Identifying Genetic Loci Whose Effect on Phenotype are Influenced by Changes in Genetic Background

Christoph D. Rau[1], Natalia M. Gonzales[2], Joshua S. Bloom[1], Julien Ayroles[3], Danny Park[4], Abraham A. Palmer[2], Aldons J. Lusis[1] and Noah Zaitlen[1]

1: UC Los Angeles 2: UC San Diego 3: Princeton 4: UC San Francisco

Motivation

Quantitative genetic models of phenotype used to map complex traits frequently assume that allelic effect sizes are fixed in a given population and do not vary from individual to individual. Numerous examples exist, however, of epistatic interactions in which two or more alleles interact in a non-additive fashion to affect phenotypic variation. These interactions suggest that the effects of genetic and environmental perturbations may vary across a population.
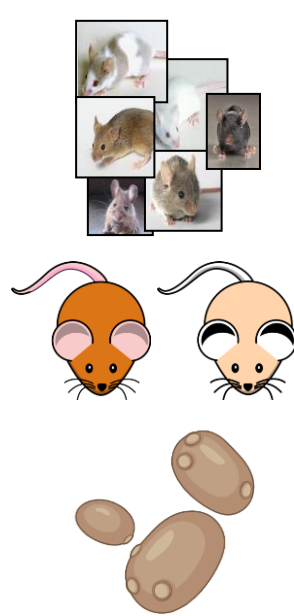
To-date, most available methods to detect epistatic interactions focus on detecting pair-wise interactions between individual polymorphisms. By contrast, we have developed a statistical test which we call Gene By Ancestry (GxΘ) that determines whether the effect of a polymorphism on a complex phenotype changes as a function of a definable ancestral background such as one found in a model organism or an admixed human population.
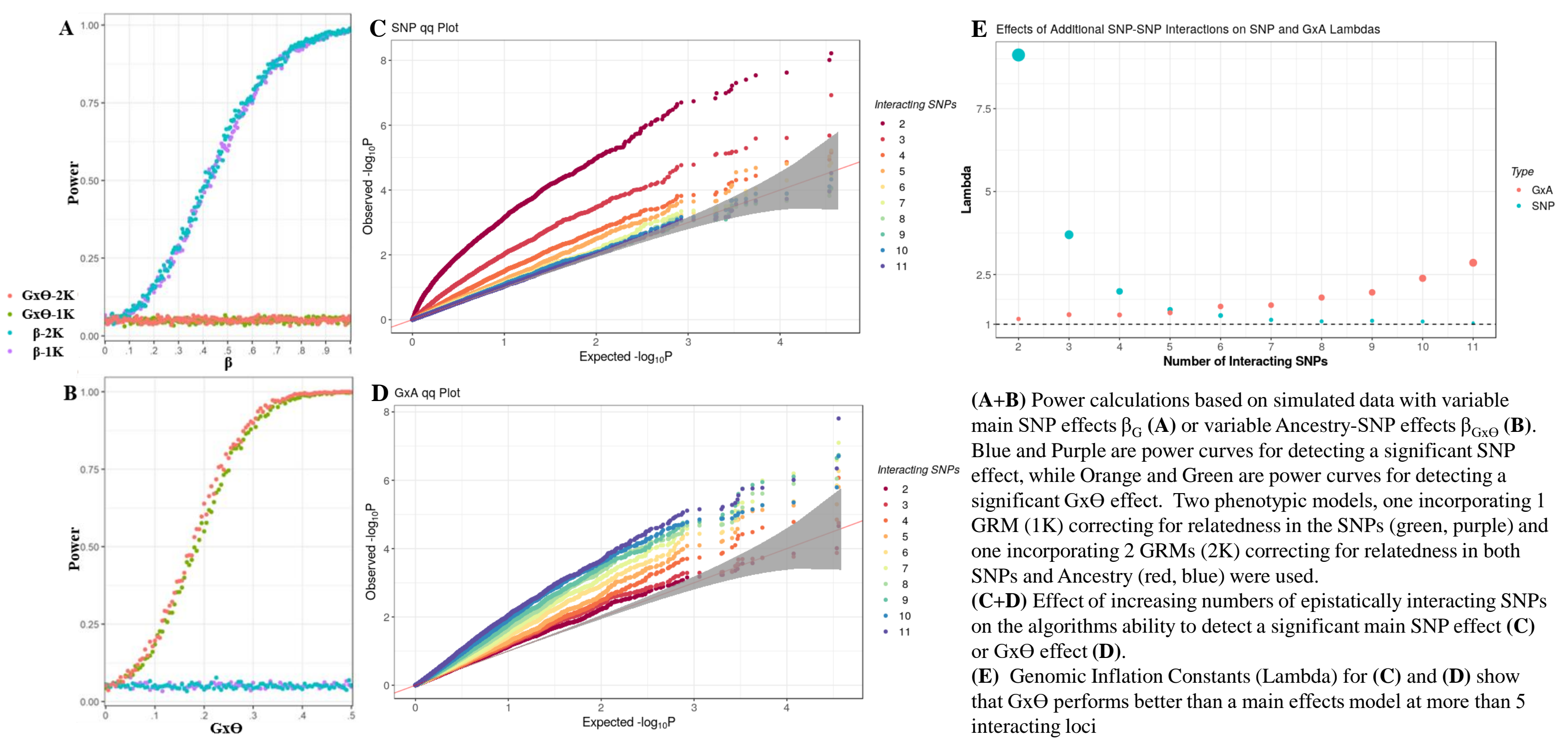
Populations

