The Effects of Mutations are Modified by Genetic Background in Mice

Christoph D. Rau, Natalia M. Gonzales, Danny Park, Abraham A. Palmer, Aldons J. Lusis, Noah Zaitlen*

Christoph D. Rau: Department of Anesthesiology, UCLA, chrau@ucla.edu

Natalia M. Gonzales: Department of Human Genetics, University of Chicago, ngon@uchicago.edu

Danny Park: Department of Medicine, UCSF, sihyung@gmail.com

Abraham A. Palmer: Department of Psychiatry, UCSD, aapalmer@ucsd.edu

Aldons J. Lusis: Department of Human Genetics, UCLA, jlusis@mednet.ucla.edu

Noah Zaitlen: Department of Neurology, UCLA, nzaitlen@mednet.ucla.edu, Corresponding Author

Abstract

Background: Commonly used genetic models of phenotype assume that alleles have similar effects in all individuals. Mounting evidence suggests that higher-order interactions, which may vary between populations, profoundly affect phenotypic variation. The effect of genetic and environmental perturbations may therefore differ across genetic backgrounds. In this manuscript, we develop a statistical test that determines whether the effect of a mutation on a complex phenotype changes as a function of ancestral background.

Results: We apply our test to two large cohorts of mice and observe 49 significant gene by ancestry interaction associations across 14 phenotypes as well as over 1,400 Bonferroni-corrected gene by ancestry interaction associations in gene expression data. We also observe evidence of rapid selection pressure on individual polymorphisms within one of the cohorts.

Conclusions: Unlike our prior work in human populations, we observe widespread evidence of ancestry-specific SNP effects, perhaps reflecting the greater divergence present in crosses among laboratory mice.

Keywords: Epistasis, Genetic Ancestry, GWAS, Mouse Genetics

Background

Genetic association studies in humans and model organisms have identified a number of significant links between individual polymorphisms and phenotypic variability. A fundamental assumption of many of these studies is that an allele will have a similar effect in each member of the population, that is, that epistatic and other higher-order interactions across the genome can largely be ignored[1–4]. While we have previously observed only modest evidence of ancestry specific genetic effects in humans[5], model organisms are often further diverged than human populations. For example, we observed radically different phenotypic consequences of null alleles of *Tcf7l2* and *Cacna1c* when expressed on different inbred strain backgrounds[6]. We therefore developed a test, designed for use in model organisms, to identify and examine heterogeneous SNP effects across different genetic backgrounds.

The field of mouse genetics provides numerous opportunities to identify and study the genetic basis of phenotypic variation in mice with obvious extensions to human populations. Many prior studies of epistasis in both mice and humans have used pairwise interactions between loci to attempt to identify interacting loci[4,7–9], yet this approach is hampered by a substantial increase in the threshold of significance due to the increased numbers of tests (from n to n^2, where n are the number of SNPs tested), which in turn requires larger cohorts to overcome. In this study, we approached epistasis in a different way by examining two distinct mouse cohorts and considering the effects of interactions between individual polymorphisms and global ancestry (Θ), which we defined as the percentage of the genome inherited from one of the parental lines of each cohort. We utilized a model we recently developed[5] which explored these interactions in human populations and extended this idea to model organisms by incorporating a linear mixed model capable of accounting for the highly structured relatedness of recombinant inbred (**RI**) lines[10]. Although our initial model was designed to identify both

genetic background and environmental effects, the controlled environment in which mouse cohorts live is intended to minimize sources of environmental variance. This test, which we have called GxΘ, acts as a surrogate measure of the concurrent action of many other SNPs, and allowed us to ask the question of whether or not SNP effect sizes change as a function of overall genetic ancestry. Our observed significant associations validate this hypothesis. In contrast to a pairwise epistatic test where significance means a detected interaction between two specific loci, a natural interpretation of a significant GxΘ interaction is widespread epistasis with the tested genotype G with many loci across the genome, as we showed previously[5].

We first evaluated the statistical properties of the method on simulated phenotypes from real recombinant inbred line data sets. The GxΘ test successfully distinguishes between effects dependent only on SNP or ancestry and effects which arise from their interaction. We then applied the test to two different populations of mice: recombinant inbred (**RI**) lines that are a subset of the Hybrid Mouse Diversity Panel (**HMDP**)[11], and a 50th generation intercross between the inbred LG/J and SM/J mouse strains[12], whose ancestry proportions can be clearly and precisely determined[12] Next, we conducted a replication study demonstrating that our approach is able to replicate findings across similar populations.

Finally, we examined the ancestry distribution at specific sites in the RI lines and observed that for many positions there is a statistically significant depletion or increase of individual ancestries. We interpret this as evidence of selection during the process of RI strain derivation and identify regions of the genome with strong selection up to 4 Mb. These regions are evidence of strong GxΘ interactions which inhibit or promote the transmission of these specific alleles to subsequent generations. These regions are enriched for genes involved in cancer and organogenesis and are enriched (P=1.8E-4) for metallopeptidases, which play key roles in fertility and neo/perinatal lethality[13–15]. We also observed several cases in which differential effect sizes were observed for SNP-ancestry interactions. This

highlights the important of epistasis in model organism genetics and validates the result of several prior studies which demonstrate profound differences in phenotypes and phenotypic responses to stressors (e.g. Park et al., 2018) based on ancestry.

Results

Method Overview:

Our objective is to determine whether a SNP has a different effect size a function genetic background in a mouse cross. As described in our previous work<cite>, this is effectively achieved through an interaction test between each SNP and global genetic ancestry Θ. Due to differences in relatedness structures of individuals in human populations and mouse crosses, we extended our approach with a linear mixed model to accommodate more complex relatedness structures (see Methods).

