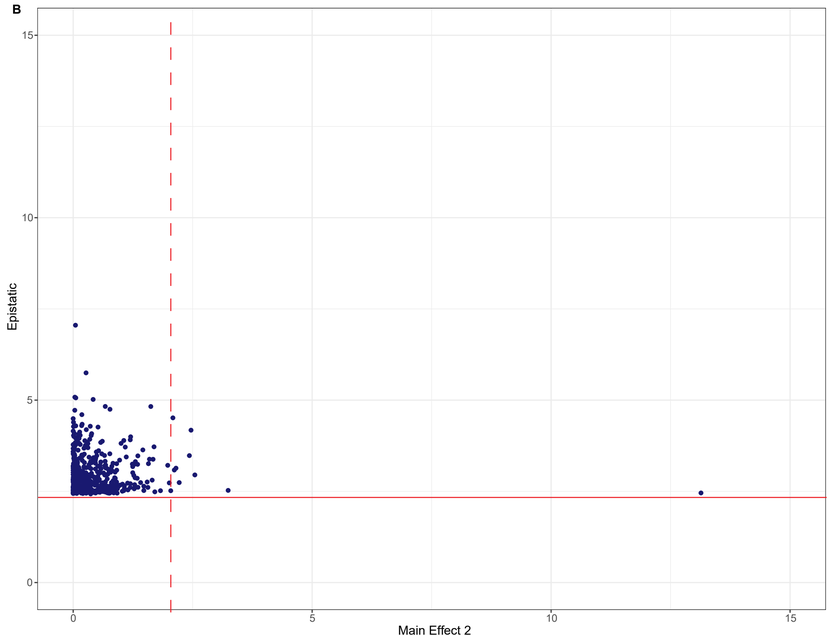
Zhao, F., D. Mamatyusupu, Y. Wang, H. Fang, H. Wang *et al.*, 2016 The Uyghur population and genetic susceptibility to type 2 diabetes: potential role for variants in CAPN10, APM1 and FUT6 genes. *J Cell Mol Med* 20 (11):2138-2147.

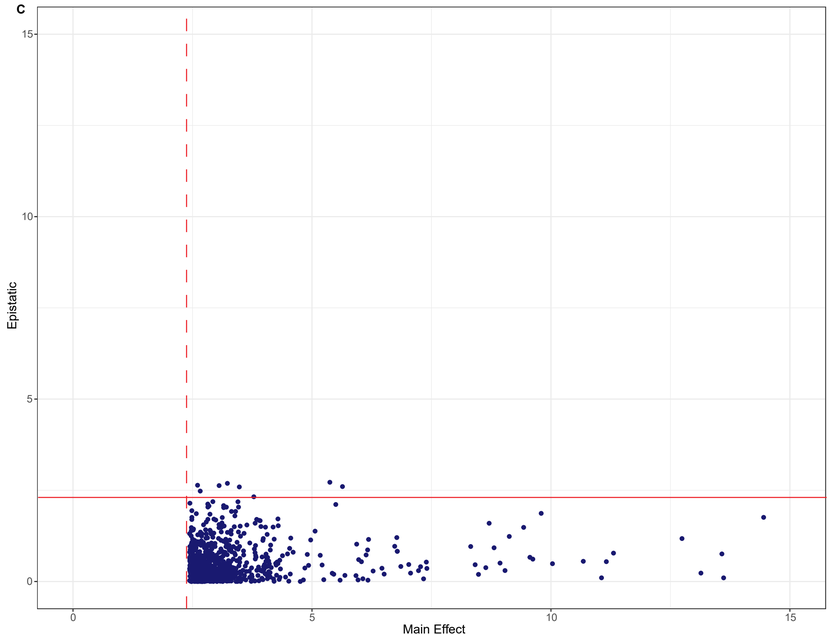
Zhao, W., P. Wang, H. Jia, M. Chen, X. Gu *et al.*, 2017 Lymphocyte count or percentage: which can better predict the prognosis of advanced cancer patients following palliative care? *BMC Cancer* 17 (1):514.