122 Recombinant Inbred Strains
Of the **Hybrid Mouse Diversity Panel**

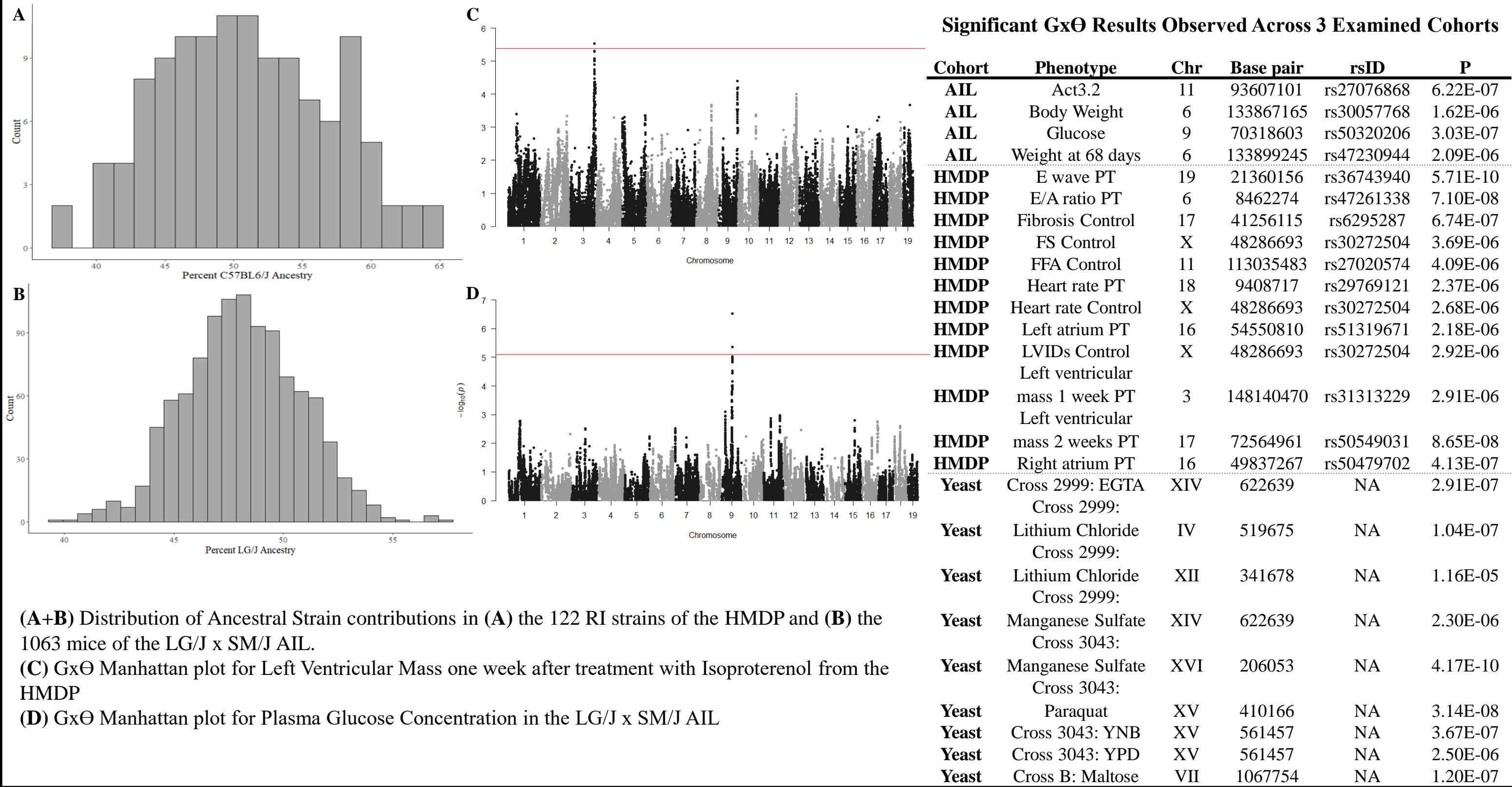1063 mice from an AIL between
LG/J and **SM/J**