Simulated Data:

To examine the properties of our approach in a mouse cross, we applied the method to phenotypes simulated using real genotypes that reflect the underlying relatedness present within the recombinant inbred (RI) lines of the Hybrid Mouse Diversity Panel (HMDP). We tested Θ, the proportion of all SNPs arising from C57BL/6J mice for interaction with each individual SNP (see Methods). Power was calculated over 200 values of either a main SNP effect β_G or a SNP-ancestry interaction effect $\beta_{G\times\Theta}$ with 1,000 simulations each (200,000 total simulations).

Changes to either β_G or $\beta_{G\times\Theta}$ did not have an effect on the likelihood of identifying an association in the other term, indicating that these two terms are estimated correctly

independently of one another. We also explored incorporating a second genetic relationship matrix (GRM) accounting for population structure arising from descent from a single ancestor (the B6 mouse).[17] However, we observed no significant improvement in power when incorporating this second GRM (Figure 1). Consequently, we used a 1 GRM model in our analyses of real phenotypic data.

Real Data:

Mouse Populations

Next, we applied the method to two large panels of mice to identify GxΘ effects, where a given mutation interacts epistatically with one or many other loci, captured in the model by Θ, the global ancestry. Our first cohort is the HMDP, a set of 150+ commercially available inbred strains[18]. Numerous GWAS have been performed in the HMDP, including several using PYLMM, which forms the core of our algorithm[19–21]. The largest component of the HMDP is comprised of 122 RI strains (28 AxB, 71 BxD, 12 BxH, 11 CxB). Each RI was constructed from re-derivation of novel inbred lines via brother-sister mating following an F_2 cross between the sub-panel's parental lines. In the case of the HMDP, one of the parental strains for each RI strain was the commonly studied strainC57BL/6J (**B6**). Using B6 as the 'ancestral line', we calculated thetas for each RI strain (Fig 2a). As expected, the average ancestry attributable to B6 was roughly 50% (50.63%) and roughly normally distributed. We removed a single outlier, BXD32/TyJ, whose B6 Ancestry of 25.41% reflects a previously known additional backcross to DBA/2J, resulting in a strain that is 75% DBA and 25% B6. Each study using the HMDP uses a different subset of the entire panel, and we selected a study on heart failure induced by the chronic beta adrenergic agonist isoproterenol[22] for analysis as it used the most RI strains compared to other published HMDP data. We used 123 clinical phenotypes in conjunction with microarray-derived gene expressions measured in the left ventricle and on average tested 67 RI strains per phenotype.

The second cohort consists of 1,063 animals from the F_{50} – F_{56} generation of an advanced intercross line (**AIL**) created by crossing the LG/J and SM/J inbred mouse strains [12]. Unlike the RI strains from the HMDP, AILs are maintained in a manner that minimizes inbreeding. We arbitrarily set LG/J as the 'ancestral' strain of interest and calculated thetas for each of the 1,063 mice in the panel (Fig 2b). For this study, we focused on a diverse set of 133 phenotypes that had been measured in these mice. We describe the results of the GxΘ associations in each panel before demonstrating replication of signal in a phenotypic trait as well as in expression data.

Evidence of SNP x Ancestry Interactions in Phenotypic Data

We first applied the GxΘ method to the 123 observed heart-failure related phenotypes from the RI strains. We observed well-calibrated statistics, with λ_{GC} equal to 0.978 and $\lambda_{GxΘC}$ equal to 1.045. We observed 44 significant GxΘ loci across 9 phenotypes: E/A Ratio, Free Fatty Acid content in the blood, Cardiac Fibrosis, Fractional Shortening of the heart during contraction, Heart Rate, Internal Diameter of the Left Ventricle, left ventricular mass and left and right atrial weights (Table S1). These GxΘ loci were largely distinct from previously reported GWAS loci in the same phenotypes in this panel of mice[19,23], yet contained a number of highly relevant genes, as discussed below.

By way of example, we focus on two important phenotypes from the HMDP panel. Cardiac fibrosis is a marker of cardiac dysfunction. Genes identified through the GxΘ screen (Fig 3a) as potential candidates include: *Crisp2* (rs6295287 , p=6.74E-7), a secreted biomarker of cardiovascular disease[24], *Top2b* (rs31538570, p=2.14E-06) which plays a cardioprotective role in response to stress[25], *Rarb* (rs31538570, p=2.14E-06) a known regulator of inflammation with unknown function in the heart[26] and *Fibrosin* (rs33146511, p=2.54E-6) a major component of the fibrosis pathway[27]. Left ventricular mass increase in response to catecholamine challenge is the primary marker of cardiac hypertrophy in

the HMDP. A single GxΘ locus(rs31313229, p=2.91E-06) on chromosome 3 (Fig 3b) contains the gene *Lphn2*, which has a role in the promotion of cellular adhesion in response to external stimuli.[28]

We next applied the GxΘ method to 133 phenotypes measured in the LG/J x SM/J AIL. As with the HMDP data, we observed well-calibrated statistics in this cross, with λ_{GC} equal to 0.995 and $\lambda_{GxΘC}$ equal to 1.033. Despite the larger sample size, we observed only 5 significant loci across 4 phenotypes (activity levels in a saline-injected animal on day 3 of a conditioned place preference test; average weight of animals across 5 different time points roughly a week apart; glucose (mg/dL) in blood after a 4 hour fast [Fig 3c]; weight at ~68 days of age) (Table S1). This smaller number of GxΘ interactions could be due to the differences in strain as well as differences in phenotypes.