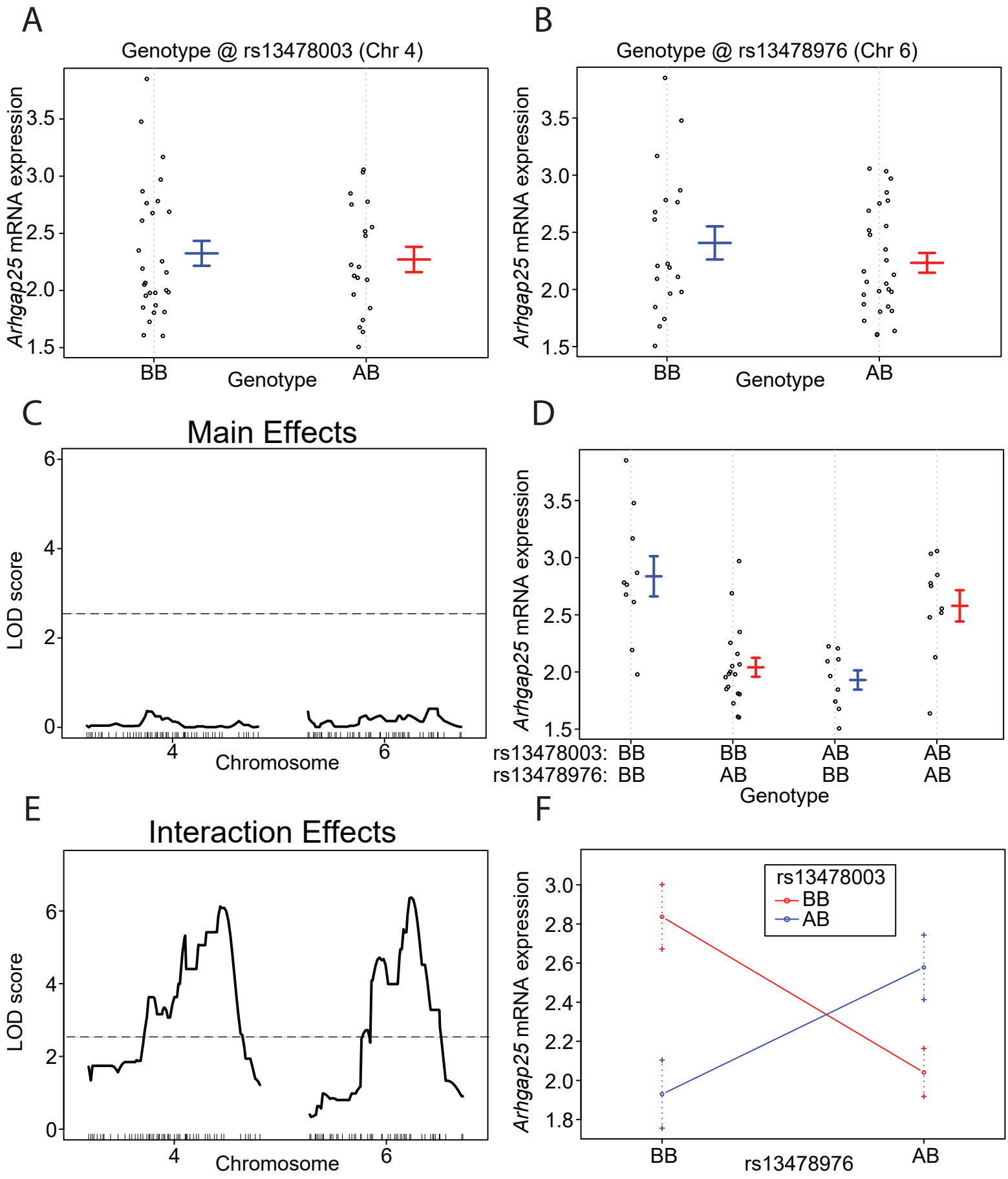

A
Chr. 4
Chr. 6
X
B
Chr. 4
Chr. 6
1
2
3
4
5
6
7
A/J-derived
C57BL/6J-derived

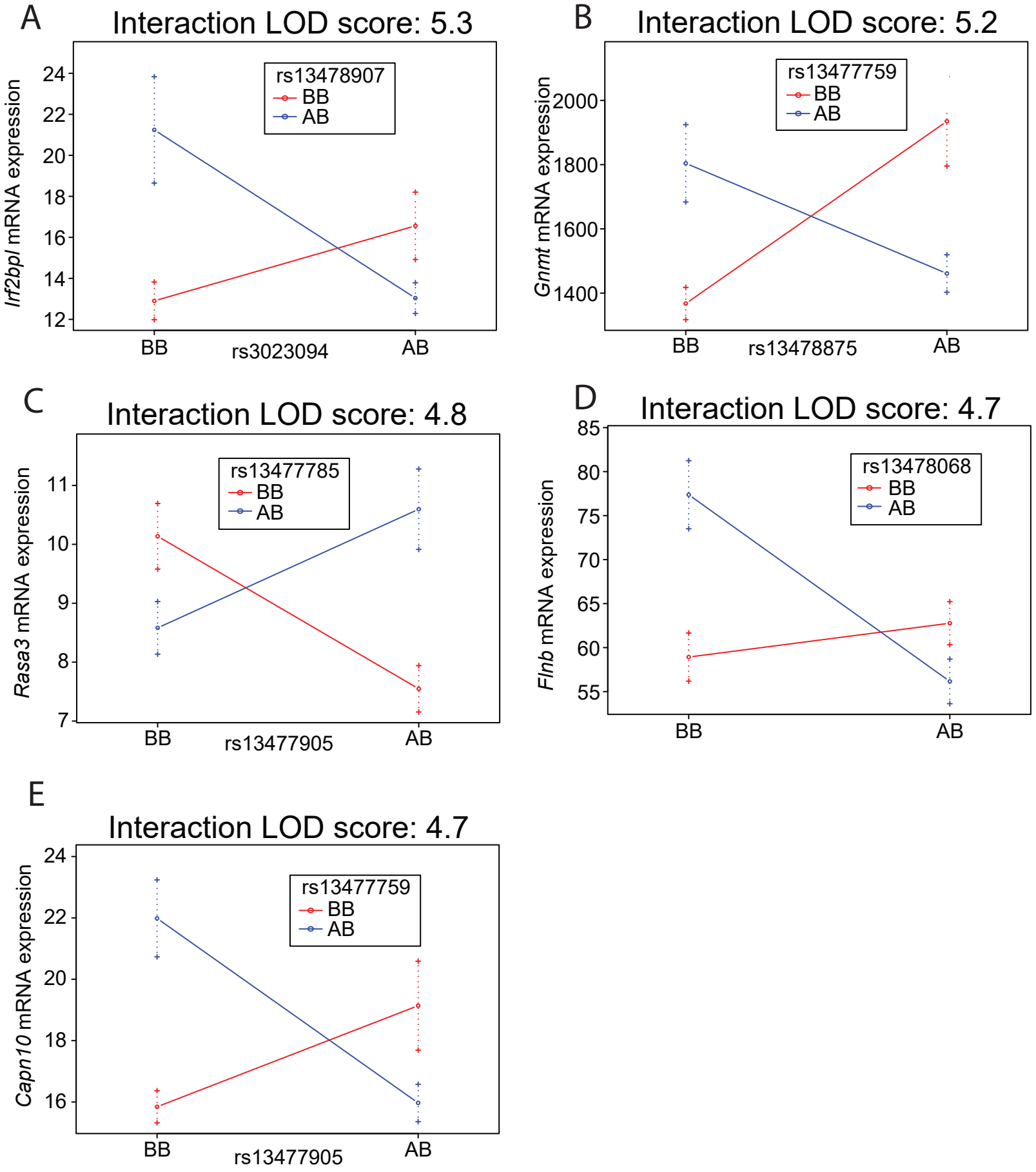

A
Genotype @ rs13477644 (Chr 4)
Lymphocyte (%)
90
85
80
75
70
BB
AB
Genotype
B
Genotype @ rs13478739 (Chr 6)
Lymphocyte (%)
90
85
80
75
70
BB
AB
Genotype
C
Main Effects
LOD score
5
4
3
2
1
0
4
6
Chromosome
D
Lymphocyte (%)
90
85
80
75
70
rs13477644: BB BB AB AB
rs13478739: BB AB BB AB
Genotype
E
Interaction Effects
LOD score
5
4
3
2
1
0
4
6
Chromosome
F
Lymphocyte (%)
86
85
84
83
82
81
80
79
rs13477644
BB
AB
Lorem ipsum
BB
AB
rs13478739


A
Epistatic
Main Effect 1
0
5
10
15


B
Epistatic
Main Effect 2
0
5
10
15


C
Epistatic
Main Effect
0
5
10
15


A
Genotype @ rs13478003 (Chr 4)
Arhgap25 mRNA expression
3.5
3.0
2.5
2.0
1.5
BB
Genotype
AB
B
Genotype @ rs13478976 (Chr 6)
Arhgap25 mRNA expression
3.5
3.0
2.5
2.0
1.5
BB
Genotype
AB
C
Main Effects
LOD score
6
4
2
0
4
Chromosome
6
D
Arhgap25 mRNA expression
3.5
3.0
2.5
2.0
1.5
rs13478003: BB BB AB AB
rs13478976: BB AB BB AB
Genotype
E
Interaction Effects
LOD score
6
4
2
0
4
Chromosome
6
F
rs13478003
BB
AB
Arhgap25 mRNA expression
3.0
2.8
2.6
2.4
2.2
2.0
1.8
BB
rs13478976
AB