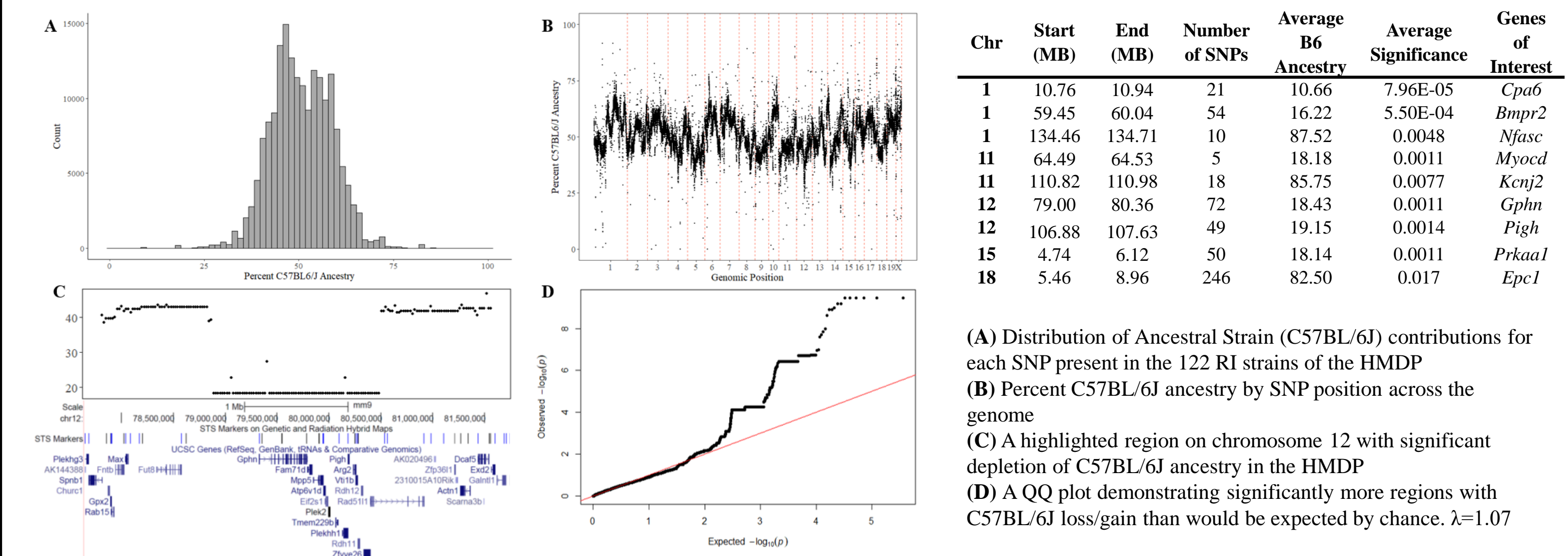15 **yeast crosses** each with between
650 – 950 progeny per cross

Model

The standard model to determine the effect of a set of SNPs on a phenotype can written as

$$y_k = \mu + \sum_{i=1}^{M} \beta_i X_i + \varepsilon_k$$

Where y_k is the phenotype of individual k, μ is the mean phenotypic value, M is the number of markers, β are the weights on the SNPs, X is the *m by n* array of SNP genotypes and ε is the combined error term. The effect of an individual SNP i on a phenotype can then be written as

$$y = \mu + \beta_i X_i + u + e$$

Where y is the vector of all phenotypes and The random effect u accounts for relatedness of individuals based on SNPs.

Motivated by the above, we add an Ancestry term (Θ), defined in our model as the percentage of a definable ancestral population (eg a known mouse strain) identified within each individual, as well as an interaction term between Θ and genetics X.

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

Where δ is the global weight of the ancestry effect, Θ are the ancestries for all N individuals and ϕ are the weights of the GxΘ effect. We want to identify SNPs where $\varphi_i \neq 0$ as these are sites where Ancestry is interacting with our genotypes. Motivated by our model above, we can write a new model for the effect of a single SNP i on a phenotypic trait as:

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

Here, Θ is the column vector of ancestries, and Θ * X is the element-wise product. Our GxΘ test is then a LRT test with a null of $\varphi_i = 0$ and an alternate of $\varphi_i \neq 0$.

Simulation

(A+B) Power calculations based on simulated data with variable main SNP effects β_G (A) or variable Ancestry-SNP effects $\beta_{GxΘ}$ (B). Blue and Purple are power curves for detecting a significant SNP effect, while Orange and Green are 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.
(C+D) Effect of increasing numbers of epistatically interacting SNPs on the algorithms ability to detect a significant main SNP effect (C) or GxΘ effect (D).
(E) Genomic Inflation Constants (Lambda) for (C) and (D) show that GxΘ performs better than a main effects model at more than 5 interacting loci.