GxΘ Associations in HMDP Cohort Gene Expression

We next examined gene expression in the hearts of the HMDP cohort, where we had transcriptome microarrays from both a control and treated condition. Each cohort consists of approximately 70 RI lines, with 66 lines overlapping between the two cohorts (full lists of strains in Table S2). We examined 13,155 expressed and varying (CV > 5%) genes from the left ventricles of the HMDP cohorts using the ~170k SNPs (MAF <= 0.05). We observed 1,486 significant associations with 18 genes at a Bonferroni-corrected P value of 3.2E-10 (135,130 associations with 1,350 genes at GW-significant threshold of 4.2E-6) in the control cohort and 597 significant associations with 39 genes at the same threshold in the treated cohort (32,043 associations with 1,042 genes at 4.2E-6). The reduction of significant associations in the treated cohort is likely due to reduced power due to increased phenotypic variance caused by the treatment itself. Genes in the control cohort with significant GxΘ associations are enriched for mitochondrial genes (P=1.8E-10 (P=.032 in treated)), suggesting a role of GxΘ interactions in the regulation of mitochondrial dynamics. Genes in the treated cohort with significant

associations were enriched for genes involved in post-transcriptional modifications to RNA including RNA splicing (P=7.4E-3), highlighting the importance of alternative splicing to the response to catecholamine challenge[29].

Replication of GxΘ Associations across eQTLs from HMDP Cohort

Although the catecholamine treatment used to induce heart failure in this study does affect gene expression, the majority (12178, 92.6%) of genes were not significantly (Student's T-test) affected by the drug. To demonstrate that the method is able to replicate GxΘ results across cohorts, we examined the reproducibility of expression GxΘ QTLs in the treated and untreated RI lines of the HMDP. Of the 1,486 associations observed in the control data, we observe **305** (21%) with a GxΘ signal (P<.05) in the treated cohort (36 at FDR< 5%), suggesting strong replication between the two cohorts despite differences in genetic background and the reduced power in the treated cohort caused by the effects of the catecholamine drug.

Fitness

Strong GxΘ effects have the potential to affect overall organismal fitness. For example, strong GxΘ effects may lead to a differential frequency of strain-specific genotypes at individual loci, as those loci interact with the rest of the genome to cause changes to fitness and, consequently, retention of those sites. Alternately, it is possible that these variations may occur due to absolute differences in fitness of that SNP without GxΘ effects: however, both of these results are interesting and evidence of selection after admixture.

We conducted a test to search for individual loci with enriched or depleted B6 ancestry in the RI strains. As expected, the average B6 ancestry across all SNPs was 50.68% +/- 8.1%, which is indistinguishable from the ancestry by strain average of 50.63% +/- 6.7% (Fig 4a). At the level of individual loci, however, we observed significant variation in B6 ancestry across the genome (Fig 4b),

with some SNPs displaying very low or very high frequency (Fig 4c). We calculated the statistical likelihood of detecting the observed ancestries at all loci across the genome (λ=1.07, Fig 4d) and identified 614 SNPs from the 170k original SNPs with significantly altered ancestries at a FDR of 5% (Table S3). Nine loci were identified in which 5 or more SNPs were located together (Table 1). These 9 regions contain the majority (525, 86%) of all significant affected loci. These loci are evidence of an association between a gene near their peak and a loss or gain of overall organismal fitness. One region of particular interest lies on chromosome 18 and contains 246 SNPs (40% of all significantly altered SNPs) with an average B6 ancestry of 82.5%. SNPs in this region are found between B6 and C3H/HeJ and B6 and BALB/cJ. This locus contains a number of interesting candidate loci relating to organismal fitness, the most striking of which is *Epc1*, a transcription factor linked to DNA repair, muscle differentiation and cancer suppression and the only gene within the locus which has suggestive *cis*-GxΘQTL (P=0.014)[30,31].

When all genes within 2 MB (the average LD block size in the HMDP[32]) are examined as a whole, we observe significant enrichments for genes involved in cancer (Ps range from 1.3E-8 to 1.6E-7), respiratory tube development (2.7E-4), heart development (7.6E-4), likely mediated through changes in Matrix Metalloproteinase activity (3.9E-8). These enzymes are canonically responsible for the regulation of the extracellular matrix, but have been linked to the modulation of responses to many bioactive molecules and have important roles in fertility, embryonic development and neo/perinatal mortality in addition to many other diseases[15].

Discussion

We present a test which we call GxΘ, that leverages admixed populations such as inbred mouse strains to identify epistatic interactions between a SNP and an unknown number of other loci summarized into a single genomic ancestry score, Θ (available at https://github.com/ChristophRau/GxTheta). One major

advantage of this approach is that, unlike other epistasis-focused association approaches, it does not increase the number of tests when compared to a typical GWAS, resulting in similar genome-wide significance thresholds.

The existence of epistatic interactions and their role in human diseases and phenotypes have been known for many years, with notable examples of gene-gene interactions in Alzheimers[33], Bardet-Biedl syndrome[34] as well as classic interactions governing hair color and skin pigmentation. Despite these examples and many more like them, epistatic interactions have proven difficult to find using association techniques[35], with several authors going so far as to claim that epistatic and other non-additive effects play at best a minor role in determining phenotypic variance[2,4,36]. One notable exception has been the success of using consomic strains of mice, where one or more chromosome in strain A is replaced with a chromosome from another strain B[37]. This approach, which increases or decreases the contribution to phenotypic variance by a given strain in a clear and controlled manner has resulted in numerous identifications of epistatic interactions[38,39].