A
Interaction LOD score: 5.3
rs13478907
BB
AB
Irf2bpl mRNA expression
24
22
20
18
16
14
12
BB
rs3023094
AB
B
Interaction LOD score: 5.2
rs13477759
BB
AB
Gnmt mRNA expression
2000
1800
1600
1400
BB
rs13478875
AB
C
Interaction LOD score: 4.8
rs13477785
BB
AB
Rasa3 mRNA expression
11
10
9
8
7
BB
rs13477905
AB
D
Interaction LOD score: 4.7
rs13478068
BB
AB
Flnb mRNA expression
85
80
75
70
65
60
55
BB
AB
E
Interaction LOD score: 4.7
rs13477759
BB
AB
Capn10 mRNA expression
24
22
20
18
16
BB
rs13477905
AB

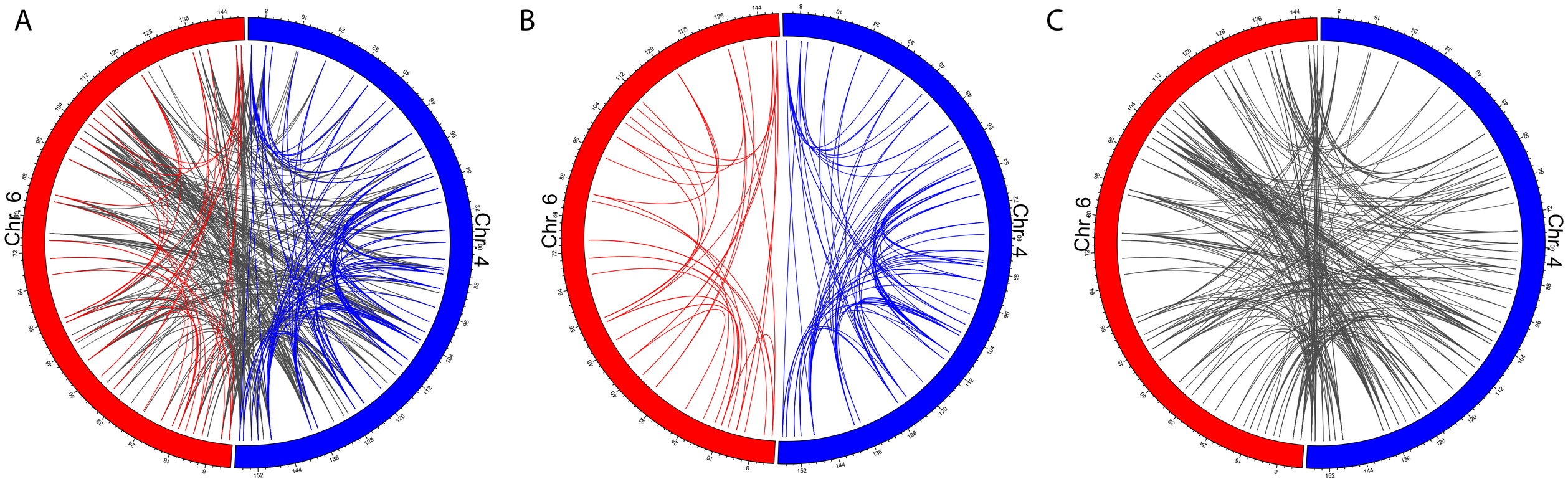

A
Chr 6
Chr 4
B
Chr 6
Chr 4
C
Chr 6
Chr 4

TABLE 1. 1-LOD SUPPORT INTERVALS FOR MAIN EFFECT TRAIT QTLS

	Chromosome	Marker	Position (Mb)	LOD	p-value	BMI LOD	BMI p-value
Fat Pad Weight	6	rs13478720	38.85	1.17		1.19	
	6	**rs13478811**	**70.31**	**2.18**	**0.030**	**2.24**	**0.030**
	6	rs13478947	104.84	1.06		1.08	
Platelets	4	rs13477808	83.18	1.56		1.55	
	4	**rs13477911**	**111.29**	**2.61**	**0.018**	**2.64**	**0.010**
	4	rs13477973	128.75	1.41		1.36	
Granulocyte (%)	6	rs3023094	125.04	1.61		1.49	
	6	**rs13479068**	**137.48**	**2.64**	**0.006**	**2.57**	**0.02**
	6	rs3090690	146.46	1.55		1.55	