GxΘ Identifies Loci that Interact with Genetic Ancestry

(A+B) Distribution of Ancestral Strain contributions in (A) the 122 RI strains of the HMDP and (B) the 1063 mice of the LG/J x SM/J AIL.
(C) GxΘ Manhattan plot for Left Ventricular Mass one week after treatment with Isoproterenol from the HMDP
(D) GxΘ Manhattan plot for Plasma Glucose Concentration in the LG/J x SM/J AIL

Significant GxΘ Results Observed Across 3 Examined Cohorts

Cohort	Phenotype	Chr	Base pair	rsID	P
AIL	Act3.2	11	93607101	rs27076868	6.22E-07
AIL	Body Weight	6	133867165	rs30057768	1.62E-06
AIL	Glucose	9	70318603	rs50320206	3.03E-07
AIL	Weight at 68 days	6	133899245	rs47230944	2.09E-06
HMDP	E wave PT	19	21360156	rs36743940	5.71E-10
HMDP	E/A ratio PT	6	8462274	rs47261338	7.10E-08
HMDP	Fibrosis Control	17	41256115	rs6295287	6.74E-07
HMDP	FS Control	X	48286693	rs30272504	3.69E-06
HMDP	FFA Control	11	113035483	rs27020574	4.09E-06
HMDP	Heart rate PT	18	9408717	rs29769121	2.37E-06
HMDP	Heart rate Control	X	48286693	rs30272504	2.68E-06
HMDP	Left atrium PT	16	54550810	rs51319671	2.18E-06
HMDP	LVIDs Control	X	48286693	rs30272504	2.92E-06
HMDP	Left ventricular mass 1 week PT	3	148140470	rs31313229	2.91E-06
HMDP	Left ventricular mass 2 weeks PT	17	72564961	rs50549031	8.65E-08
HMDP	Right atrium PT	16	49837267	rs50479702	4.13E-07
Yeast	Cross 2999: EGTA	XIV	622639	NA	2.91E-07
Yeast	Cross 2999: Lithium Chloride	IV	519675	NA	1.04E-07
Yeast	Cross 2999: Lithium Chloride	XII	341678	NA	1.16E-05
Yeast	Cross 2999: Manganese Sulfate	XIV	622639	NA	2.30E-06
Yeast	Cross 3043: Manganese Sulfate	XVI	206053	NA	4.17E-10
Yeast	Cross 3043: Paraquat	XV	410166	NA	3.14E-08
Yeast	Cross 3043: YNB	XV	561457	NA	3.67E-07
Yeast	Cross 3043: YPD	XV	561457	NA	2.50E-06
Yeast	Cross B: Maltose	VII	1067754	NA	1.20E-07

Identification of Selection in HMDP RI Strains

Chr	Start (MB)	End (MB)	Number of SNPs	Average B6 Ancestry	Average Significance	Genes of Interest
1	10.76	10.94	21	10.66	7.96E-06	*Cpa6*
1	59.45	60.04	54	16.22	5.50E-04	*Bmpr2*
1	134.46	134.71	10	87.52	0.0048	*Nfasc*
11	64.49	64.53	5	18.18	0.0011	*Myocd*
11	110.82	110.98	18	85.75	0.0077	*Kcnj2*
12	79.00	80.36	72	18.43	0.0011	*Gphn*
12	106.88	107.63	49	19.15	0.0014	*Pigh*
15	4.74	6.12	50	18.14	0.0011	*Prkaa1*
18	5.46	8.96	246	82.50	0.017	*Epc1*

(A) Distribution of Ancestral Strain (C57BL/6J) contributions for each SNP present in the 122 RI strains of the HMDP
(B) Percent C57BL/6J ancestry by SNP position across the genome
(C) A highlighted region on chromosome 12 with significant depletion of C57BL/6J ancestry in the HMDP
(D) A QQ plot demonstrating significantly more regions with C57BL/6J loss/gain than would be expected by chance. λ=1.07

Conclusions

We have applied GxΘ to data from two model organisms. We observe that the number of identifiable GxΘ effects appears to decrease as a function of phenotypic and organismal complexity, with the greatest number of hits observed in phenotypes relating to organismal survival in yeast (14/15 crosses and 38/40 phenotypes), then gene expression in mice, then physiological traits in mice and finally behavioral traits in mice. Analysis of the Recombinant Inbred panels of the HMDP reveal evidence of regions of the C57BL/6J genome which are selected for or against during RI strain derivation, another example of epistatic interactions leading to significant phenotypic effects.
GxΘ is available at: https://github.com/ChristophRau/GxTheta