Our method sought to circumvent the traditional pitfall in Epistasis GWAS by examining each SNP only once for an interaction with a global genomic ancestry. We examined two populations of mice, the recombinant inbred lines of the Hybrid Mouse Diversity Panel[32] and an AIL based on LG/J and SM/J[12]. Despite nearly ten times the number of genetically distinct mice, as well as a larger number of phenotypes and genotypes in the AIL, we observed approximately nine times more significant GxΘ peaks in the HMDP phenotypes (44 vs 5). Several possible reasons exist for this difference. First, outbred mice have significantly lower power due to increased phenotypic variance caused by high rates of heterozygosity at alleles while inbred populations have increased power due to a lack of this variance. This can also be seen in our regular GWAS results (see Supp)[12,22,40]. Second, the different phenotypes studied in each cohort will have different genetic architecture and experimental noise (e.g. the AIL study includes many behavioral traits while the HMDP does not). Third, the genetic backgrounds of the two

cohorts are different, which might also contribute to differences in observed numbers of GxΘ interactions. Although individually, LG and SM are only slightly less genetically diverged from one another as any pair of strains that make up the RI panels of the HMDP (Table S4), the presence of five ancestral lines in the HMDP compared to only two in the F50 cross results in much more genetic diversity in the HMDP compared to the F50 mice. Finally, we observe a higher variance in ancestral background in the HMDP when compared to the AIL. As our method relies on differences in ancestral background to identify sites with different effect sizes in different genetic contexts, the reduced variance in the AIL lines necessarily corresponds to a decrease in the power to detect GxΘ interactions. Taken together, our method is best suited to datasets with relatively low heterozygosity, clear and numerous differences in genetic background, and with higher variance in the percentage of SNPs attributable to a given ancestry.

Conclusion

The results of our study suggest that heterogenous SNP effects due to differing ancestries is pervasive in mouse populations, especially in diverse populations such as the HMDP. This observation matches prior observation of significant epistatic interactions in inbred strains of mice as well as examples in human studies[5]. Further analyses of these heterogenous-effect SNPs may reveal novel epistatic interactions which drive phenotypic expression, and suggests that careful attention to genetic ancestry should be considered when studying the role of an individual polymorphism on a phenotype.

Methods

Mouse Populations and Ancestry Calling

Mouse data were drawn from previously reported studies[12,41]. The LGxSM AIL consists of 1,063 G50-56 mice derived from an original F1 intercross between the LG and SM inbred lines. The Hybrid Mouse Diversity Panel consists of over 150 strains of commercially available inbred mice[32], of which 122 strains were recombinant inbred lines and suitable for our study. SNPwise ancestries were determined by identifying all SNPs which differed between parental lines (AIL: LG and SM or HMDP:C57BL/6 and A, C3H, DBA/2 or BALB/c). Genotypes from the G50-56 or RI lines were filtered for these SNPs and ancestries calculated using either LG or C57BL/6 as the strain background of interest.

Simulation Framework

We created sets of simulated phenotypes based on the genotypes of the HDMP RI panel, which is an admixed population in which the B6 strain, on average, contributes 50% of each strain's DNA. For each simulated phenotypes, we drew a SNP (MAF > 5%) at random from the HMDP genotypes and created a phenotype based on β, the genetic effect size, , φ the effect size of the interaction between global ancestry (Θ) and the chosen SNP and three variance terms: σ_g^2, the proportion of variance attributable to genetic effects σ_Θ^2 , the proportion of variance attributable to GxΘ effects and σ_ε^2, the residual proportional variance attributable to all combined sources of error and variation not considered in this study. Phenotypes were generated both with and without the GxΘ variance term to ascertain the necessity of incorporating a second GRM ($\mathbf{K}^A$) into the algorithm.

We simulated four distinct phenotypes for our analysis

1) Phenotypes generated by including a SNP Effect

$$y = \beta X + mvn\left(0, \sigma_g^2 \boldsymbol{K} + \sigma_e^2 \boldsymbol{I}\right) \text{ or}$$

$$y = \beta X + mvn(0, \sigma_g^2 \boldsymbol{K} + \sigma_\theta^2 \boldsymbol{K^A} + \sigma_e^2 \boldsymbol{I})$$

2) Phenotypes generated by including a GxΘ Effect

$$y = \varphi\theta * X + mvn(0, \sigma_g^2 \boldsymbol{K} + \sigma_e^2 \boldsymbol{I}) \text{ or}$$

$$y = \varphi\theta * X + mvn(0, \sigma_g^2 \boldsymbol{K} + \sigma_\theta^2 \boldsymbol{K^A} + \sigma_e^2 \boldsymbol{I})$$

In each phenotype, σ_g^2 was set to 0.4. When incorporated, σ_θ^2 was set to 0.2 and σ_ε^2 was set to the remainder of the variance (0.6 or 0.4). The power of our model and independence of our β and φ terms were queried by varying either β or φ from 0 to 1 (200 values set 0.005 apart) with 1,000 simulated phenotypes at each step (200,000 total simulations per phenotype).

GxΘ Model

The equation to determine the effects of a SNP and a SNP x Ancestry term on a phenotype can be written as:

$$\boldsymbol{y_k = \mu + \sum_{i=1}^{M} \beta_i X_i + \delta\theta_K + \sum_{i=1}^{M} \varphi_i\, \theta_i X_{ik} + \varepsilon_k}$$

Where y_k is the phenotype of person k, μ is the mean phenotypic value, M is the number of markers, β are the weights on the SNPs, X is the *mxn* array of SNP genotypes, δ is the global weight of the ancestry effect, Θ is the ancestries for all N individuals. φ are the weights of the GxΘ effect and ε is the combined error term. We want to identify SNPs where $\varphi_i \neq 0$ as these are sites where Ancestry is interacting with our genotypes.

We can rewrite the above as:

$$\boldsymbol{y} = \mu + \beta_i \boldsymbol{X_i} + \delta\boldsymbol{\theta} + \varphi_i \boldsymbol{\theta} * \boldsymbol{X_i} + \boldsymbol{u} + \boldsymbol{e}$$

For an individual SNP i. Here, Θ is the column vector of ancestries, and Θ * X is the element-wise product. The random effect u accounts for relatedness of individuals based on SNPs. Our GxΘ test is then a LRT test with a null of $\varphi_i = 0$ and an alternate of $\varphi_i \neq 0$.