*Peak SNPs are in bold with the SNPs defining the 1-LOD support interval shown above and below the peak SNP. The LOD and P-value columns include sex in the model, The BMI LOD and BMI p-value columns include sex and BMI in the model

TABLE 2. 1-LOD SUPPORT INTERVALS FOR EPISTATIC TRAIT QTLS

	Chromosome	Marker (nearest SNP)	Position (Mb)	LOD	p-value
Lymphocytes (%)					**0.032**
Interval 1					
	4	rs13477536	5.09	2.86	
	4	**rs13477644**	**35.28**	**4.23**	
	4	rs3695339	104.51	2.72	
Interval 2					
	6	rs13478720	38.85	2.29	
	6	**rs13478739**	**47.03**	**3.85**	
	6	rs13478897	93.25	1.95	
Red Cell Distribution Width					**0.036**
Interval 1					
	6	rs13459096	3.49	3.10	
	6	**rs13478633**	**12.25**	**5.69**	
	6	rs13478641	16.06	0.90	
Interval 2					
	6	rs13478638	14.36	3.70	
	6	**rs13478641**	**16.06**	**5.95**	
	6	rs3023064	17.72	3.91	

*Peak SNPs are in bold with the SNPs defining the 1-LOD support interval shown above and below both peak SNPs

TABLE S3. TRAIT DATA SUMMARY STATISTICS

Trait	Female (mean(sd))	Male (mean(sd))	T-Test	Wilcoxon Test
Fasted Weight (g)	14.09 (0.99)	16.94 (1.73)	<2.2E-16	<2.2E-16
Fasting Glucose (mg/dL)	83.58 (20.39)	98.79 (24.30)	6.26E-05	6.19E-06
Length (cm)	88.80 (1.95)	92.28 (2.73)	1.28E-15	2.04E-14
Fat Pad Weight (g)	0.043 (0.02)	0.078 (0.04)	1.54E-11	1.88E-10
White Blood Count (K/uL)	6.31 (2.20)	5.95 (2.33)	3.60E-01	3.03E-01
Lymphocytes (K/uL)	5.15 (1.62)	4.90 (1.93)	4.14E-01	2.30E-01
Monocytes (K/uL)	0.31 (0.14)	0.29 (0.14)	4.86E-01	4.89E-01
Lymphocytes (%)	83.03 (4.47)	82.28 (4.14)	3.20E-01	3.36E-01
Basophils (%)	0.12 (0.13)	0.17 (0.18)	5.42E-01	5.00E-01
Red Blood Count (M/uL)	9.32 (1.00)	9.28 (0.97)	7.88E-01	7.54E-01
Hemoglobin (g/dL)	14.79 (1.58)	14.56 (1.73)	4.30E-01	4.83E-01
Hematocrit (%)	43.63 (4.08)	44.34 (4.29)	3.31E-01	9.99E-02
Mean Corpus Volume (fL)	46.88 (1.68)	47.84 (1.69)	1.28E-03	1.81E-04
Mean Corpuscular Hemoglobin (pg)	15.85 (0.36)	15.68 (0.88)	1.59E-01	8.23E-01
Mean Corpuscular Hemoglobin Concentration (g/dL)	33.87 (1.30)	32.86 (2.46)	3.44E-03	1.73E-05
Red Cell Distribution Width (%)	19.11 (1.00)	19.24 (1.09)	4.88E-01	3.14E-01
Platelets (K/uL)	562.90 (125.60)	656.14 (152.13)	1.73E-04	6.25E-06
Mean Platelet Volume (fL)	5.19 (0.39)	5.16 (0.44)	7.06E-01	5.19E-01

Granulocyte (K/uL)	0.75 (0.38)	0.75 (0.37)	9.95E-01	7.48E-01
Granulocyte (%)	12.6 (4.06)	13.17 (3.88)	4.11E-01	1.47E-01
RDWa	33.62 (1.00)	34.77 (1.41)	2.26E-06	6.63E-06

TABLE S4. MAIN EFFECT LOD SCORES FOR THE INTERVALS CONTAINING EPISTATIC TRAIT QTLS

Trait	Chromosome	SNP	Position (Mb)	LOD	p value
Lymphocytes (%)	4	rs13477644	35.28	0.393	0.182
	6	rs13478739	47.03	9.535e-5	0.982
Red Cell Distribution Width	6	rs13478633	12.25	0.316	0.966
	6	rs13478641	16.06	0.345	0.926