Declarations:

Ethics Approval and Consent to Participate: Not Applicable

Consent for Publication: Not Applicable

Availability of Data and Materials:

Data from the HMDP may be accessed at https://systems.genetics.ucla.edu/

Data from the AIL cross may be accessed at http://palmerlab.org/protocols-data/

The GxΘ algorithm may be found at https://github.com/ChristophRau/GxTheta

Competing Interests: The authors declare they have no competing interests

Funding: This work was supported by NIDA (AAP: R01DA021336; NMG: F31DA036358), NIGMS (NMG: T32GM007197), and the NIH (CDR: K99HL138301, NZ: K25HL121295, U01HG009080, and R01HG006399)). The funding bodies played no role in the study other than funding.

Author Contributions:

CDR designed the project, created the GxΘ software, acquired the HMDP data and analyzed and interpreted it and the AIL data and drafted the manuscript. NMG assisted in the design of the AIL cohort, acquired the AIL data and substantially edited the manuscript. DP assisted in conceptualizing the GxΘ algorithm and substantially revised the manuscript. AAP assisted in the design of the AIL cohort and substantially revised the manuscript. AJL assisted in the design of the HMDP cohort and substantially revised the manuscript. NZ conceptualized and designed the GxΘ algorithm, interpreted

the data and helped to draft and substantially revise the manuscript. All authors read and approved the final manuscript.

Acknowledgements:

The authors would like to thank Yibin Wang for his constructive critiques of the algorithm and the manuscript.

References

1. Polderman TJC, Benyamin B, de Leeuw C a, Sullivan PF, van Bochoven A, Visscher PM, Posthuma D. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*. 2015;47(7):702–709.
2. Hill WG, Goddard ME, Visscher PM. Data and theory point to mainly additive genetic variance for complex traits. *PLoS genetics*. 2008;4(2):e1000008.
3. Fish AE, Capra JA, Bush WS. Are Interactions between cis-Regulatory Variants Evidence for Biological Epistasis or Statistical Artifacts? *American journal of human genetics*. 2016;99(4):817–830.
4. Mäki-Tanila A, Hill WG. Influence of gene interaction on complex trait variation with multilocus models. *Genetics*. 2014;198(1):355–67.
5. Park DS, Eskin I, Kang EY, Gamazon ER, Eng C, Gignoux CR, Galanter JM, Burchard E, Ye CJ, Aschard H, Eskin E, Halperin E, Zaitlen N. An ancestry-based approach for detecting interactions. *Genetic Epidemiology*. 2018;42(1):49–63.
6. Sittig LJ, Carbonetto P, Engel KA, Krauss KS, Barrios-Camacho CM, Palmer AA. Genetic Background Limits Generalizability of Genotype-Phenotype Relationships. *Neuron*. 2016;91(6):1253–1259.
7. Tyler AL, Donahue LR, Churchill GA, Carter GW. Weak Epistasis Generally Stabilizes Phenotypes in a Mouse Intercross. *PLoS genetics*. 2016;12(2):e1005805.
8. Logsdon B a, Hoffman GE, Mezey JG. A variational Bayes algorithm for fast and accurate multiple locus genome-wide association analysis. *BMC bioinformatics*. 2010;11:58.
9. Varón-González C, Navarro N. Epistasis regulates the developmental stability of the mouse craniofacial shape. *Heredity*. 2018:1.
10. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. Efficient control of population structure in model organism association mapping. *Genetics*. 2008;178(3):1709–1723.
11. Lusis AJ, Seldin MM, Allayee H, Bennett BJ, Civelek M, Davis RC, Eskin E, Farber CR, Hui ST, Mehrabian M, Norheim F, Pan C, Parks B, Rau CD, Smith DJ, et al. The Hybrid Mouse Diversity Panel: A Resource for Systems Genetics Analyses of Metabolic and Cardiovascular Traits. *Journal of lipid research*. 2016;58(12):7250–7.
12. Gonzales NM, Seo J, Hernandez-Cordero AI, Pierre CL St., Gregory JS, Distler MG, Abney M, Canzar S, Lionikas A, Palmer AA. Genome wide association analysis in a mouse advanced intercross line. *Nature Communications*. 2018:in press.
13. Jha SK, Rauniyar K, Jeltsch M. Key molecules in lymphatic development, function, and identification. *Annals of Anatomy - Anatomischer Anzeiger*. 2018;219:25–34.

14. Afratis NA, Selman M, Pardo A, Sagi I. Emerging insights into the role of matrix metalloproteases as therapeutic targets in fibrosis. *Matrix Biology*. 2018;68–69:167–179.
15. Cerdà-Costa N, Gomis-Rüth FX. Architecture and function of metallopeptidase catalytic domains. *Protein science : a publication of the Protein Society*. 2014;23(2):123–44.
16. Park S, Ranjbarvazirj S, Lay FD, Zhao P, Miller MJ, Dhaliwal JS, Huertas-Vazquez A, Wu X, Qiao R, Soffer JM, Mikkola HKA, Lusis AJ, Ardehali R. Genetic Regulation of Fibroblast Activation and Proliferation in Cardiac Fibrosis. *Circulation*. 2018:CIRCULATIONAHA.118.035420.
17. Sul JH, Bilow M, Yang W-Y, Kostem E, Furlotte N, He D, Eskin E. Accounting for Population Structure in Gene-by-Environment Interactions in Genome-Wide Association Studies Using Mixed Models. Schork NJ, ed. *PLOS Genetics*. 2016;12(3):e1005849.
18. Ghazalpour A, Rau CD, Farber CR, Bennett BJ, Orozco LD, Van Nas A, Pan C, Allayee H, Beaven SW, Civelek M, Davis RC, Drake TA, Friedman RA, Furlotte N, Hui ST, et al. Hybrid mouse diversity panel: A panel of inbred mouse strains suitable for analysis of complex genetic traits. *Mammalian Genome*. 2012;23(9–10).
19. Rau CD, Wang J, Avetisyan R, Romay MC, Martin L, Ren S, Wang Y, Lusis AJ. Mapping genetic contributions to cardiac pathology induced by beta-adrenergic stimulation in mice. *Circulation: Cardiovascular Genetics*. 2015;8(1).
20. Bennett BJ, Farber CR, Orozco L, Kang HM, Ghazalpour A, Siemers N, Neubauer M, Neuhaus I, Yordanova R, Guan B, Truong A, Yang W, He A, Kayne P, Gargalovic P, et al. A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome research*. 2010;20(2):281–290.
21. Parks BW, Sallam T, Mehrabian M, Psychogios N, Hui ST, Norheim F, Castellani LW, Rau CD, Pan C, Phun J, Zhou Z, Yang W-P, Neuhaus I, Gargalovic PS, Kirchgessner TG, et al. Genetic architecture of insulin resistance in the mouse. *Cell Metabolism*. 2015;21(2).
22. Rau CD, Wang J, Avetisyan R, Romay MC, Martin L, Ren S, Wang Y, Lusis AJ. Mapping genetic contributions to cardiac pathology induced by beta-adrenergic stimulation in mice. *Circulation: Cardiovascular Genetics*. 2015;8(1):40–49.
23. Wang JJ-CC, Rau C, Avetisyan R, Ren S, Romay MC, Stolin G, Gong KW, Wang Y, Lusis AJ. Genetic Dissection of Cardiac Remodeling in an Isoproterenol-Induced Heart Failure Mouse Model. *PLoS Genetics*. 2016;12(7):1–30.
24. Gibbs GM, Scanlon MJ, Swarbrick J, Curtis S, Gallant E, Dulhunty AF, O'Bryan MK. The cysteine-rich secretory protein domain of Tpx-1 is related to ion channel toxins and regulates ryanodine receptor Ca2+ signaling. *The Journal of biological chemistry*. 2006;281(7):4156–63.
25. Zhang S, Liu X, Bawa-Khalfe T, Lu L-S, Lyu YL, Liu LF, Yeh ETH. Identification of the molecular basis of doxorubicin-induced cardiotoxicity. *Nature Medicine*. 2012;18(11):1639–1642.
26. Albino-Sanchez M, Vazquez-Hernandez J, Ocadiz-Delgado R, Serafin-Higuera N, León-Galicia I, Garcia-Villa E, Hernandez-Pando R, Gariglio P. Decreased RARβ expression induces abundant inflammation and cervical precancerous lesions. *Experimental Cell Research*. 2016;346(1):40–52.
27. Wyler DJ. Fibrosin, a Novel Fibrogenic Protein: Discovery, Cloning and Implications for Fibrotic Disorders. *International Archives of Allergy and Immunology*. 1996;111(4):326–329.
28. White GR, Varley JM, Heighway J. Genomic structure and expression profile of LPHH1, a 7TM gene variably expressed in breast cancer cell lines. *Biochimica et biophysica acta*. 2000;1491(1–3):75–92.
29. Lara-Pezzi E, Gómez-Salinero J, Gatto A, García-Pavía P. The Alternative Heart: Impact of Alternative Splicing in Heart Disease. *Journal of Cardiovascular Translational Research*. 2013;6(6):945–955.
30. Wang Y, Alla V, Goody D, Gupta SK, Spitschak A, Wolkenhauer O, Pützer BM, Engelmann D. Epigenetic factor EPC1 is a master regulator of DNA damage response by interacting with E2F1 to

silence death and activate metastasis-related gene signatures. *Nucleic Acids Research*. 2016;44(1):117–133.

31. Kee HJ, Kim J-R, Nam K-I, Park HY, Shin S, Kim JC, Shimono Y, Takahashi M, Jeong MH, Kim N, Kim KK, Kook H. Enhancer of Polycomb1, a Novel Homeodomain Only Protein-binding Partner, Induces Skeletal Muscle Differentiation. *Journal of Biological Chemistry*. 2007;282(10):7700–7709.
32. Ghazalpour A, Rau CD, Farber CR, Bennett BJ, Orozco LD, van Nas A, Pan C, Allayee H, Beaven SW, Civelek M, Davis RC, Drake T a, Friedman R a, Furlotte N, Hui ST, et al. Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. *Mammalian genome : official journal of the International Mammalian Genome Society*. 2012;23(9–10):680–92.
33. Combarros O, Cortina-Borja M, Smith AD, Lehmann DJ. Epistasis in sporadic Alzheimer's disease. *Neurobiology of Aging*. 2009;30(9):1333–1349.
34. Badano JL, Leitch CC, Ansley SJ, May-Simera H, Lawson S, Lewis RA, Beales PL, Dietz HC, Fisher S, Katsanis N. Dissection of epistasis in oligogenic Bardet–Biedl syndrome. *Nature*. 2006;439(7074):326–330.
35. Mackay TF, Moore JH. Why epistasis is important for tackling complex human disease genetics. *Genome Medicine*. 2014;6(6):125.
36. Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L. Finding the sources of missing heritability in a yeast cross. *Nature*. 2013;494(7436):234–237.
37. Buchner DA, Nadeau JH. Contrasting genetic architectures in different mouse reference populations used for studying complex traits. *Genome research*. 2015;25(6):775–91.
38. Ackerman KG, Huang H, Grasemann H, Puma C, Singer JB, Hill AE, Lander E, Nadeau JH, Churchill GA, Drazen JM, Beier DR. Interacting genetic loci cause airway hyperresponsiveness. *Physiological Genomics*. 2005;21(1):105–111.
39. Street VA, Kujawa SG, Manichaikul A, Broman KW, Kallman JC, Shilling DJ, Iwata AJ, Robinson LC, Robbins CA, Li J, Liberman MC, Tempel BL. Resistance to Noise-Induced Hearing Loss in 129S6 and MOLF Mice: Identification of Independent, Overlapping, and Interacting Chromosomal Regions. *Journal of the Association for Research in Otolaryngology*. 2014;15(5):721–738.
40. Wang JJ-C, Rau C, Avetisyan R, Ren S, Romay MC, Stolin G, Gong KW, Wang Y, Lusis AJ. Genetic Dissection of Cardiac Remodeling in an Isoproterenol-Induced Heart Failure Mouse Model. *PLoS Genetics*. 2016;12(7).
41. Rau CD, Wang J, Avetisyan R, Romay MC, Martin L, Ren S, Wang Y, Lusis a. J. Mapping Genetic Contributions to Cardiac Pathology Induced by Beta-Adrenergic Stimulation in Mice. *Circulation: Cardiovascular Genetics*. 2015;8(1):40–49.

Figure 1. Power Calculations Using the GxΘ algorithm. Power calculations based on simulated data with variable main SNP effects β_G (Left panel) or variable Ancestry-SNP effects $\beta_{GxΘ}$ (Right panel). Blue and Purple power curves are the power curves for detecting a significant SNP effect, while the orange and green curves are the power curves for detecting a significant GxΘ effect. Two phenotypic models, one incorporating 1 GRM (1K) correcting for relatedness in the SNPs (green, purple) and one incorporating 2 GRMs (2K) correcting for relatedness in both SNPs and Ancestry (red, blue) were used.

Figure 2. Ancestral Strain Contributions by Strain. A) The 122 strains of the RI panel of the HMDP, **B)** The 1063 animals in the F_{50} – F_{56} generation of the Palmer Cross

Figure 3. GxΘ Results in the HMDP and LGxSM AIL Cohorts. A) Cardiac Fibrosis B) Left Ventricular Weight C) Glucose. Significance Threshold in HMDP (A,B)= 4.2E-6[20], AIL (C)=9.1E-6 (FDR 5%)

Figure 4. Fitness of individual sites across the genome. A) A histogram of B6 ancestry across the geneome. **B)** Genome-wide plot of ancestry **C)** Zoomed in region on chr 12 with dramatic depletion of B6 allele. **D)** QQ plot of all Pvalues. λ=1.07

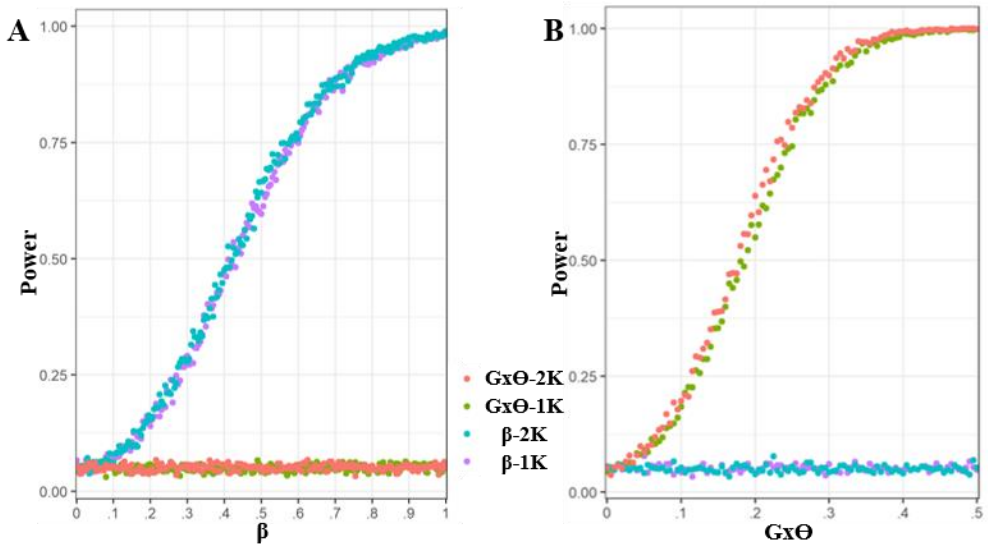
A
B
Power
β
GxΘ
GxΘ-2K
GxΘ-1K
β-2K
β-1K
1.00
0.75
0.50
0.25
0.00
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1

Figure 1

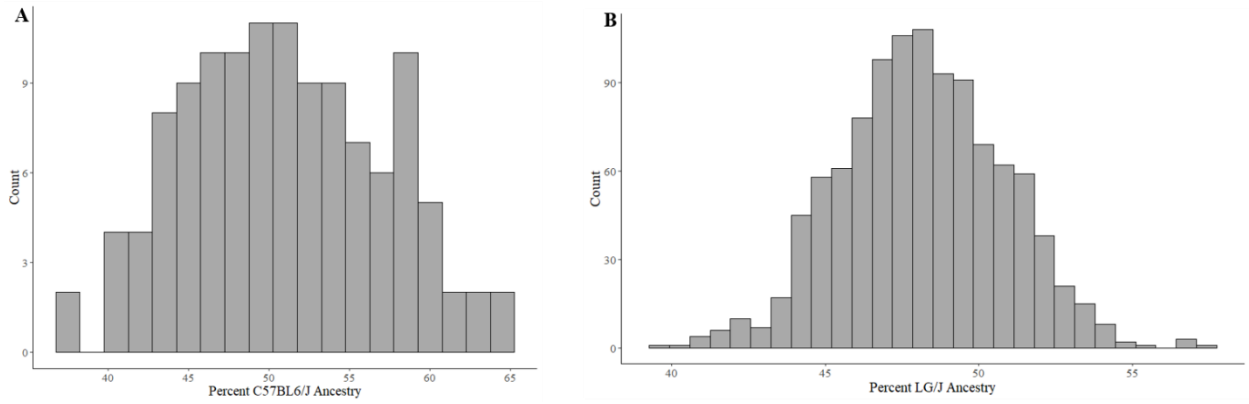
A
Count
0
3
6
9
40
45
50
55
60
65
Percent C57BL6/J Ancestry
B
Count
0
30
60
90
40
45
50
55
Percent LG/J Ancestry

Figure 2

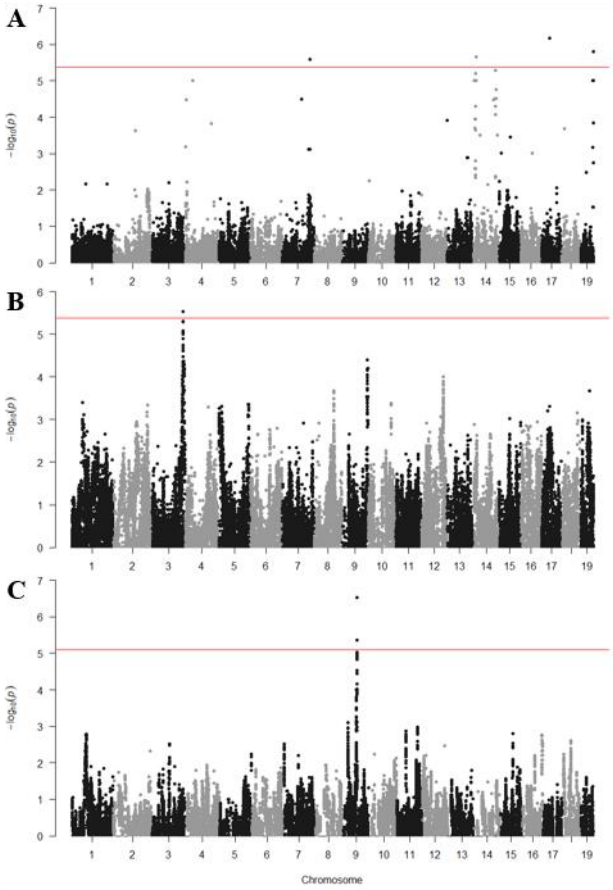

Figure 3

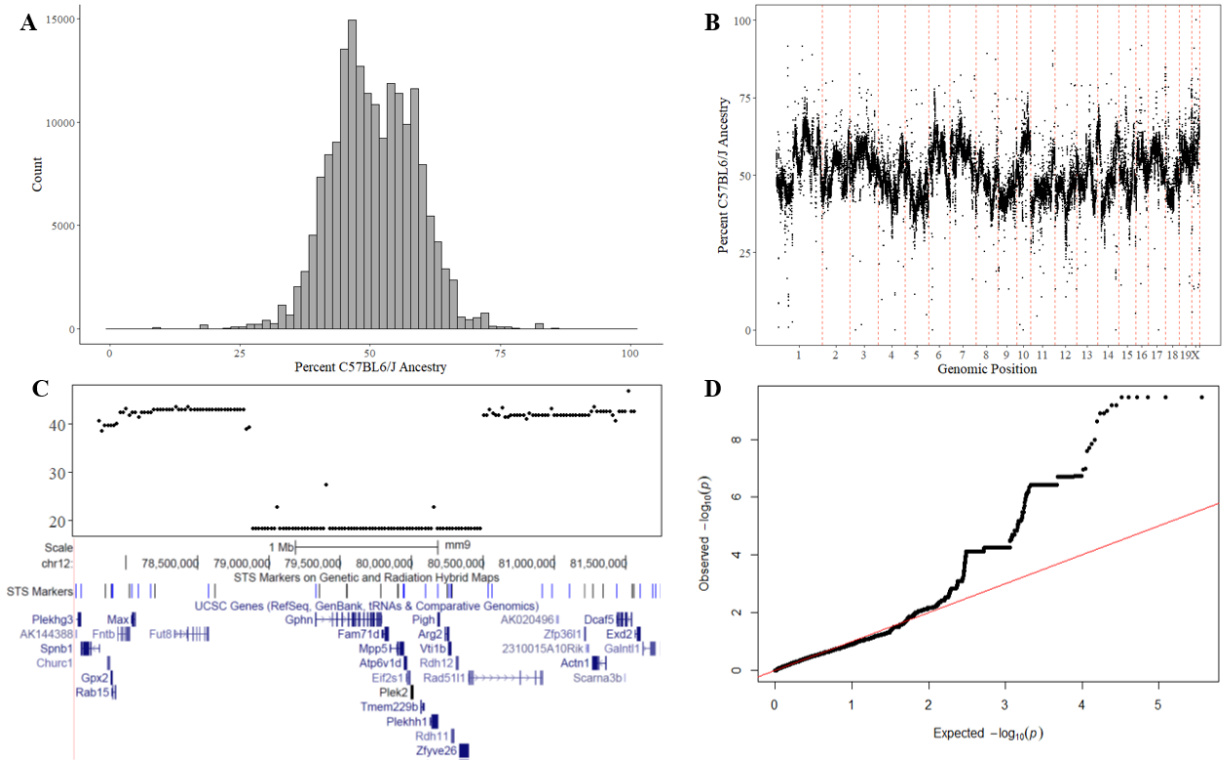
A
Count
Percent C57BL6/J Ancestry
B
Percent C57BL6/J Ancestry
Genomic Position
C
Scale
chr12:
1 Mb
mm9
STS Markers
STS Markers on Genetic and Radiation Hybrid Maps
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
D
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

Figure 4

Table 1. **Regions with Significant and Sustained B6 Depletion or Selection**

Chromosome	Start	End	Number of SNPs	Average B6 Ancestry	Genes of Interest
1	10763382	10945149	21	10.66	*Cpa6*
1	59458339	60046978	54	16.22	*Bmpr2*
1	134463208	134712156	10	87.52	*Nfasc*
11	64495476	64538963	5	18.18	*Myocd*
11	110824338	110984075	18	85.75	*Kcnj2*
12	79002481	80362465	72	18.43	*Gphn*
12	106889886	107633641	49	19.15	*Pigh*
15	4743169	6121864	50	18.14	*Prkaa1*
18	5465201	8965173	246	82.50	*Epc1*